Designation: E 2185 – 01

An American National Standard

Standard Specification for Transferring Digital Voice Data Between Independent Digital Dictation Systems and Workstations[1]

This standard is issued under the fixed designation E 2185; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ϵ) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This specification covers the format and content of digitally recorded voice data files and their identifying data. The object is to enable transfer between independent digital dictation systems and workstations, regardless of manufacturer and protocols for ensuring reliability. This specification does not cover the transmission of voice data files and their identifying data within digital dictation systems and workstations or their transcription into text files.

1.2 This specification may be applied to either the transmission of data over medium- to high-speed data communication networks or to the transmission of data by recording on, and later playing back from, magnetic or optical digital storage media. It defines the blocked stream of data, called a message, which is transmitted over a network connection or recorded on a storage medium. It does not define the hardware or software network protocols or storage media formats needed for message transmission (for example, see ISO 8072-1986) or the formats used to store data internally by the sender or receiver.

1.3 Since some standardization in storage media format and network protocols would help to promote the exchange of data between computer systems with diverse hardware and software, it is suggested that readily available universal media and formats be used for data exchange when possible.

1.4 Any considerations regarding the security of the digital dictation file or its components as defined herein are outside the scope of this specification. Such measures as encryption of files (either at rest or in transit), authentication of users or originators, assignment and control of file access permissions, and backup or recovery of files which may be necessary to meet institutional policies or governmental regulations are not addressed in this specification. Guidance for security of dictated health records can be found in Guide E 1902.

2. Referenced Documents

2.1 *ASTM Standards:*

E 1762 Guide for Electronic Authentication of Health Care Information[2]

E 1902 Guide for Management for the Confidentiality and Security of Dictation, Transcription, and Transcribed Health Records[2]

E 1985 Guide for User Authentication and Authorization[2]

E 2084 Specification for Authentication of Healthcare Information Using Digital Signatures[2]

2.2 *ANSI Standards:*

X3.172-1990 Dictionary of Information Processing[3]

X3.4-1986 Coded Character Sets—American National Standard Code for Information Exchange (7 bit ASCII)[4]

2.3 *ISO Standard:*

ISO 8072-1986 Network Standards[3]

2.4 *Other Standards:*

Health Level Seven Standard (HL7), Version 2.3.1[5]

Resource Interchange File Format (RIFF) Standard[6]

ITU G7.11[7]

3. Terminology

3.1 *Definitions:*

3.1.1 *degradation of sound quality*—in speech storage and reproduction, loss of intelligibility, reduced signal quality, and reduced ability to identify the speaker. The sound quality is a major factor in transcription, where the user must listen to speech for extended periods. The listener may have a reduced ability to identify the speaker due to the degradation of sound quality.

3.1.2 *destination system*—digital dictation system that receives dictation messages from another system.

---

[1] This specification is under the jurisdiction of ASTM Committee E31 on Healthcare Informatics and is the direct responsibility of Subcommittee E31.22 on Health Information Transcription and Documentation.

Current edition approved Dec. 10, 2001. Published March 2002.

[2] *Annual Book of ASTM Standards*, Vol 14.01.

[3] Available from International Organization for Standards (ISO), 1 rue de Varembe, Case Postale 56, CH-1211, Geneve 20, Switzerland.

[4] Available from American National Standards Institute (ANSI), 25 W. 43rd St., 4th floor, New York NY 10036.

[5] Available from Health Level Seven (HL7), 3300 Washtenaw Ave., Suite 227, Ann Arbor, MI 48104-4261.

[6] A multimedia standard published as a joint design document by IBM and Microsoft, available from Microsoft Corporation, One Microsoft Way, Redmond, WA 98052.

[7] Available from International Telecommunications Union (ITU) (formerly CCITT), Palais des Nations, CH-1211, Geneva 10, Switzerland.

3.1.3 *dictate workstation (or dictate station)*—device for input of voice dictation and data that is captured and subsequently stored.

3.1.4 *digital dictation recorder*—device designed specifically to accept and store digital dictation from multiple discrete sources. The recorder can usually be accessed by a variety of means including, but not limited to, touch-tone telephones, direct-wire input stations, and PC-based dictation applications.

3.1.5 *digital dictation server*—device designed as a repository for digital dictation. This server may function as the management point for workflow and routing of the dictation files.

3.1.6 *digital dictation system*—system that receives, stores, manages, and transmits digital sound files and their associated data. It may be shared between a number of remote dictate and transcribe stations. The system may contain various combinations of dictate workstations, dictation recorders, dictation servers, transcribe workstations, and other related components.

