ANSI/NIST-ITL 1-2011 SUPPLEMENT:

VOICE RECORD

Draft

22 August, 2012

Version A-1c

Contents

Introduction 3

Investigatory Voice Biometric Committee (IVBC) Membership 5

Definitions of Specialized Terms Used in this Document 5

Transactions Supported by a Type-11 Record 8

Scope of the Type-11 Record 13

Source Documents 14

Administrative Metadata Requirements 14

Speaker and Content Metadata Requirements 15

Audio Technology Metadata Requirements 16

Audit Logs 16

General Organization of the Type-11 Record 17

Record Type-11: Voice record 18

Summary

The idea of automated and semi-automated (human-assisted) speaker recognition for forensic, investigatory and related applications goes back to World War II. Considerable government and private sector monies have been spent over the intervening 70 years in developing technical approaches, speech databases and testing programs. Missing from these efforts, however, has been the development of a forensic voice recording interchange format comparable to the interchange formats that currently exist for fingerprint, palmprint, face, iris, scar/mark/tattoo, and DNA data used for the purpose of human recognition. The Investigatory Voice Biometrics Committee (IVBC) was created by the Federal Bureau of Investigation (FBI) in early 2011 to take on the task of initiating development of a voice recording format to allow the interchange of voice data conformant to the data structure specified in the standard “American National Standard for Information Systems, Data Format for the Interchange of Fingerprint, Facial & Other Biometric Information, ANSI/NIST-ITL 1-2011”[[1]](#footnote-2), That standard is the current *de facto* international standard for exchange of biometric data for law enforcement and national security applications. This supplement to that standard is based upon a report of those. A first draft of this supplement was presented by IVBC to the ANSI/NIST-ITL Voice Working Group (ANVWG) at its first meeting on March 9, 2012 at the National Institute of Standards and Technology (NIST). The ANVWG was open to participation from all interested parties. Further work on the draft of the Supplement was conducted by the ANVWG and the IVBC working jointly for presentation to the canvassees.

The Type-11 record is modeled roughly after existing record types in the 2011 version of the ANSI/NIST-ITL standard, ANSI/NIST-ITL 1-2011, which is referred to as ANSI/NIST-ITL in this supplement. It allows exchange of both digital and analog data using both electronic and physical media. This Type-11 record is designed to be used within ANSI/NIST-ITL formatted transactions for law enforcement and homeland security-type speaker recognition and other closely-related speech applications. It is not specifically designed for speaker recognition within logical or physical access control, “time-and-attendance”, point-of-sale, or other consumer applications. This standard does not specify which techniques will be used in any human-assisted, automated or mixed voice processing application and does not specify how findings of forensic voice comparisons will be quantified or recorded. Supplemental text is included in several sections of the standard, as detailed in this document.

# Introduction

Speaker recognition presents some unique challenges not found in other forms of human recognition, such as fingerprint, iris or face. The human voice, generally carrying both speech and non-speech sounds, propagates varying distances through air (principally) or another medium to reach acoustic transducers (usually microphones) of varying amplitude and phase response. For purposes of the Type-11 record, a “speaker” is any person producing “vocalizations” from the throat or oral cavity, which may be voiced (activating the vocal cords) or unvoiced (such as aspirations, whispers, tongue clicks and other similar sounds). The current state of technology for speaker recognition usually requires vocalizations containing some speech (linguistic content) . An automated interlocutor is considered to be a “speaker” in for purposes of this record type, since the intent is to directly mimic human speech, although such a speaker will not be the primary subject of a speaker recognition transaction.

When voice sounds carry speech, that speech usually occurs within a social context involving more than one speaker. Consequently, a speech signal collected *in situ* will generally contain the voices of multiple speakers, each voice signal with its own transfer function between the speaker and the transducer. Segmenting and de-conflicting overlapped voice signals (“speaker separation”) through automation is currently an unsolved problem in the general case, thus implying that many operational applications of speaker recognition technology will involve audio recordings containing multiple speakers and multiple acoustic transmission paths.

The ANSI/NIST-ITL standard was originally developed for the interchange of fingerprint data, whether collected from latent prints lifted from crime scenes, scanned off of ink-based fingerprint cards or taken directly from electronic “live” scanners. The standard, therefore, is explicitly restricted to cases where, “All records in a transaction shall pertain to a single subject”. This restriction presents special challenges for use of the standard for interchange of natural voice signals, containing both speech and non-speech sounds, collected in a social, multi-speaker context and stored either digitally or in analog form and either electronically or on physical media. Therefore, a voice record type will have to accommodate:

1) bespoke recordings of single speaker voice signals for the specific purpose of speaker recognition;

2) conversational and interview scenario voice signals, digitized and segmented into clips, or “snips”, restricted to speech from the single speaker of interest (the voice data subject);

3) unsegmented natural voice signals on digital or analog media, with or without an accompanying timing diary of the segments attributable to speech from the single speaker of interest;

4) unannotated speech segment(s) for input to annotation work-flow tools. In all cases, the voice samples referred to in the Type-11 record must accommodate signals collected non-continuously and stored in multiple segments, a requirement that has been encountered before in other ANSI/NIST record types. For example, the Type-14 (variable-resolution fingerprint images) record has the capacity to carry multiple fingerprint samples in one image with segment boundary information for each finger in the image, albeit from a single individual, and serves as a model in this regard.

There are other challenges facing a speaker recognition standard. The most significant ones include:

* Voice signals generally contain both speech and non-speech elements, either of which might be useful in speaker recognition applications.
* Unlike other modalities, voice signals are collected in time, not spatial, dimensions and will not have a single “time of collection”.
* In mobile applications, even a single segment of a voice signal may not be linkable to a single geographic location.
* Voice signals containing speech have direct informational content. Unlike other forms of biometric recognition, the speech itself means something and, even if stripped of all personally identifiable information including the acoustic content itself, may require protection for privacy or security reasons.
* Unlike other modalities, voice signals may reflect the social and behavioral conditions of the collection environment, including the relationship between the data subject and any interlocutors.

Consequently, creating a Type-11record for voice signal transmission with the ANSI/NIST-ITL context is more complicated than simply copying an existing ANSI/NIST record type and changing terminology ( for example, substituting “voice” for “fingerprint” and “signal” for “image”). In the case of DNA Type-18 records, the standard has previously shown significant flexibility in dealing with record types which carry non-spatial data with significant content beyond that required for the recognition of individuals.

# Investigatory Voice Biometric Committee (IVBC) Membership

Joseph Campbell, MIT

Carson Dayley, FBI

Craig Greenberg, NIST

Peter Higgins, Consultant

Alysha Jeans, FBI

Ryan Lewis, FBI

Jim Loudermilk, FBI

Kenneth Marr, FBI

Alvin Martin, Consultant

Hirotaka Nakasone, FBI

Mark Przybocki, NIST (IVBC Chair)

Vince Stanford, NIST

Pedro Torres-Carrasquillo, MIT

James Wayman, Consultant

Bradford Wing, NIST

**ANSI/NIST-ITL Voice Working Group (ANVWG)**

In addition to the above members of the IVBC, the following persons participated in the ANVWG:

Bonny Scheier, Saber

Walter Tewes, Forensic Odontology Partners

Martin Herman, NIST

 **TO BE ADDED TO AS NEEDED (CHECK SIGN-IN LIST)**

# Definitions of Specialized Terms Used in this Document

The following definitions are supplemental to Section 4 of ANSI/NIST-ITL 1-2011

Acoustic signal

Pressure waves in a media with information content.

Audio signal

Information in analog or digital form that contains acoustic content (voice or otherwise)

Audio recording

A stored audio signal capable of being transduced into an audible acoustic signal.

Note: By “audible” means “capable of being heard by humans”.

Contemporaneous

Existing at or occurring at the same period of time.

Note: In this record type, the phrase “contemporaneous capture of a voice signal” indicates recording of the voice signal at the time of the speaker vocalization.

Diary

List giving the start and stop times of speech segments of interest pertaining to the primary voice subject within the voice signal.

Note: Diarization of segments from multiple speakers requires multiple Type-11 records, one for each speaker. These multiple Type-11 records may be contained in a single transaction, as long as the transaction is focused upon a single subject.

Known Voice Signal

A voice signal from an individual who has been “identified”, or individuated in a way that allows linking to additional, available information about that individual.

Metadata

Documentation about the biometric sample necessary or helpful in supporting the types of transactions likely to be encountered in law enforcement and homeland security applications.

Physical medium

Any external storage material of the voice signal and content information in either analog or digital form. Examples include reel-to-reel recording tape, cassette tape, Compact Disc, and phonograph record.

Quality

An estimate of the usefulness of a biometric sample for the purpose of recognition.

Questioned Voice Signal

A voice signal from an individual who is unknown and has not yet been linked to any previously encountered individual. Note: The task of speaker identification is to link a questioned voice sample to a known voice sample through determination of a common speaker.

Record (n)

An ANSI/NIST-ITL biometric data format type, in its entirety, within an ANSI/NIST-ITL transaction.

Note 1:In this document, this will be the Type-11 record unless otherwise stated.

Note 2: An ANSI/NIST-ITL transaction might contain multiple Type-11 records, as well as other record types, including the mandatory Type-1 record.

Record (v)

The act of converting an acoustic voice signal directly from an individual into a storage media, perhaps through contemporaneous, intermediate (transient) signal types.

Note: This definition is retained because of its entrenchment in natural language use. Consequently, a record (n) is not recorded, it is created.

Note: Transcoding is the term used for further processing of the voice signal and any digital or analog representation of that signal.

Record creation

The act of creating a record contained in an ANSI/NIST-ITL transaction.

Recording (n)

A stored acoustic signal in either analog or digital form.

Redaction

Over-writing of segments of a voice signal for the purpose of masking speech content in a way that does not disrupt the time record of the original recording.

Snip (n)

A segment of a voice signal extracted from a larger voice signal recording.

Note: Also called a “clip” or a “cut” in some communities.

Snip (v)

Extraction of segments of a voice signal in a way that disrupts the continuity and time record of the original recording.

Speaker

A vocalizing human, whether or not the vocalizations contain speech.

 Note: An interlocutor might be a synthesized voice, which can be considered a “speaker” within the context of this supplement.

Speech

Audible vocalizations made with the intent of communicating information through linguistic content.

 Note 1: Nonsensical vocalizations with linguistic content will be considered as speech.

 Note 2: Speech can be made by humans, by machine synthesizers, or by other means.

Subject of the record

The person to whom the data in the record applies.

 Note: The subject of the record need not be the subject of the transaction, because a transaction can include Type-11 records for interlocutors and others not named as the subject of the transaction.

Subject of the transaction

The person to whom the transaction applies.

 Note: The subject of a record need not be the subject of the transaction.

Transaction

A transmission between sites or agencies comprised of records, types of which are defined in ANSI/NIST-ITL.

Note: An ANSI/NIST-ITL transaction is called a file in Traditional encoding and an Exchange Package in XML encoding.

Transcoding

Any transfer, compression, manipulation, re-formatting or re-storage of the original recorded material.

Note 1: Transcoding is not the first recording of the acoustic signal.

Note 2: Transcoding can be lossless or lossy.

Voice data file

The digital, encoded file primarily containing the sounds of vocalizations of both speech and non-speech content, convertible to an acoustic signal replicating the original acoustic signal.

Note 1: A voice data file is extracted from an audio recording, but not all audio recordings contain voice signals and not all voice data is speech.

Note 2: A physical medium, such as a cassette tape, contains a voice signal but is not a voice data file.

Voice recording

A signal, stored on a digital or analog medium, of vocalizations containing both speech and non-speech content.

Voice subject

The single speaker of interest in the Type-11 record.

Note 1: This may not be the subject of the transaction.

Note 2: The voice signal subject may be known or unknown.

# Transactions Supported by a Type-11 Record

This record type was designed to support specific applications encountered in a forensic or investigatory environment. These applications may be supported by the ANSI/NIST-ITL standard by utilizing a combination of record types, as illustrated below. The Type-11 record was designed to be flexible enough to handle both the data for submission to a laboratory or agency as well as the response of that laboratory or agency to the submitting organization.

The Type-1 record within an ANSI/NIST-ITL transaction contains Field 1.004 for specifying the “type of transaction” (**TOT**) – the purpose for which the transaction was generated

The **information designation character /** **IDC** value(s) in Field 1.003 are selected to be unique for each submitted Type-11 record and entered in Field 11.002 as such. Note that Field 1.010 **Transaction control reference / TCR** for each response transmission would contain the same value as was sent in Field 1.009 **Transaction control number / TCN** in the original request. In this manner, requests are linked to responses.

For submissions, the transactions may include:

\* Type-2 record: if required by the application profile

\* Type-20 record(s): optionally, one for each source from which the biometric sample(s) in a Type-11 record or records is derived.

\* Type-21 record(s): optionally, one for each associated record (such as a copy of a warrant authorizing the recording of a person)

\* Type-98 record: optionally, to provide an audit log and to provide information assurance that the data has not been altered.

Note: additional record types may be included concerning the subject of the transaction, such as a facial image (Type-10).

Potential scenarios using a Type-11 record:

1. Voice model creation and storage for a known speaker
	1. Request submission
		1. Type-11 record(s): one for each recording of speech of the known speaker
		2. Other record types as appropriate
	2. Response (optional)
		1. Type-1 record
			1. **TOT** field 1.004 specified appropriately for the specific application profile indicating that a voice model has been created.
		2. Type-2 record: if required by the application profile
2. Voice model creation and storage for an unknown speaker.

This would follow a similar pattern to scenario 1 (above), but with a different **TOT** selected in Record Type-1. Note that the subject of the transaction is unknown, but is a single person. The original recording may contain more than one person’s voice, but segmentation could be used to indicate the portion of the recording that contains the voice of the subject of the transaction.

1. Comparison of the speakers in two audio recordings.

This is structurally a unique case. If the submitting organization has the capability to segment the audio, a separate transaction shall be created for each of the speakers in each of the two audio recordings. The laboratory / receiving agency would then compare the separate transactions relating to the first recording to each separate transaction relating to the second recording. The response would indicate which segmented transaction from the first transaction matched which (if any) segmented transaction from the second recording.

If the submitting organization does not have the capability to segment the audio recording, a single transaction is created with two type-11 records -- one for each recording. The subject of the transaction is considered ‘Unknown.’ The receiving laboratory / agency performs the segmentation, and comparison of the audio signals. The laboratory / receiving agency would then prepare a separate transaction for each identified individual in response to the submitting organization.



**Figure V1. Canonical speaker comparison example, scenario number 3** **(comparison of the speakers in two audio recordings)**. **The question “?” (“Does the question voice recording share the same source as a known voice recording?”) is addressed (with appropriate metadata, context, analysis, and caveats) by a certified examiner in a report delivered externally.**

1. Comparison of the voice in an audio recording to the voice models from a list of known speakers.

This scenario is similar to the first scenario on the submission (with the appropriate **TOT**) but the response would include a Type-11 record indicating a candidate. Each potential candidate would be handled by sending a separate transaction from the laboratory / receiving agency to the submitting organization.

1. Converting an analog audio recording into digitized voice data file(s).

This is handled by expressing the process as a particular **TOT** in the Type-1 record. The subject of the transaction is considered to be ‘Unknown’ unless specified by the submitting organization.

1. Duplicating or transcoding an audio recording.

This is handled by expressing the process as a particular **TOT** in the Type-1 record. The subject of the transaction and of the Type-11 record is considered to be ‘Unknown’ unless specified by the submitting organization.

1. Finding and isolating voice signals in an audio recording.

This is handled by expressing the process as a particular **TOT** in the Type-1 record. The subject of the transaction and of the Type-11 is considered to be ‘Unknown’ unless specified by the submitting organization.

1. Finding and isolating speech signals within an audio recording.

This is handled by expressing the process as a particular **TOT** in the Type-1 record. The subject of the transaction and of the Type-11 record is considered to be ‘Unknown’ unless specified by the submitting organization.

1. Determination of the distinct speakers in an audio recording.

The submitting organization creates a transaction with the subject of the transaction as ‘Unknown’ and uses the appropriate TOT as defined by the application profile.

1. Indexing an audio recording into voice segments attributable to distinct speakers.

The submitting organization creates a transaction with the subject of the transaction as ‘Unknown’ and uses the appropriate **TOT** as defined by the application profile. The receiving laboratory / agency performs the segmentation. It is recommended that the laboratory / receiving agency create a Type-98 record with a log of the operations performed. The segmented recording would be returned to the submitting organization in a Type-11 record. The original recording may be also sent back in a Type-20 record linked to the Type-11 record.

1. Creation of a diary, attributing speech segments to a speaker of interest.

This operation is similar to the above scenario, with the addition of specific references to the appropriate segments in the Type-11 record in the response from the laboratory / receiving agency.

1. Creation of word or phonic level transcriptions, in the language spoken, of segments of speech attributable to a single speaker.

This operation is similar to scenario 10, with the addition of specific references to the appropriate segments in the Type-11 record from the response from the laboratory / receiving agency.

1. Redaction of an audio recording to remove sensitive speech segments.

This operation is similar to scenario 10, with the addition of specific references to the redactions in the Type-11 record in the response from the laboratory / receiving agency.

1. Snipping of an audio recording to remove segments of non-speech, speech not attributable to the subject of interest, or speech not of interest to the transaction.

This operation is similar to scenario 10, with the addition of specific references to the snippings in the Type-11 record in the response from the laboratory / receiving agency.

1. Enhancing the speech segments in an audio recording for return to the submitting agency for use in human-assisted or automated speaker recognition applications.

This operation is similar to scenario 10, with the addition of specific references to the enhancements in the Type-11 record in the response from the laboratory / receiving agency.

1. Authentication of an audio recording as containing the continuous speech of a single speaker without deletions or insertions.

In this case, the submitting organization would create a Type-11 record, and use the appropriate **TOT** in the Type-1 record. The response from the receiving laboratory / agency would be a Type-1 record with the appropriate **TOT** indicating the response (Yes or No), and at least one other record.

1. Transfer voice recording to an archive for permanent storage

The submitting organization would generate the appropriate Type-1 record, and any other necessary records. The Type-11 record would reference or include the voice recording along with the associated metadata. No response is required from the receiving laboratory / agency.

The above list is not meant to be exhaustive of potential uses of the Type-11 record, but to illustrate its relationship with other record types. As noted, the **TOT** field of the Type-1 record plays an extremely important role in the use of the Type-11 record(s).

This voice signal record type supports all of these transactions originating from submitting agencies with little or no capability in digitizing audio signals or in speech analysis, as well as inter- and intra-laboratory transmissions on fully or partially processed voice recordings. Further, the type of transactions ultimately to be performed on the voice recording might not be fully known at the time the Type-11 record is created. Therefore, the voice recordings referred to in the Type-11 record must be accompanied by documentation, when available, to support a very wide variety of potential transactions. This documentation (“metadata”) is of four basic types:

* Administrative metadata: who initiated the transaction, for what purpose, and with what authority?
* Speaker metadata: what is known about the speaker of interest and their physical and psychological condition at the time of the speech?
* Content metadata: what language is being spoken, when was the original content spoken under what conditions, and what content information is available that might help in the speaker recognition process?
* Audio technology metadata: how was the voice signal collected, stored and processed and what technical parameters will help in the faithful reproduction and analysis of the signal within the storage medium?

Some of this metadata, such as the time and date of the original recording, might only be known from external sources. Some of the metadata, such as the language being spoken, might be discernible from the voice recording itself. Much of the metadata might not be known or available to the various agencies creating the audio recording, the Type-11 record and the ANSI/NIST-ITL transaction. All of the metadata, however, could be useful in the processing of the audio recording given the potential for widely varying transactions and, therefore, should be made readily available to the receiving agency without requiring the reprocessing of the audio recording. Consequently, the goal in creating this supplement was to create as many non-redundant metadata fields as possible to permit transmission of documentation of potential future interest, even if the metadata could potentially be recovered from the audio recording itself. Most of these fields are optional because much of the potentially relevant metadata may be unknown to the various agencies involved in the transaction.