3.1.7 *digital voice file audio element*—digitized audio portion of a dictation message, excluding the data elements.

3.1.8 *digital voice file data elements*—information associated with a digital voice file, excluding the audio component. The data elements contain information that describes the voice file or its contents, or both.

3.1.9 *digital voice format*—technical description of the digital representation of the voice.

3.1.10 *interoperability*—capability of a functional unit to operate normally in different data processing environments in a way that requires users to have little or no knowledge of the unique characteristics of those units. (See ANSI X3.172-1990, page 60.)

3.1.11 *source system*—digital dictation system that sends dictation messages to another system.

3.1.12 *transcribe workstation (or transcribe station)*—device utilized to access stored data or voice dictation, or both, in order to convert it into text or integrate it, or both, with existing text.

3.1.13 *voice file compression*—sound is an analog wave that can be converted into a digital form for computer storage or nondegradable transmission. The digital voice file is generated by sampling the sound wave at predetermined intervals, and then representing each sample with bits or bytes of data. When the voice needs to be heard, a digital-to-analog conversion is performed in order to reconstitute the actual sound of the voice. The resulting sound quality will be related to: *(1)* the accuracy of the sampling algorithm, *(2)* the amount of data originally collected (the number of samples per unit of time times the number of bits per sample), and *(3)* the impact of any data compression. It will also be affected by the hardware. Compression can be "lossless" or "lossy." A lossless compression reduces the amount of data required to represent the original digital voice file but has absolutely no impact on sound quality. The original file can be replicated, precisely, at any time. Lossy compression actually loses some information, resulting in degradation of the sound quality inherent in the original voice file and an inability to precisely regenerate that original file. However, "compression ratios" are typically greater with this type of compression. Suitable lossy compression algorithms are able to reduce data storage and transmission requirements while maintaining a sufficient voice quality for the sound file's intended use.

3.1.14 *voice file transfer/transmission*—movement of a discrete amount of digitally recorded voice information over a transfer media.

3.1.14.1 *Discussion*—The voice information is in digital form and includes a header of descriptive textual data about the voice information. The amount of voice information depends on the length in time of the voice episode and the method of encoding the voice. Transfer media types include telecommunication media (like phone lines), local area networks, discrete storage media, and wireless means. The data interchange transmission may be between computer systems in a given medical institution, between medical institutions, or between a healthcare institution and a business partner, such as a transcription service.

3.2 *Definitions of Terms Specific to This Standard:*

3.2.1 *dictation message (or digital voice file)*—unit of information that consists of both audio and data elements. The audio (voice) component generally represents one entire dictated job, and the associated data elements are applicable to that particular job. In the context of this specification, a dictation message to be transferred from one system to another is generally contained within a single file.

3.2.2 *document*—voice file with its associated data elements (see *dictation message*).

3.2.3 *intelligibility*—how well the speech is understood by human subjects. It is measured by having human subjects listen to reproduced speech and answer questions to determine how well the speech is understood. Speech may be intelligible, but may be unpleasant to hear and have background noises.

3.3 *Acronyms:*

ADPCM	Adaptive Delta Pulse Code Modulation
ASCII	American Standard Code for Information Interchange
CPR	Computer-based Patient Record
MIME	Multipurpose Internet Mail Extensions
MS-DOS	Microsoft Disk Operating System
PCM	Pulse Code Modulation
RIFF	Resource Interchange File Format

4. Significance and Use

4.1 *General Approach*—This specification defines a general and flexible mechanism to enable the interoperability of digital dictation systems and workstations, including the transfer of digital voice files and their identifying data between disparate systems, regardless of manufacturer. The specification identifies required and suggested data elements associated with voice files utilized in a healthcare environment. The required elements shall be provided by any industry standard digital dictation system. The elements in the suggested category may accompany a voice file using a digital dictation system coupled with another system (like the CPR system). This specification addresses not only those data elements that are required when the job is created but also information that a destination system might need to provide back to the source system for status updates. The combined data elements and voice data object are referred to as the dictation message.