# Scope of the Type-11 Record

The following updates Section 5.3.11 of ANSI/NIST-ITL 1-2011

 Type-11 records shall support the transmission of audio recordings containing speech by one or more speakers, including noise (data of no interest to the transaction, whether speech, non-speech voice data, or non-voice data) for forensic and investigatory purposes in the context of an ANSI/NIST- ITL transaction pertaining to a single, perhaps unknown, individual. These transmissions support transactions related to detecting and recognizing speakers, extracting from an audio recording speech segments attributable to a single speaker, and linking speech segments by speaker, whether these functions are to be accomplished through automated means (computers), human experts, or hybrid human-assisted systems. Related functions, such as redaction, authentication, phonetic transcription and enhancement, while also supported, are not the primary concern of this record type, although audio recordings supporting these related functions may be transmitted via Type-11 records. This standard does not specify which techniques will be used in any human expert, automated or hybrid voice processing application and does not specify the form of the examination report. Although not designed for use in logical or physical access control, “time-and-attendance”, “point-of-sale”, or other consumer or commercial applications, nothing in this record type should be construed as preventing its application in these or other transaction types not specifically addressed here. This record type does not support streaming transactions. This record does not define the transmission of features or models extracted from voice data, but does allow the user to define specific fields to contain such information, in acordance with an implementation domain or application profile. Fields that may be used for user-specific purposes are specified as such in this supplement. This record type does not restrict the media by which the audio recording will be transmitted, but will support digital transmission of transaction information regardless of the audio recording media.

# Source Documents

The following is added to Annex I of ANSI/NIST-ITL 1-2011

1. Collaborative Digitization Program, Digital Audio Working Group, “Digital Audio Best Practices”, version 2.1, October, 2006, http://ucblibraries.colorado.edu/systems/digitalinitiatives/docs/digital-audio-bp.pdf
2. Audio Engineering Society, “AES standard for audio metadata - Audio object structures for preservation and restoration”, AES57-2011, Sept. 21, 2011
3. Audio Engineering Society, “AES standard for audio metadata -Core audio metadata”, AES60-2011, Sept. 22, 2011

# Administrative Metadata Requirements

The following are usually requirements for administrative metadata in transactions containing audio recordings:

Requirement 1: Point-of-Contact (POC) Name

Requirement 2: Agency

Requirement 3: Phone number

Requirement 4: Originating agency case ID

Requirement 5: Transaction ID

Requirement 6: Embed Case ID

Requirement 7: Email address of submitter

Requirement 8: Alternative POC

Requirement 9: User defined fields, such as “Receiving agency case/other case”

The above information is normally required to promote traceability of the audio recordings. There are at least three levels of traceability – to the submitting, compiling/post-processing and collecting agencies. It is possible for all three agencies to be the same in some transactions, but they will often be different.

# Speaker and Content Metadata Requirements

The relevant metadata requirements for metadata about the data subject and the subject’s speech are listed below. The distinction between “long-term” and “short-term” attributes might be elusive in many cases.

1. Identifier
2. Long-term Attributes
	1. gender
	2. accent[[2]](#footnote-3)
	3. date of birth
	4. native language/language biography
	5. educational level
	6. primary location where data subject grew up
	7. speech pathology (may be intermittent)
3. Short-term Attributes
	1. Impairment/intoxication
	2. Language being spoken
	3. Language proficiency Health status
	4. Intelligibility
	5. Style (public speech, conversation, read, prompted, interview, other)
	6. Emotional state/vocal effort
	7. Citizenship

Note that much of the above information may be unknown to the originator of a transaction.

# Audio Technology Metadata Requirements

The relevant metadata for the audio technology is:

1. Overall/Preliminary signal quality
2. Duration of signal measured in seconds
3. Duration of signal measured in samples
4. Encoding/container format
5. Sampling rate
6. Bit depth (may be encoding dependent)
7. Recording method (conversion of temporary to permanent storage)
8. Time/date of recording
9. Where recorded
10. Type of recorder
11. Make/model/serial number of recorder
12. Transducer characteristics
13. Transducer type: array, earbud, wire, microphone, handset, speaker phone,…
14. Channel information

Note that much of the above information may be unknown to the originator of a transaction.

# Audit Logs

The Record Type-98, “Information assurance record”, allows special data protection procedures to ensure the integrity of the transmitted data and allows for the maintenance of an audit log. **Field 98.900 (Audit log / ALF)** may be used to indicate how and why a transaction was modified. The **ALF** is of particular use when a transaction is sent from one location to a second, where additional information is included, before sending the transaction to a final destination for processing. In the case of a voice recording, the **ALF** will be used to indicate how and why redaction, snipping and diarization information was created or edited. See ANSI/ NIST ITL 1-2011**, Section 8.22** and the “[Type 98 Best Practice Implementation Guidance](http://biometrics.nist.gov/cs_links/standard/Type_98_Best_Practice_Guidance_v1.3.pdf).”[[3]](#footnote-4)

An example might be that a local police lab sends a transaction with multiple Type-11 records, containing voice signals of both known and unknown persons, to the appropriate central area command center, where additional information could be added. The command center might also redact case-sensitive speech from the voice recordings referenced in the Type-11 records before sending the transaction to another forensic unit for additional redaction. The forensic unit may then forward the updated transaction to the national Forensic Audio, Video and Image Analysis Unit. which may create diaries of the questioned voice samples in the Type-11 records, indicating which segments were from the speaker of interest, as recorded in a known voice sample in additional Type-11 records. The diaries might be revised after additional supervisory review. All of this would be documented in a Type-98 record included in the evolving transaction.

In contrast to the Type-98 record, which presents an audit log at the level of the entire transaction, **Field 11.902** provides an audit log at the level of the Type-11 record. **Field 11.902** lists the operations, such as redaction, snipping or diarization, performed on the original voice recording in order to prepare it for inclusion in the record type. See **Section 7.4.1** of the ANSI/NIST-ITL standard**.**

# General Organization of the Type-11 Record

The Type-11 record is organized into 6 parts: I) mandatory fields; II) initial global fields, applying to the entire voice data record; III) indication of presence and definition of segments within the voice data record; IV) fields applying to the individual segments; V) additional global fields modeled on other Types in the ANSI/NIST standard; VI) fields containing or pointing to the voice recording.

1. Mandatory fields:

01 Record header

02 Information designation character

1. The initial global fields are:

03 Audio object descriptor (internal or external digital file, external physical media containing digital/analog/unknown recording)

04 Voice laboratory setting (source of the voice recording, phone numbers and POCs)

05 Role of voice recording (known sample, unknown single speaker, unknown multiple speakers)

06 Recorder (hardware/software)

07 Type-11 record creation date

08 Voice recording creation date

09 Total recording duration

10 Physical media object (tape, CD, phonograph record,...)

11 Container Format (wav, ogg, mp3/4)

12 Codec (PCM types)

13 Preliminary signal quality (multiple quality metrics possible)

14-20 Fields reserved for future use

1. The presence and definition of segments within the audio file follow.

21 Redaction (yes/no, by whom?)

22 Redaction diary (where and why redaction occurred)

23 Snipping (yes/no, by whom?)

24 Snipping diary (separate snips/clips/cuts are numbered and identified by relative start/end times, comments)

25 Diarization (yes/no, by whom?)

26 Segment diary (segments are numbered with relative start/end times, labels of attributes attributed to the speech and speaker of each segment, and comments.)

27-30 Reserved for future use

1. Repeating sets of sub-fields labeled by segment numbers as designated in the diarization. (If the segment number is "0", that becomes the default for all segments not otherwise listed.)

31 Date/time of recording of segment/snip and labeled date/time of recording

32 Geolocation of data subject of this Type-11 record at start of segment/snip

33 Segment/snip quality values (possible multiple values for each segment)

34 Vocal collision indicator (two or more persons speaking at once)

35 Processing priority of the segment/snip

36 Segment content (language, prompted/read/conversation, word transcript, phonetic transcript, translations)

37 Segment/snip speaker characteristics (impairment, intelligibility, health, emotion, vocal effort, vocal style, language proficiency)

38 Segment channel (transducer, capture environment, channel type)

39-50 Fields reserved for future use

1. More global fields modeled on other record types in ANSI/NIST ITL 2011:

51 Global comments

52 – 901 Fields reserved for future use

902 Annotation information

903 Device Unique Identifier

904 Make/Model/Serial

905-992 Fields reserved for future use

1. Source Agency Name
2. The voice recording or pointers to that recording:

994 External file reference

995 Associated context reference (Type 21 record)

996 Voice data file hash

997 Source representation reference (Type 20 record with original audio)

998 Field reserved for future use

999 Voice data file

The following is a replacement for Section 8.11 of ANSI/NIST-ITL 1-2011

# Record Type-11: Voice record

The Type-11 record shall be used to exchange a single voice data file or a physical medium containing a digital or analog voice recording, together with fixed and user-defined textual information fields (referred to in this standard as “metadata”) pertinent for understanding and processing the voice signal.

A voice signal is defined in this standard as any audible vocalizations emanating from the human mouth or throat with or without speech content. The Type-11 record references a recording of a voice signal stored as a digital voice data file within the record, or a recording external to the transaction.. Information regarding the recording type, the voice data file size, and other parameters or comments required to process the voice data file are given as fields within the Type-11 record. If the Type-11 record references a voice recording contained in a physical medium (i.e., an analog tape, a digital tape, a CD, a phonograph record), the label and location of that medium shall be indicated in this Type-11 record, along with the information necessary to render the stored recording as acoustic output.

A transmitted voice recording may be processed by the recipient agencies to isolate the voice signal of interest and to extract the desired feature or model information required for voice comparison, speaker detection, or speech attribution purposes.

A single ANSI/NIST-ITL transaction might contain multiple voice recordings, each as a separate Type-11 record within the transaction. Although the transaction pertains to a single person, the individual voice recordings in each of the Type-11 records required for the transaction may contain the speech of multiple speakers.

If there are multiple speakers of interest in a voice recording supported by a Type-11 record, then a separate ANSI/NIST-ITL transaction may be created for each individual of interest, each transaction possibly containing the same Type-11 records. If the voice recording included in or pointed to by a Type-11 record has been extracted from a longer source recording, that source recording may be included in digital form within the transaction as a Type-20 record, or referred to as an external source in either digital or analog format in the Type-20 record. Voice models or features extracted from voice data are not explicitly accommodated in this record, but may be transmitted in user-defined fields.

**Table V-1 Type-11 record layout**

Key for Character type: N=Numeric; A=Alphabetic; AN=Alphanumeric; B=Binary or Base64; U=Unicode

Key for Cond. code: M=Mandatory; O=Optional; D = Dependent upon another value or condition described in the text; M↑=Mandatory if the field/subfield is used; O↑=Optional if the field/subfield is used.

| **Field****Number** | **Mnemonic** | **Content Description** | **Cond**  **code** | **Character** | **Value Constraints** | **Occurrence** |
| --- | --- | --- | --- | --- | --- | --- |
| Type | Min# | Max# | Min**#** | Max**#** |
| 11.001 |  | **RECORD HEADER**  | M | encoding specific: see **Annex B: Traditional encoding** or **Annex C: NIEM-conformant encoding rules** | encoding specific: see **Annex B: Traditional encoding** or **Annex C: NIEM-conformant encoding rules** | 1 | 1 |
| 11.002 | **IDC** | **INFORMATION DESIGNATION CHARACTER** | M | N | 1 | 2 | 0 ≤ IDC ≤ 99integer | 1 | 1 |
| 11.003 | **AOD** | **AUDIO OBJECT DESCRIPTOR** | M | N | 1 | 1 | See Table V-20 ≤ AOD ≤ 4 | 1 | 1 |
| 11.004 | **VRI** | **VOICE RECORDING INFORMATION**  | O |   | 0 | 1 |
| SRT | source recorder type | O↑ | A | 1 | 1 | LTY = G, I, P, O or U | 0 | 1 |
| NOO | name of original source | O↑ | U | 1 | 400 | none | 0 | 1 |
| POC | point of contact | O↑ | U | 1 | 200 | none | 0 | 1 |
| CSC | code of sending country | O↑ | AN | 1 | 3 | value from*ISO-3166-1* | 0 | 1 |
| 11.005 | **ROL** | **ROLE OF VOICE RECORDING** | M | N | 1 | 2 | See Table V-30 ≤ ROL ≤ 99 | 1 | 1 |
| 11.006 | **REC** | **RECORDER** | M |   | 1 | 1 |
| RTP | recorder type | O | U | 1 | 4000 | none | 0 | 1 |
| MAK | recorder make  | O | U | 1 | 50 | None | 0 | 1 |
| MOD | recorder model | O | U | 1 | 50 | None | 0 | 1 |
| SER | recorder serial number | O | U | 1 | 50 | None | 0 | 1 |
| AQS | acquisition source | M | AN | 1 | 2 | value from **Table 83 except for 1 through 6 inclusive or 11;** **or AQS = MS** | 1 | 1 |
| COM | comment  | O | U | 1 | 4000 | None | 0 | 1 |
| 11.007  | **RCD** | **RECORD CREATION DATE** | M | See Section 7.7.2.4 Local date and time; encoding specific: see Annex B: Traditional encoding or Annex C: NIEM-conformant encoding rules | See Section 7.7.2.4 Local date and time; encoding specific: see Annex B: Traditional encoding or Annex C: NIEM-conformant encoding rules  | 1 | 1 |
| 11.008 | **VRD** | **VOICE RECORDING CREATION DATE** | O | See Section 7.7.2.4 Local date and time; encoding specific: see Annex B: Traditional encoding or Annex C: NIEM-conformant encoding rules | See Section 7.7.2.4 Local date and time; encoding specific: see Annex B: Traditional encoding or Annex C: NIEM-conformant encoding rules  | 0 | 1 |
| 11.009 | **TRD** | **TOTAL RECORDING DURATION** | O |  | 0 | 1 |
| TIM | total time | O↑ | N | 1 | 11 | 1 ≤ TIM ≤ 99999999999 (in microseconds)(no commas) | 0 | 1 |
| CBY | compressed bytes | O↑ | N | 1 | 14 | 1 ≤ CBY ≤ 99999999999999(no commas) | 0 | 1 |
| TSM | total samples | O↑ | N | 1 | 11 | 1 ≤ SMP ≤ 99999999999999(no commas) | 0 | 1 |
| 11.010 | **PMO** | **PHYSICAL MEDIA OBJECT** | D |  | 0 | 1 |
| MTP | media type | M↑ | U | 1 | 300 | None | 1 | 1 |
| RSP | recording speed | O↑ | NS | 1 | 9 | 0.9999999 ≤ RSP ≤ 999999999value may include a decimal point or be an integer (no commas) | 0 | 1 |
| RSU | recording speed units | D↑  | A | 1 | 300 | None | 0 | 1 |
| EQ | equalization | O↑ | AN | 1 | 100 | None | 0 | 1 |
| TRK | tracks | O↑ | N | 1 | 2 | 1 ≤ TRK ≤ 99 | 0 | 1 |
| STK | speaker track | O↑ | NS | 1 | 200 | values between 1 and 99 inclusive that are separated by commas | 0 | 99 |
| COM | comments | O↑ | U | 1 | 4000 | None | 0 | 1 |
| 11.011 | **CFT** | **CONTAINER FORMAT** | O | N | 1 | 2 | See Table V-4 | 0 | 1 |
| 11.012 | **CDC** | **CODEC** | D |  | 0 | 1 |
| CDT | codec type | M↑ | N | 1 | 3 | see **See Table V-5** | 1 | 1 |
| SRT | sampling rate | O↑ | NS  | 1 | 5 | 0 ≤ SRT < 100,000  (kHz) value may include a decimal point or be an integer0 = variable or unknown | 0 | 1 |
|  | BIT | bit depth | O↑ | N | 1 | 2 | 0 ≤ BIT ≤60positive integer0 = variable or unknown | 0 | 1 |
| EDN | endian | O↑ | N | 1 | 1 | 0=big; 1=little; 2=native | 0 | 1 |
| PNT | fixed point | O↑ | N | 1 | 1 | 0=floating point1=fixed point | 0 | 1 |
| NCH | number of channels | O↑ | N | 1 | 2 | 1 ≤ NCH ≤ 99 | 0 | 1 |
| COM | comment | D | U | 1 | 4000 | None | 0 | 1 |
| 11.013 | **PSQ** | **PRELIMINARY SIGNAL QUALITY** | O |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** |  | 1 | 9 |
| QVU | quality value | M↑ | N | 1 | 3 | 0 ≤ QVU ≤ 100 or 255= quality not assessed; Integer | 1 | 1 |
| QAV | algorithm vendor identification | M↑  | H | 4 | 4 | 0x00 ≤ QAV ≤ FFFF | 1 | 1 |
| QAP | algoorithm product identification | M↑ | N | 1 | 5 | 0 ≤ QAP ≤ 65534 positive integer | 1 | 1 |
| COM | comments | D | U | 1 | 300 | None | 0 | 1 |
| 11.014--11.020 |  | **RESERVED FOR FUTURE USE only by ANSI/NIST-ITL** |  |
| 11.021 | **RED** | **REDACTION** | O |  | 0 | 1 |
| RDI | redaction indicator | M↑ | B | 1 | 1 | 0=no1=yes | 1 | 1 |
| RDA | redaction authority | O↑ | U | 1 | 300 | None | 0 | 1 |
| COM | comment | O↑ | U | 1 | 4000 | None | 0 | 1 |
| 11.022 | **RDD** | **REDACTION DIARY** | O |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** | M↑ | 1 |  600,000 |
| RID | redaction identifier | M↑ | N | 1 | 6 | 1 ≤ RID ≤ 600000 | 1 | 1 |
| TRK | tracks  | D↑ | NS | 1 | 297 | List of integers separated by commas | 0 | 1 |
| RST | relative start time | M↑ | N | 1 | 11 | 1≤ RST ≤ 99999999998  | 1 | 1 |
| RET | relative end time | M↑ | N | 1 | 11 | 99999999999 ≥ RET > RST  | 1 | 1 |
| COM | comment | O↑ | U | 1 | 4000 | None | 0 | 1 |
| 11.023 | **SNP** | **SNIPPING SEGMENTA-TION** | O |  | 0 | 1 |
| SGI | snipping indicator | M↑ | B | 1 | 1 | 0=no1=yes | 1 | 1 |
| SPA | snipping authority | O↑ | U | 1 | 300 | None | 0 | 1 |
| COM | comment | O↑ | U  | 1 | 4000 | None | 0 | 1 |
| 11.024 | **SPD** | **SNIPPING DIARY** | O |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** |  | 1 | 600000 |
| SPI | snip identifier | M↑ | N | 1 | 6 | 1 ≤ SPI ≤ 600000 | 1 | 1 |
| TRK | tracks  | D↑ | NS | 1 | 297 | List of integers separated by commas | 0 | 1 |
| RST | relative start time | M↑ | N | 1 | 11 | 99999999998≥RST ≥ 0  | 1 | 1 |
| RET | relative end time | M↑ | N | 1 | 11 | 99999999999> RET > RST  | 1 | 1 |
| COM | comment | O↑ | U | 1 | 4000 | None | 1 | 1 |
| 11.025 | **DIA** | **DIARIZATION** | D |  | 0 | 1 |
| DII | diarization indicator | M↑ | B | 1 | 1 | 0=no1=yes | 1 | 1 |
| DAU | diarization authority | O↑ | U | 1 | 300 | None | 0 | 1 |
| COM | comment | O↑ | U  | 1 | 4000 | None | 0 | 1 |
| 11.026 | **SGD** | **SEGMENT DIARY**  | D |  | 0 | 1 |
|  | ***subfields: repeating sets of information items*** | M↑ |  | 1 | 600,000 |
| SID | segment identifier | M↑ | N | 1 | 6 | 1 ≤ SID ≤600,000 | 1 | 1 |
| TRK | tracks | D↑ | NS | 1 | 297 | List of integers separated by commas | 0 | 1 |
| RST | relative start time | M↑ | N | 1 | 11 | 99999999998 ≥ RST ≥ 0 | 1 | 1 |
| RET | relative end time | M↑ | N | 1 | 11 | 99999999999 ≥ RET > RST  | 1 | 1 |
| COM | comment | O↑ | U | 1 | 10000 | None | 0 | 1 |
| 11.027 –11.030 |  | **RESERVED FOR FUTURE USE only by ANSI/NIST-ITL** |  |
| 11.031 | **TME** | **TIME OF SEGMENT RECORDING**  | D |  | 0 | 1 |
|  |  | ***Subfield: repeating sets of information items*** | M↑ |  | 1 | \* |
| DIA | diary identifier | M↑ | B | 1 | 1 | 0=snip diary1=segment diary | 1 | 1 |
| SID | segment identifier | M↑ | N | 1 | 5 | 1 ≤ SID ≤ 600,000 | 1 | 1 |
| DOR | date of original recording | O↑ | encoding specific: see **Annex B** or **Annex C** | encoding specific: see **Annex B** or **Annex C** | 0 | 1 |
| TDT | tagged date  | O↑ | encoding specific: see **Annex B** or **Annex C** | encoding specific: see **Annex B** or **Annex C** | 0 | 1 |
| SRT | start time of segment recording | O↑ | encoding specific: see **Annex B** or **Annex C** | encoding specific: see **Annex B** or **Annex C** | 0 | 1 |
| TST | tagged start time | O↑ | encoding specific: see **Annex B** or **Annex C** | encoding specific: see **Annex B** or **Annex C** | 0 | 1 |