4.2 *Level of Implementation*:

TABLE 1 Supported Wave Voice File Formats

Format Type	<wformatTag> (hexadecimal)	Channels	Bits/ Sample	Sample Rate (/second)
PCM	0x0001	mono (1)	8 & 16	8 000, 11 025, 22 050
ITU / CCITT A Law	0x0006	mono (1)	8	8 000, 11 025, 22 050
ITU / CCITT mu Law	0x0007	mono (1)	8	8 000, 11 025, 22 050
OKI ADPCM	0x0010	mono (1)	4	6 000, 8 000
IMA (DVI) ADPCM	0x0011	mono (1)	3 & 4	8 000, 11 025, 22 050
TrueSpeech	0x0022	mono (1)	1.067	8 000
GSM 610	0x0031	mono (1)	1.6	8 000

4.2.1 This specification is specifically targeted for the definition of a message encapsulating both the data elements and actual voice file encoded in a standard compression algorithm. While given consideration, the transport mechanism is outside the scope of this specification.

4.2.2 An example suitable for the transmission of large amounts of digital voice data would be the use of industry-standard magnetic tape or digital audio tape, with ANSI standard-type labels, using variable length blocked records (lines) with a maximum block size of 4092 bytes. Individual lines within each block could be terminated by a carriage return character (ASCII code 13). As another example, for the transmission of moderate amounts of digital voice data, floppy disks written in MS-DOS[8] format or another commonly used directory and file structure could be used; the data would be contained within a single sequential file on the disk, with lines within the file delimited by carriage return (ASCII 13) or carriage return followed by line feed (ASCII 10) characters. An example of network hardware and software suitable for the transmission of digital voice data would be Ethernet[9] and the TCP/IP[10] protocol. The files could be sent from system to system using the HL7 data exchange specification.[5]

4.3 *Voice File Formats*—This specification provides the voice file compression algorithms that are commonly available in the industry. This specification considers standard voice files to be those voice file formats that are widely used and cover the current media types that are anticipated to be used in the healthcare industry in the foreseeable future. The compression algorithms referenced are widely used and supported by common operating environments. It is expected that the algorithms referenced may need to be converted to another format before being utilized by proprietary systems.

5. Description of the Digital Voice File Formats

5.1 The digital voice formats defined by this specification were selected to meet the following criteria to ensure interoperability in different data processing environments: *(1)* the voice file format must be self-descriptive, *(2)* conversion must not require proprietary hardware, and *(3)* the format must be commonly available in the industry.

5.2 The digital voice file format is Resource Interchange File Format (RIFF). RIFF is a tagged-file specification used to define formats for multimedia files. Tagged-file structure helps prevent compatibility problems that often occur when file-format definitions change over time. Because each piece of data in the file is identified by a header, an application that does not recognize a given data element can skip over the unknown information. A WAV file uses the RIFF format and has a .WAV filename extension. RIFF format voice files are self-descriptive; that is, the voice file format is defined within the file. This standard supports the WAV voice file formats in Table 1.

5.3 *Voice File Format Descriptions*:

5.3.1 In PCM (pulse code modulation), the digitized data consists of a series of numeric values at some regular rate (the stated sampling rate). Each sampled value is either 8 or 16 bits, with linear quantization. Sampling rates of 8000, 11 025, and 22 050 per second are supported. This format is supported on most standard platform computer operating systems.

5.3.2 ITU/CCITT mu Law and A Law are versions of PCM with a piece-wise linear approximation of logarithmic encoding. Mu law is a standard for North American telephony. A Law is used in Europe. See ITU G.711 for details. Sampling rates of 8000, 11 025, and 22 050 per second are supported. This format is supported on most standard platform computer operating systems.

5.3.3 OKI ADPCM is a version of Adaptive Delta PCM, which encodes the difference between samples with a method in which the same number of bits per sample can sometimes represent very small changes in the input signal and at other times represent much larger changes. This encoding reduces the voice file size for a given sample rate. The encoder/decoder specification is defined in Dialogic Application Note AN009. Sampling rates of 6000 and 8000 per second are supported.

5.3.4 IMA (DVI) ADPCM is a version of Adaptive Delta PCM, which encodes the difference between samples with a method in which the same number of bits per sample can sometimes represent very small changes in the input signal and at other times represent much larger changes. This encoding reduces the voice file size for a given sample rate. The encoder/decoder specification is defined in the RIFF standard. Sampling rates of 8000, 11 025, and 22 050 per second are supported. This format is supported on many standard platform computer operating systems, including Microsoft Windows 95 and Windows NT.

5.3.5 TrueSpeech is DSP Group's TrueSpeech 8.5 speech compression algorithm. A sampling rate of 8000 per second is

[8] Microsoft Disk Operating System, Microsoft Corporation.

[9] Ethernet LAN is a local area network system developed by Digital Equipment Corporation, Intel Corporation, and Xerox Corporation.