| END | end time of segment recording | O↑ | encoding specific: see **Annex B** or **Annex C** | encoding specific: see **Annex B** or **Annex C** | 0 | 1 |
| TET | tagged end time | O↑ | encoding specific: see **Annex B** or **Annex C** | encoding specific: see **Annex B** or **Annex C** | 0 | 0 |
| STM | source of time | O↑ | U | 1 | 300 | none | 0 | 1 |
| COM | comment | O↑ | U | 1 | 4000 | none | 0 | 1 |
| 11.032 | **GEO** | **SEGMENT GEOGRAPHIC-AL INFORMATION** (about person of interest at start of segment) | D |  | 0 | 1 |
|  |  | ***Subfields: repeating sets of information items*** | M↑ |  | 1 | \* |
| DIA | diary identifier | M↑ | B | 1 | 1 | 0=snip diary1=segment diary | 1 | 1 |
| SID | segment identifiers | M↑ | NS | 1 | \* | 0 or a list of integers separated by commas | 1 | 1 |
| SCT | segment cell phone tower code | O↑ | U | 1 | 100 | none | 0 | 1 |
| LTD | latitude degree value | D | NS  | 1 | 9 | -90 ≤ LTD ≤ 90 | 0 | 1 |
| LTM | latitude minute value | D | NS  | 1 | 8 | 0 ≤ LTM < 60 | 0 | 1 |
| LTS | latitude second value | D | NS  | 1 | 8 | 0 ≤ LTS < 60 | 0 | 1 |
| LGD | longitude degree value | D | NS  | 1 | 10 | -180 ≤ LGD ≤ 180 | 0 | 1 |
| LGM | longitude minute value | D | NS  | 1 | 8 | 0 ≤ LGM < 60 | 0 | 1 |
| LGS | longitude second value | D | N  | 1 | 2 | 0 ≤ LGS < 60positive integer | 0 | 1 |
| ELE | Elevation | O↑ | NS  | 1 | 8 | -422.000 < ELE < 8848.000real number | 0 | 1 |
| GDC | geodetic datum code | O↑ | AN  | 3 | 6 | value fromTable 6 | 0 | 1 |
| GCM | geographic coordinate universal transverse mercator zone | D | AN  | 2 | 3 | one or two integers followed by a single letter | 0 | 1 |
| GCE | geographic coordinate universal transverse mercator easting | D | N  | 1 | 6 | integer | 0 | 1 |
| GCN | geographic coordinate universal transverse mercator northing | D | N  | 1 | 8 | integer | 0 | 1 |
| GRT | geographic reference text | O↑ | U  | 1 | 150 | none | 0 | 1 |
| OSI | geographic coordinate other system identifier ( or landmark)  | O↑ | U  | 1 | 10 | none | 0 | 1 |
| OCV | geographic coordinate other system value | D | U  | 1 | 126 | none | 0 | 1 |
| 11.033 | **SQV** | **SEGMENT QUALITY VALUES** | D |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** | M↑ | 1 | \* |
| DIA | diary identifier | M↑ | B | 1 | 1 | 0=snip diary1=segment diary | 1 | 1 |
| SID | segment identifiers | M↑ | NS | 1 | \* | 0 or a list of integers separated by commas | 1 | 1  |
| QVU | quality value | M↑ | N  | 1 | 3 | 0 ≤ QVU ≤ 100or 255 = quality not assessed;Integer | 1 | 1 |
| QAV | algorithm vendor identification | M ↑ | H | 4 | 4 | 0x00 ≤ QAV ≤ FFFF | 1 | 1 |
| QAP | algorithm product identification | M ↑ | N | 1 | 5 | 0 ≤ QAP ≤ 65534 positive integer | 1 | 1 |
| COM | comment | D | U | 1 | 300 | none  | 1 | 1 |
| 11.034 | **VCI** | **VOCAL COLLISION INDICATOR** | D |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** |  | 1 | 2 |
| DIA | diary identifier | M↑ | B | 1 | 1 | 0=snip diary1=segment diary | 1 | 1 |
| SID | segment identifiers | M↑ | NS | 1 | \* | 0 or a list of integers separated by commas | 1 | 1 |
| 11.035 | **PPY** | **PROCESSING PRIORITY** | D |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** |  | 1 | \* |
| DIA | diary identifier | M↑ | B | 1 | 1 | 0=snip diary1=segment diary | 1 | 1 |
| SID | segment identifiers | M↑ | N | 1 | \* | 0 or a list of integers separated by commas | 1 | 1 |
| PTY | Priority | ↑ | N | 1 | 1 | 1 ≤ PTY ≤ 9 | 1 | 1 |
| 11.036 | **SCN** | **SEGMENT CONTENT** | D |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** | M↑ | 0 |  TBD |
| DIA | diary identifier | M↑ | B | 1 | 1 | 0=snip diary1= segment diary | 1 | 1 |
| SID | segment identifiers | M↑ | N | 1 | \* | 0 or a list of integers separated by commas | 0 | 1 |
| TRN | Transcript | O↑ | U | 1 | 100,000 | none | 0 | 1 |
| TRA | transcript authority | O↑ | U | 1 | 10,000 | none | 0 | 1 |
| 11.037 | **SCC** | **SEGMENT SPEAKER CHARACTERISTIC** | D |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** | M↑ | 1 |  TBD |
| DIA | diary identifier | M↑ | B | 1 | 1 | 0=snip diary1=segment diary | 1 | 1 |
| SID | segment identifiers | M↑ | NS | 1 | \* | 0 or a list of integers separated by commas | 0 | 1 |
| IMP | impairment | O↑ | N | 1 | 1 | 0 ≤ IMP ≤ 5 | 0 | 1 |
| LBS | language being spoken | O↑ | A | 3 | 3 | Value from *ISO 639-3* | 0 | 1 |
| LPF | language proficiency | O↑ | N | 1 | 1 | 0 ≤ LPF ≤ 9 | 0 | 1 |
| STY | style of speech | O↑ | N | 1 | 2 | See **Table V-6** | 0 | 1 |
| INT | intelligibility  | O↑ | N | 0 | 1 | 0 ≤ INT ≤ 9 | 0 | 1 |
| ITM | intimacy | O↑ | N | 0 | 1 | 0 ≤ ITM ≤ 5 | 0 | 1 |
| HST | health status | O↑ | U | 0 | 4000 | None | 0 | 1 |
| EM | emotional state | O↑ | N | 1 | 2 | See **Table V-7** | 0 | 1 |
| VEF | vocal effort  | O↑ | N | 1 | 1 | 0 ≤ VEF ≤ 5 | 0 | 1 |
| VSY | vocal style | O↑ | N | 1 | 2 | See **Table V-8** | 0 | 1 |
| AWR | awareness of recording process | O↑ | N | 1 | 1 | 0=unknown1=aware2=unaware | 0 | 1 |
| SCR | script | O↑ | U | 0 | 9999 | None | 0 | 1 |
| COM | comment | O↑ | U | 1 | 4000 | None | 0 | 1 |
| 11.038 | **SCH** | **SEGMENT CHANNEL** | D |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** | M↑ | 1 | TBD |
| DIA | diary identifier | M↑ | B | 1 | 1 | 0=snip diary1=segment diary | 1 | 1 |
| SID | segment identifiers | M↑ | N | 1 | \* | 0 or a list of integers separated by commas | 0 | 1 |
| TYP | transducer type | O↑ | N | 1 | 2 | See **Table V-9** | 0 | 1 |
| TRN | transducer | O↑ | N | 1 | 1 | unknown=0carbon=1electret=2other=3 | 0 | 1 |
| ENV | capture environment | O↑ | AN | 1 | 4000 | Text | 0 | 1 |
| DST | distance to transducer | O↑ | N | 1 | 5 | 0 ≤ DST ≤ 99999Integer | 0 | 1 |
| ACS | acquisition source | O↑ | N | 1 | 2 | See **Table 83** | 0 | 1 |
| ALT | alteration | O↑ | U | 1 | 400 | None | 0 | 1 |
| COM | comment | O↑ | U | 1 | 4000 | None | 0 | 1 |
| 11.039-11.050 |  | **RESRRESERVED FOR FUTURE USE only by ANSI/NIST-ITL** |  |
| 11.051 | COM | **COMMENT** | O↑ | U | 1 | 4000 | none | 0 | 1 |
| 11.052-11.099 |  | **RESERVED FOR FUTURE USE only by ANSI/NIST-ITL** | Not to be used |
| 11.100-11.900 | **UDF** | **USER-DEFINED FIELDS** | O | user-defined | user-defined | user-defined |
| 11.901 |  | **RESERVED FOR FUTURE USE only by ANSI/NIST-ITL** | Not to be used |
| 11.902 | **ANN** | **ANNOTATION INFORMATION** | O |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** | M↑ | 1 | \* |
| GMT | Greenwich mean time | M↑ | encoding specific: see **Annex B** or **Annex C** | encoding specific: see **Annex B** or **Annex C** | 1 | 1 |
| NAV | processing algorithm name version | M↑ | U  | 1 | 64 | none | 1 | 1 |
| OWN | algorithm owner | M↑ | U  | 1 | 64 | none | 1 | 1 |
| PRO | process description  | M↑ | U  | 1 | 255 | none | 1 | 1 |
| 11.903 | **DUI** | **DEVICE UNIQUE IDENTIFIER** | O | ANS | 13 | 16 | first character = M or P | 0 | 1 |
| 11.904 | **MMS** | **MAKE/MODEL/SERIAL NUMBER** | O |  | 0 | 1 |
| MAK | make | M↑ | U  | 1 | 50 | none | 1 | 1 |
| MOD | model | M↑ | U  | 1 | 50 | none | 1 | 1 |
| SER | serial number | M↑ | U  | 1 | 50 | none | 1 | 1 |
| COM | comment | O↑ | U | 1 | \* | none |  |  |
| 11.905-11.992 |  | **RESERVED FOR FUTURE USE only by ANSI/NIST-ITL** | Not to be used |
| 11.993 | **SAN** | **SOURCE AGENCY NAME** | O | U | 1 | 125 | none | 0 | 1 |
| 11.994 | **EFR** | **EXTERNAL FILE REFERENCE** | D | U  | 1 | 200 | none | 0 | 1 |
| 11.995 | **ASC** | **ASSOCIATED CONTEXT** | O |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** | M ↑ | 1 | 255 |
| CAN | associated context number | M↑ | N | 1 | 3 | 1 ≤ ACN ≤ 255positive integer | 1 | 1 |
| ASP | associated segment position | O↑ | N | 1 | 2 | 1 ≤ ASP ≤ 99positive integer | 0 | 1 |
| 11.996 | **HAS** | **HASH** | O | H | 64 | 64 | none | 0 | 1 |
| 11.997 | **SOR** | **SOURCE REPRESENTA-TION**  | O |  | 0 | 1 |
|  | ***Subfields: Repeating sets of information items*** | M↑ | 1 | 255 |
| SRN | source representation number | M↑ | N | 1 | 3 | 1 ≤ SRN ≤ 255positive integer | 1 | 1 |
| RSP | reference segment position | O↑ | N  | 1 | 2 | 1 ≤ RSP ≤ 99positive integer | 0 | 1 |
| 11.998 |  | RESERVED FOR FUTURE USE only by ANSI/NIST-ITL | Not to be used |
| 11.999 | **DATA** | **VOICE DATA** | D | B  | 1 | 22 | None | 0 | 1 |

1. **Field 11.001: Record header**

The content of this mandatory field is dependent upon the encoding used. See the relevant annex of this standard for details. See **Section 7.1**.

1. **Field 11.002: Information Designation Character / IDC**

This mandatory field shall contain the **IDC** assigned to this Type-11 record as listed in the information item **IDC** for this record in **Field 1.003 Transaction content / CNT.** See **Section 7.3.1.**

1. **Field 11.003: Audio Object Descriptor/AOD**

This mandatory field shall be a numeric entry selected from the attribute code column of Table V2. Only one value is allowed and indicates the type of audio object containing the voice recording which is the focus of this Type-11 record. Attribute code 0 indicates that the audio object of this record is a digital voice data file in the **Field 11.999**. Attribute code 1 indicates that the audio object is a digital voice data file at the location specified in **Field 11.994**. Attribute codes 2-4 indicate that the audio object is a physical media object at a location described in **Field 11.994**.

If the Type-11 record contains only metadata (such as in a response to a sample submission), attribute code 5 shall be selected.

**Table V-2**

**Audio Object Descriptor**

|  |  |
| --- | --- |
| **Audio Object** | **Attribute Code** |
| Internal digital voice data file | 0 |
| External digital voice data file | 1 |
| Physical Media Object containing digital data | 2 |
| Physical Media Object containing analog signals | 3 |
| Physical Media Object containing unknown data or signals | 4 |
|  No audio object associated with this record | 5 |

1. **Field 11.004: Voice Recording Information/VRI**

This is an optional field and shall contain information about the site or agency that created the voice recording pointed to or included in this record. In the case of files created from previous recordings, this is not necessarily the source of the original transduction of the acoustic vocalizations from the person to whom the Type-11 record pertains. This need not be the same as the **Source agency / SRC** or the Originating agency of **Field 1.008** or the Destination agency of **Field 1.007.** The first information item, the **source recorder type / SRT,** is optional. There may be no more than one occurrence of this item. When present, this information item contains a single character describing the site or agency that created the voice recording:

U = Unknown

P = Private individual

I = Industry / Commercial

G = Government

O = Other

* The second information item (**name of original source/ NOO**) is optional and shall be the name of the group, organization or agency that created the voice recording. There may be no more than one occurrence for this item. This is an optional information item in Unicode characters and is limited to 400 characters in length.
* The third information item is the **point of** c**ontact / POC** who composed the voice recording. This is an optional information item that could include the name, telephone number and e-mail address of the person or persons responsible for the creation of the voice recording. This information item may be up to 200 Unicode characters.
* The fourth information item is optional. It is the *ISO-3166-1* **code of the sending country / CSC**. This is the code of where the voice recording was created – not necessarily the nation of the agency entered in **Field 11.993: Source agency / SRC** . All three formats specified in *ISO-3166-1* are allowed (Alpha2, Alpha3 and Numeric). A country code is either 2 or 3 characters long.
1. **Field 11.005: Role of Voice Recording/ROL**

This is a mandatory field and shall be a numeric entry selected from the “attribute code” column of **Table V-3**. Only one value is allowed and indicates the role of the voice recording (known or questioned) within the transaction.

**Table V-3**

**Role of the Voice Recording**

|  |  |
| --- | --- |
| **Role** | **Attribute Code** |
| No information | 0 |
| Known sample, single speaker, subject of transaction | 1 |
| Known sample, single speaker, interlocutor | 2 |
| Known sample, single speaker, other | 3 |
| Known sample, multiple speakers, including subject of transaction | 4 |
| Known sample, multiple speakers, excluding subject of transaction | 5 |
| Questioned sample, single speaker | 6 |
| Questioned sample, multiple speakers | 7 |
| Audio recording with unknown voice content | 8 |
| Other | 9 |
| User-defined | 10-99 |

1. **Field 11.006: Recorder / REC**

This field is mandatory and shall indicate information about the recording equipment that created the voice recording contained in or pointed to by this record. There may be no more than one occurrence of this field.

NOTE: As recordings or data files may be transcoded from previously recorded or broadcast content, this equipment may or may not be the equipment used to record the original acoustic vocalization of the person to whom Type-11 record pertains.

* The first information item (**recorder type/RTP)** is an optional text field of up to 4000 characters describing the recording equipment that created the voice recording. An example would be “Home telephone answering device”.

* The second, third and fourth information items (**recorder make/MAK, recorder model/MOD, recorder serial number/SER)** are optional items of up to 50 characters each and shall contain the make, model and serial number, respectively, for the recording device. There may be no more than one entry for this item. See Section 7.7.1.2 for details.
* The fifth information item (**acquisition source/AQS**) is mandatory and is an alpha-numeric item. If all of the audio signal in the voice recording comes from a single acquisition source, the item shall be a numeric entry selected from the “attribute code” column of **Table 83** of the Type-20 record. When multiple sources are used for various voice segments in the voice recording, the code “MS” shall be used and individual sources will be given in the following comment item. If “12” from **Table 83** is chosen indicating an analog recording, then **Field 11.003** will indicate “3”, the recording will be described in **Field 11.010**, and the location of the physical medium will be recorded in **Field 11.994**. Note that codes 1 through 6 and 11 from **Table 83** are inapplicable, and shall not be used as a value in this information item.
* The sixth information item (**comments/COM**) is an optional text string of a maximum length of 4000 characters that may contain any additional information about the recorder used to create the voice recording, including information about the recording software. If **AQS** indicates multiple sources, “MS”, this field should be used to summarize the known sources from which the voice recording was created.

**Table 83[[4]](#footnote-5)**

**Acquisition Source**

|  |  |
| --- | --- |
| **Acquisition source type** | **Attribute code** |
| Unspecified or unknown | 0 |
| Static digital image from an unknown source | 1 |
| Static digital image from a digital still-image camera | 2 |
| Static digital image from a scanner | 3 |
| Single video frame from an unknown source | 4 |
| Single video frame from an analog video camera | 5 |
| Single video frame from a digital video camera | 6 |
| Video sequence from an unknown source | 7 |
| Video sequence from an analog video camera, stored in analog format | 8 |
| Video sequence from an analog video camera, stored in digital format | 9 |
| Video sequence frame from a digital video camera | 10 |
| Computer screen image capture  | 11 |
| Analog audio recording device; stored in analog form (such as a phonograph record) | 12 |
| Analog audio recording device; converted to digital | 13 |
| Digital audio recording device | 14 |
| Landline telephone – both sender and receiver | 15 |
| Mobile telephone – both sender and receiver | 16 |
| Satellite telephone – both sender and receiver | 17 |
| Telephone – unknown or mixed sources | 18 |
| Television – NSTC | 19 |
| Television – PAL | 20 |
| Television – Other | 21 |
| Voice-over-internet protocol (VOIP) | 22 |
| Radio transmission: short-wave (specify single side band or continuous wave in FDN) | 23 |
| Radio transmission: amateur radio (specify lower side band or continuous wave in FDN) | 24 |
| Radio transmission: FM (87.5 MHz to 108 MHz) | 25 |
| Radio transmission: long-wave (150 kHz to 519 kHz) | 26 |
| Radio transmission: AM (570 kHz to 1720 kHz) | 27 |
| Radio transmission: Aircraft frequencies | 28 |
| Radio transmission: Ship and coastal station frequencies | 29 |
| Vendor specific capture format | 30 |
| Other | 31 |

1. **Field 11.007: Record Creation Date/RCD**

This mandatory field shall contain the date and time of creation of this Type-11 record. This date will generally be different from the voice recording creation date and may be different from the date at which the acoustic vocalization originally occurred. See **Section 7.7.2.4** **Local date and time** for details.

1. **Field 11.008: Voice Recording Creation Date/VRD**

This optional field shall contain the date and time of creation of the voice recording contained in the record. If pre-recorded or transcoded materials were used, this date may be different from the date at which the acoustic vocalization originally occurred. See **Section 7.7.2.4** **Local date and time** for details.

1. **Field 11.009: Total Recording Duration / TRD**

This field is optional and gives the total length of the voice recording in time, compressed bytes and total samples. At least one of the three information items must be entered if this field is used.