[10] Comer, D., "Transmission Control Protocol/Internet Protocol (TCP/IP), "Internetworking with TCP/IP; Principles, Protocols, and Architecture, Prentice-Hall, Englewood Cliffs, NJ 07632.

supported. This format is supported on many standard platform computer operating systems, including Microsoft Windows 95 and Windows NT.

5.3.6 GSM 610 is a speech compression algorithm. This format is supported on many standard platform computer operating systems, including Microsoft Windows 95 and Windows NT.

5.3.7 There are other registered RIFF WAV formats that may not meet this specification's requirements as stated in 6.1 and, therefore, they are not included in this specification.

6. Proposed Structure for Voice File Transfer

6.1 *Basic RIFF Audio File Format*—The WAV file format is a subset of Microsoft's RIFF specification for the storage of multimedia files. A RIFF file starts out with a file header followed by a sequence of data chunks. A WAV file is often just a RIFF file with a single WAV chunk which consists of two sub-chunks: *(1)* a format sub-chunk specifying the data format and *(2)* a data sub-chunk containing the actual audio data. This format is described in Table 2.

6.2 Along with the format and data sub-chunks, additional sections may be added to the WAV file to store related information. It is proposed that a fourth section be added to the basic WAV file to include the digital voice file data elements. The layout of this data element sub-chunk is described in Table 3.

6.3 Fields identified as *Required* must be supplied for each digital dictation file sent from one system to another. Fields identified as *Optional* may be sent as necessary to satisfy the particular requirements of a given implementation. Fields identified as *Conditional* are typically dependent on other fields or conditions, as explained. Resolution of any issues arising from differences in the availability of or requirements for these optional and conditional fields between systems involved in the transfer of digital dictation files will be the responsibility of the parties concerned.

6.4 Fields within the data element sub-chunk have the following format:

FIELD NAME=FIELD VALUE|…

6.4.1 The FIELD NAME=FIELD VALUE entries are pipe delimited by default; however, the delimiter character is configurable. If no field value is available, the delimiter will immediately follow the '=' sign. The last entry in the list is not terminated with a delimiter.

6.4.2 An example of the data portion of a data element sub-chunk follows. A carriage return and line feed is included after each delimiter to increase readability here, but there would be no such characters in the transmitted file.

```
WorkType=HP|
ForeignOriginatorID=98765|
ForeignOriginatorName=Alexander|
ForeignDocumentID=1234|
ForeignCreationDate=20000216143201|
ForeignDifficultyFactor=0|
OriginatingInputDevice=|
OriginatingFacility=General Hospital|
OriginatingSystem=GH1|
State=WorkflowComplete|
PriorityDate=|
CreationDate=20000217174353|
ID=1387259|
ArchiveStatus=0|
Originator=Robert C Armstrong|
Editor=|
DictationDate=20000217174356|
LastModified=20000217174353|
CreateDocTime=20000217174401|
CloseDocTime=20000217174356|
ProviderID=888657|
ProviderName=Robert C Armstrong|
TranAssignTime=20000218192556|
TranBeginTime=20000218192623|
TranCompleteTime=20000218192706|
TranAssignID=999543|
TranAssignName=Kathy Sturgeon|
TranCompleteID=999543|
TranCompleteName=Kathy Sturgeon|
TranTurnAroundSecs=543|
AudioFileSecs=53|
MedRecNum=001234|
AcctNum=99774321|
PatLastName=Stone|
PatFirstName=Rebecca|
PatMiddleName=Anne
```

7. Keywords

7.1 destination system; dictation message; digital dictation; digital dictation recorder; digital dictation server; digital dictation system; digital voice; encryption; interoperability; security; source system; voice file; voice file compression; voice file transfer

TABLE 2 Standard WAV File

Field Name	Field Size (bytes)	Field Offset (bytes)	Basic Form of a RIFF File for the Audio
Chunk ID	4	0	The RIFF chunk will identify this as a WAV file, which requires two additional chunks—format and data.
Chunk Size	4	4	
Format	4	8	
Sub-chunk1 ID	4	12	The format sub-chunk contains the details about the makeup of the audio portion of the file. The audio is contained in the next WAV sub-chunk.
Sub-chunk1 Size	4	16	
Audio Format	2	20	
Number of Channels	2	22	
Sample Rate	4	24	
Byte Rate	4	28	
Block Align	2	32	
Bits per Sample	2	34	
Sub-chunk2 ID	4	36	The data sub-chunk specifies the length of the audio portion and contains the raw audio data.
Sub-chunk2 Size	4	40	
Data	Variable	44	