* The first information item (**time/TIM**) is optional and gives the total time of the voice recording in microseconds. The size of this item is limited to 11 digits, limiting the total time duration of the signal to 99,999 seconds, which is approximately 27 hours.
* The second information item (**compressed bytes/CBY**) is optional and gives the total number of compressed bytes in the voice data file. Consequently, this information item applies only to digital voice recordings stored as voice data files. The size of this item is limited to 14 digits, limiting the total size of the voice data file to 99 terabytes.
* The third information item (**total samples/TSM**) is optional and gives the total number of samples in the voice data file after any decompression of the compressed signal. This information item applies only to digital voice recordings stored as voice data files. The size of this item is limited to 11 digits.
1. **Field 11.010: Physical Media Object/ PMO**

This field is optional and identifies the characteristics of the physical media containing the voice recording. There can be only one physical media object per Type-11 record, but multiple Type-11 records can point to the same physical media object. This field only applies if Field 11.003 has an attribute code of 2, 3 or 4. The location of the physical media object is given in Field 11.994.

* The first information item (**media type/MTP**) is mandatory if this field is used and contains text of up to 300 characters describing the general type of media (i.e., analog cassette tape, reel-to-reel tape, CD, DVD, phonograph record) upon which the voice recording is stored. If an analog media is used for storage, and **AQS** of Field 11.006 is 14, then a description of the digital to analog procedure should be noted in **Field 11.902** and the reasons for such a conversion noted in **COM** of Field 11.010.
* The second information item (r**ecording speed/RSP**) is optional and gives a numerical value to the speed at which the physical media object must be played to reproduce the voice signal content. This value may be integer or floating point and shall not exceed 9 characters.
* The third information item (r**ecording speed units/RSU**) is mandatory if the second information item, **RSP**, is entered and contains text of up to 300 characters to indicate the units of measure to which **RSP** refers.
* The fourth information item (e**qualization/EQ**) is optional and indicates the equalization that should be applied for faithful rendering of the voice recording on the physical media object.
* The fifth information item (t**racks/TRK**) is an optional integer between 1 and 99, inclusive, that gives the number of tracks on the physical media object. For example, a stereo phonograph record will have 2 tracks.
* The sixth information item (s**peaker track/STK**) is an optional list of integers which indicate which tracks carry the voices of the speaker(s).
* The seventh information item **(comment/COM) i**s optional and allows for additional comments of up to 4000 Unicode characters in length describing the physical media object.
1. **Field 11.011: Container Format/CFT**

This is an optional field (**container format/CFT**) that gives information about the container format, if any, which encapsulates the audio data of the electronic file used to carry the voice data in the digital recording. This field is not used if the voice recording is stored on a physical media object as an analog signal. If present, this field overrides the **CDC** Field 11.012. This field does not accommodate multiple Container Formats in a single Type-11 record. The Container Format shall be entered as the appropriate integer code from the Table below.

Typically these are files with headers describing the data and its encoding. Container files contain the audio samples and the audio specifications to properly decode the audio (or video), such as the codec; codec parameters; number of channels; sample rate; bit/byte depth; big, little, or native endian (which byte goes first) that are typically stored in the form of a header. More generally, the container formats may specify a codec or may encapsulate one or more audio channels as Linear PCM. In at least one case, Apple uses a pseudo-codec to indicate the endian sense.

A popular container format today is Microsoft’s Waveform Audio File Format (WAVE or WAV), which is a Microsoft Resource Interchange File Format (RIFF) method for storing data in chunks and is given the Windows filename extension *.wav*. The well-known Wave container specification has fields such as chunk ID, chunk size, audio format (codec), sampling rate, number of channels, space for extra parameters (for the codec or other uses). The Audio Format field within Wave can be harmonized with the Type-11 codec nomenclature defined below. Table V-4 is a list of canonical container formats, and other widely recognized container formats to be transcoded to canonical containers before inclusion in a Type-11 Record. Type-11 supports the following Container Formats. These are the most common container formats in the law-enforcement community. Rare container formats can be handled via conversion to a supported format, such as RIFF (.wav) outside the Type-11. For example, free software utilities such as MPlayer, SoX, or SUPER© can be used to convert to common audio formats.

**Table V-4**

**External Table of Audio Visual Container Types**

|  |  |  |
| --- | --- | --- |
| **Container Type** | **Extension** | **Attribute Code** |
| WAV (RIFF audio)\* | .wav | 1 |
| AVI (RIFF video) |  | 2 |
| WebM (Vorbis)\* |  | 3 |
| WebM (VP8 video) |  | 4 |
| AIFF\* | .aiff .aif | 5 |
| Vorbis (OGG audio)\* | .ogg | 6 |
| Theora (OGG video) |  | 7 |
| MPEG program stream (PS) |  | 8 |
| MPEG2 transport stream (TS) |  | 9 |
| MP4 (H.264/MPEG-4 AVC) |  | 10 |
| MKV Matroska container | .mkv | 11 |
| MXF Material eXchange Format (SMPTE std.) |  | 12 |
| ISO base media file format (3GP, MP4, ISO IEC) |  | 13 |
| ASF (MS container for wma, and wmv) |  | 14 |
| DVR-MS (MS container based upon ASF) |  | 15 |
| RMVB RealNetworks |  | 16 |
| RM (Realmedia) |  | 17 |
| QuickTime (Apple VBR-audio/video/image) | .mov .qt | 18 |
| FLV (Flash video) | .flv | 19 |
| F4V (Flash video) |  | 20 |
| Video for Windows | .avi | 21 |
| Windows Media | .wmv .wma .asf .asx | 22 |
| MPEG-1 | .mpg .mpeg .mpe | 23 |
| MPEG-2 | .vob | 24 |
| MPEG-3 | .mp3 | 25 |
| MPEG-4 | .mp4 | 26 |
| 3GP and 3G2 mobile video | .3gp .3g2 | 27 |

 \* Canonical Container File Format

Container formats evolve and new formats will be considered in future updates to this Type-11 record. Hazardous container formats, such as Flash Video (.flv) that encapsulates scripting language or code, are risky and should be avoided. Recommendation: reformat hazardous container formats to a canonical container format, such as RIFF, for transmission via a Type-11.

Raw, or headerless, files have only the audio samples and a file name to go on, in the absence of other information. All the audio characteristics required to properly interpret those samples must be gleaned elsewhere, hence, the need for the table of codecs, SRT, BIT, NCH, COM (Field 11.011), etc.

1. **Field 11.012: Codec/CDC**

This is an optional field that gives information about the codec used to encode the voice data in the digital recording. This field is not used if the voice recording is stored on a physical media object as an analog signal. This field is only used if no header is read for the digital audio file when it is opened. Information in Field 11.011 (**Container Type/CFT**) overrides this Field if both are present.

* The first information item (**Codec type/CDT)** is mandatory if this information item is used and indicates the single codec type used for all audio segments in the record. This format does not accommodate multiple codec types within a single record. It shall be a numeric entry selected from the “attribute code” column of **Table V-5**. If the codec type is identified as “other” -- a value of 4, the fifth information item (**comment/COM**) shall be used to describe the codec.

**Table V-5**

**Table of Codecs**

|  |  |
| --- | --- |
| **Codec Type** | **Attribute Code** |
| Linear PCM | 1 |
| Floating-point linear PCM | 2 |
| ITU-T G.711 (PCM): μ-law or A-law with reverse sample option | 3 |
| Other | 4 |

The second item (**sampling rate/SRT**) indicates the number of digital samples that represent a second of analog voice data upon conversion to an acoustic signal. The sampling rate is expressed in kHz and may contain a decimal point or may be an integer. Acceptable values are between 0 and 100 MHz (100,000 kHz), but unknown or variable sampling rates shall be given the value of 0. Common values of SRT are 8000, 11025, 16000, 22050, 32000, 44100, and 48000 Hz

* The third item (**bit depth/BIT**) indicates the number of bits that are used to represent a single sample of voice data. Acceptable values are between 1 and 60, inclusive. Encoders of unknown or variable bit depth shall be given the value of 0. Nothing in this field is meant to be an indication of the dynamic range of the voice data. Changes to bit depth should be logged in Type-98 or **Field 11.902** audit logs. Common values for **BIT** are 8, 16, 24, and 32 bits.
* The fourth item (**endian/EDN**) is optional and indicates which byte goes first. The values for **EDN** are 0=big, 1=little, or 2=native endian.
* The fifth item (**fixed point/PNT**) is optional and indicates the sample representation. The value is 0 if the samples are represented as floating-point or 1 if the samples are fixed-point.
* The sixth item (**number of channels/NCH**) is optional and gives the integer number of channels of data represented in the digital voice data file. The number of channels must be between 1 and 99, inclusive. If this item is not included, the voice data file will be assumed to have only one channel. Common values for **NCH** are 1 and 2 channels.
* The seventh item (**comment/COM**) is an optional unrestricted text string of up to 4000 characters in length that may contain additional information about the codec or additional instructions for reconstruction of audio output from the stored digital data. However, this information item shall be present if **CDT** = 4 (Other). This item would include description of any noise reduction processing or equalization that must be applied to faithfully render the voice recording. Codec parameters shall be specified in this field when required for unambiguous decoding.
1. **Field 11.013: Preliminary Signal Quality/PSQ**

This field is optional and gives an assessment of the general “quality” of the voice recording. There may be as many as 9 **PSQ** subfields for the audio file to indicate different types of quality assessments.

* The first information item (**quality value/QVU**) is mandatory if this field is used and shall indicate the general quality value between 0 (low quality) and 100 (high quality). A value of 255 indicates that quality was not assessed.
* A second information item is mandatory if this field is used and shall specify the ID of the vendor of the quality assessment algorithm used to calculate the quality score, which is an **algorithm vendor identification / QAV**. This 4-digit hex value (See **Section 5.5 Character types**)is assigned by IBIA and expressed as four characters. The IBIA maintains the Vendor Registry of CBEFF Biometric Organizations that map the value in this field to a registered organization. For algorithms not registered with the IBIA, the value of 0x00 shall be used.
* A third information item is mandatory if this field is used and shall specify a numeric product code assigned by the vendor of the quality assessment algorithm, which may be registered with the IBIA, but registration is not required. This is the **algorithm product identification / QAP** that indicates which of the vendor’s algorithms was used in the calculation of the quality score. This information item contains the integer product code and should be within the range 1 to 65,534. For products not registered with the IBIA, the code 0 shall be used.
* The fourth information item (**comment/COM**) is optional and should be used to give additional information about the quality assessment process. It shall be used to describe unregistered algorithms.
1. **Fields 11.014-020: Reserved Fields**

These fields are reserved for future use by ANSI/NIST-ITL.