TABLE 3 Voice File Data Elements

Field Name	Required/Optional or Conditional	Field Size (bytes)	Description
Sub-chunk3ID	R	4	Data elements (literal)
Sub-chunk3Size	R	4	Size of Sub-chunk3 (not including the 8 bytes required for the ID and Size fields)
WorkType	R	Variable	Document work type on source system
ForeignOriginatorID	C[A]	Variable	If the document did not originate on the source system, this field will contain the Originator ID from the system on which the document originated
ForeignOriginatorName	C[A]	Variable	If the document did not originate on the source system, this field will contain the Originator Name from the system on which the document originated
ForeignDocumentID	C[A]	Variable	If the document did not originate on the source system, this field will contain the Document ID from the system on which the document originated
ForeignCreationDate	C[A]	Variable	If the document did not originate on the source system, this field will contain the date/time the document was created on the originating system (YYYYMMDDHHMMSS)
ForeignDifficultyFactor	O	Variable	If the document did not originate on the source system, this field will contain the assigned transcription difficulty factor from the system on which the document originated
OriginatingInputDevice	R	Variable	Identifier for the device from which the document was created on the source system (dictate station, PC, etc.)
OriginatingFacility	R	Variable	Identifier for the facility where the document was originally created
OriginatingSystem	R	Variable	Identifier for the digital dictation system on which the document was originally created
State	O	Variable	Current document workflow state on source system
PriorityDate	O	Variable	Date/time the document was set to priority status on the source system (YYYYMMDDHHMMSS)
CreationDate	C[B]	Variable	Date/time the document was created on the source system. If the document was initially created by specifying the data elements and then adding the audio portion later, this field indicates the date/time the data elements were specified. (YYYYMMDDHHMMSS)
ID	R	Variable	Document ID on the source system
ArchiveStatus	O	Variable	Indicates whether or not the document has been archived (making the voice file available for deletion)
Originator	R	Variable	ID of the person creating (dictating) the document
Editor	O	Variable	ID of person who modifies the document
DictationDate	R	Variable	Date/time the audio portion of the document was dictated on the source system. (YYYYMMDDHHMMSS)
LastModified	C[C]	Variable	Date/time of the most recent modification of the document. (YYYYMMDDHHMMSS)
CreateDocTime	O	Variable	Date/time the document was initially opened (when dictation began) (YYYYMMDDHHMMSS)
CloseDocTime	O	Variable	Date/time the document was closed (when dictation ended) (YYYYMMDDHHMMSS)
ProviderID	O	Variable	ID of the person providing the service for which the document was dictated (may be the same as the Originator)
ProviderName	O	Variable	Name of the person providing the service for which the document was dictated
TransAssignTime	C[D]	Variable	Date/time the document was assigned to a transcriptionist (YYYYMMDDHHMMSS)
TransBeginTime	C[E]	Variable	Date/time the transcriptionist opens the document (begins playing the audio) (YYYYMMDDHHMMSS)
TransCompleteTime	C[F]	Variable	Date/time the transcriptionist marks the document to signify that it has been transcribed (YYYYMMDDHHMMSS)
TransAssignID	C[D]	Variable	ID of the person assigned to perform transcription of the document
TransAssignName	C[D]	Variable	Name of the person assigned to perform transcription of the document
TransCompleteID	C[F]	Variable	ID of the person who actually completes transcription of the document
TransCompleteName	C[F]	Variable	Name of the person who actually completes transcription of the document
TransTurnAroundSecs	C[F]	Variable	Difference between TransBeginTime and TransCompleteTime (seconds)
AudioFileSecs	R	Variable	Length of the voice file (seconds)
MedRecNum	R	Variable	Medical record number of the person who is the subject of the document
AcctNum	O	Variable	Account number of the person who is the subject of the document
PatLastName	O	Variable	Patient's last (family) name
PatFirstName	O	Variable	Patient's first (given) name
PatMiddleName	O	Variable	Patient's middle name

[A] If the document originated on a foreign system, this field is required.
[B] If the document was created "all at once" by an originator entering data elements and then immediately dictating, this field is not required.
[C] For new documents, this field will contain the date/time the document was created.
[D] Valid only if the document has been assigned to a transcriptionist on the source system.
[E] Valid only if transcription of the document has begun on the source system.
[F] Valid only if transcription of the document has been completed on the source system.