1. **Field 11.021: Redaction/ RED**

This field is optional and indicates whether the voice recording has been redacted, meaning that some of the audio record has been overwritten (“Beeped”) or erased to delete speech content without altering the relative timings within, or the length of, the segments. This field is not to be used to indicate that audio content has been snipped with the alteration of the relative timings in, or length of, the segment.

* The first information item (**redaction indicator/RDI)** is a binary indicator and is mandatory if this field is used. It indicates whether the voice recording contains overwritten or erased sections intended to remove, without altering the length of the segment, semantic content deemed not suitable for transmission or storage. 0 indicates no redaction and 1 indicates that redaction has occurred.
* The second information item (**redaction authority/RDA)** is an optional text field of up to 300 characters in length containing information about the agency that directed, authorized or performed the redaction. Agencies undertaking redaction activities on the original speech should log their actions by appending to this item and noting the change of field contents in the Type-98 record and / or **Field 11.902** of this record
* The third information item (**comment/COM)** is an optional unrestricted text string of up to 4000 characters in length that may contain text information about the redactions affecting the stored voice data.
1. **Field 11.022: Redaction Diary/RDD**

This optional field (**redaction diary/RDD**) indicates the timings with the voice recording of redacted (overwritten) audio segments. The redactions need not be dominated by speech from the subject of this transaction or record. Four items (uniquely numbering the redactions identified by recording track and giving relative start and end times of each) are mandatory if this field is used and shall repeat for each redaction. A fifth item is optional and accommodates comments on the individual redactions. The record type accommodates up to 600,000 redactions by repeating the subfield.

* The first item (**redaction identifier/RID**) is mandatory if this field is used and uniquely numbers the redactions to which the following items in the field apply. There is no requirement that the redactions be numbered sequentially. The **RID** may contain up to 6 digits. The number of redactions is limited to 600,000.
* The second item (**tracks/TRK**) is mandatory if item **PMO\_TRK** in Field 11.010 or **CDC\_NCH** of Field 11.012 is greater than one and lists all tracks or channels on the recording to which the redaction identifier applies. The track numbers are separated by commas. No value in this list should be greater than the value of **PMO\_TRK** or **CDC\_NCH**, whichever applies. For example, in the case of a two-track stereo recording where both tracks contain a redaction at the same start and end times, this item will be “1,2”
* The third item (**relative start time/RST**) is a mandatory integer for every redaction identified by an **RID** and indicates in microseconds the time of the start of the redaction relative to the beginning of the voice recording. The item can contain up to 11 digits, meaning that the start of a redaction might occur anywhere within a voice recording limited to about 27 hours. It is not expected that redactions on the same track of the audio object will overlap, meaning that the **RST** of a redaction is not expected to occur between the **RST** and **RET** of any other redaction on the same track, although this is not prohibited. If the Type-11 record refers to an analog recording, the method of determining the start time shall be given in the comment item of this field.
* The fourth item **(relative end time/RET**) is a mandatory integer for every redaction identified by an **RID** and indicates in microseconds the time of the end of the redaction relative to the beginning of the voice recording. The item can contain up to 11 digits, meaning that the end of a redaction might occur anywhere within a voice recording limited to about 27 hours. As with the **RST**, it is not expected that redactions on the same track of the audio object will overlap, although this is not prohibited.

* The fifth item (**comment/COM**) is an optional unrestricted text string of up to 4000 characters in length that allows for comments of any type to be made on a redaction.
1. **Field 11.023: Snipping Segmentation/ SNP**

This field is optional and indicates whether the voice recording referenced in this Type-11 record has had segments removed or contains segments that have been snipped from one or more longer voice recordings, in either case meaning that the voice signal is not a continuous recording in time. This field is used to indicate removal, for any reason, of audio signal from the original recording of the acoustic vocalizations in a way that disrupts time references.

* The first information item **(snip indicator/SGI)** is a binary variable and is mandatory if this field is used. It indicates whether the voice recording contains temporal discontinuities caused by snipping of segments from one or more longer recordings. 0 indicates no snipping and 1 indicates that snipping has occurred.
* The second information item (**snipping authority/SPA)** is an optional text field of up to 300 characters containing information about the agency that performed the snipping segmentation. Agencies undertaking snipping activities on the original speech should log their actions by appending to this item and noting the change of field contents in the Type-98 record and / or **Field 11.902** of this record.
* The third information item (**comment/COM)** is an optional unrestricted text string of up to 4000 characters that may contain text information about the snip activities affecting the voice recording.
1. **Field 11.024: Snipping Diary/SPD**

This optional field (**snipping diary/SPD**) allows this type to document the snips obtained from larger voice recordings, which might themselves be included in the transaction as Type-20 records. There may be up to 600,000 snips diarized in repeating subfields. Each snip shall be dominated by speech from the subject of this Type-11 record. Four items (uniquely numbering the snips by track and giving relative start and end times of each) are mandatory in each subfield. A fifth item is optional within each subfield and allows for comments on the identified snip. If there is no snipping (**Field 11.023**) indicated, then all of the data in the voice recording will be considered as in toto and the subfields will not repeat. There can be at most one snipping diary for each Type-11 record.

* The first item (**snip identifier/SPI**) is mandatory in each subfield and uniquely numbers the snip to which the following items in the subfield apply. There is no requirement that the snips be numbered sequentially. The **SPI** may contain up to 6 digits and up to 600,000 snips may be identified. If **Field 11.023** indicates snipping, the voice recording must consist of at least one snip.
* The second item (**tracks/TRK**) is mandatory if item **PMO\_TRK** in Field 11.010 or **CDC\_NCH** of Field 11.012 is greater than one and lists all tracks or channels on the recording to which the snip identifier applies. The track numbers are separated by commas. No value in this list should be greater than the value of **PMO\_TRK** or **CDC\_NCH**, whichever applies. For example, in the case of a two-track stereo recording where both tracks contain a snip at the same start and end times, this item will be “1,2”
* The third item (**relative start time/RST**) is a mandatory integer for every snip identified by an **SPI** and indicates in microseconds the time of the start of the snip relative to the beginning of the voice recording. The item can contain up to 11 digits, meaning that the **RST** might occur anywhere within a voice recording limited to about 27 hours. Because each snip is obtained independently from a larger voice recording, snips from a single track on the audio object described in Field 11.003 shall not overlap, meaning that the **RST** of a snip shall not occur between the **RST** and **RET** of any other snip on the same track. If the Type-11 record refers to an analog recording, the method of determining the start time shall be given in the comment item of this field.
* The fourth item **(relative end time/RET**) is a mandatory integer for every snip identified by an **SPI** and indicates in microseconds the time of the end of the snip relative to the beginning of the voice recording. The item can contain up to 11 digits, meaning that the snip may end anywhere within the 27 hour voice recording. Because each snip is obtained independently from a larger voice recording, snips from the same track of the audio object of Field 11.003 shall not overlap, meaning that the **RET** of a snip shall not occur between the **RST** and **RET** of any other snip from the same track.
* The fifth item **(comment/COM**) is an optional unrestricted text string of up to 4000 characters in length that allows for comments of any type to be made on a snip. This allows for comments on a snip-by-snip basis, including comments on the source of each snip. This comment field could contain word- or phonic level transcriptions, language translations or security classification markings, as specified in exchange agreements.
1. **Field 11.025: Diarization/DIA**

This field (**Diarization/DIA**) is optional and indicates whether the voice recording has been diarized, meaning that time markings are included in **Field 11.026** to indicate the speech segments of interest pertaining to the subject of this Type-11 record.

* The first information item (**diarization indicator/DII)** is mandatory if this field is used. It is a binary indicator that indicates whether the voice recording is accompanied by a segment diary in **Field 11.026** indicating speech segments from the voice signal subject of the Type-11 record. 0 indicates no accompanying diary and 1 indicates one or more accompanying diaries.
* The second information item (**diarization authority/DAU)** is an optional text field of up to 300 characters containing information about the agency that performed the diarization. Agencies undertaking diarization activities on the original speech should log their actions by appending to this item and noting the change of field contents in the Type-98 record and / or **Field 11.902** of this record
* The third information item (**comment/COM) is an optional** unrestricted text string of up to 4000 characters that may contain text information about the diarization activities undertaken on the voice data.
1. **Field 11.026: Segment Diary/SGD**

This field only appears if **Field 11.025** is present and **DII** = 1. This field (**segment diary/SDI**) contains repeating subfields that name and locate the segments within the voice recording of this Type-11 record associated with a single speaker. Within a Type-11 record, there may be only one segment diary describing a single speaker within the single voice recording. If additional diarizations of this voice recording are necessary -- for example, to locate segments of speech from a second speaker in the voice recording, additional Type-11 records must be created. Each segment diarized shall contain speech from the subject of this record, although a segment may contain speech collisions. The first four items (uniquely identifying the segments, identifying the tracks from the audio media object of Field 11.003 to which the segment number applies, and giving start and end times of each relative to the absolute beginning of the voice recording) are mandatory if this field is used and shall repeat for each speech segment identified. A fifth item is optional and accommodates comments on the individual segments. This record type accommodates up to 600,000 speech segments as repeating subfields. For voice recordings consisting of snips, the snipping diary **SPD** of Field 11.024 may be included in the **SGD** as a subset and may be identical.

* The first item (**segment identifier/SID**) is mandatory in each subfield and uniquely numbers the segment to which the following items in the subfield apply. There is no requirement that the segments be numbered sequentially in sequential subfields. The **SID** may contain up to 6 digits, but the number of segments identified in the field (the total number of recurring subfields) is limited to 600,000.
* The second item (**tracks/TRK**) is mandatory if item **PMO\_TRK** in Field 11.010 or **CDC\_NCH** of Field 11.012 is greater than one and lists all tracks or channels on the recording to which the segment identifier applies. The track numbers are separated by commas. No value in this list should be greater than the value of **PMO\_TRK** or **CDC\_NCH**, whichever applies. For example, in the case of a two-track stereo recording where both tracks contain a segment at the same start and end times, this item will be “1,2”
* The third item (**relative start time/RST**) is a mandatory integer for every segment identified and indicates in microseconds the time of the start of the segment relative to the absolute beginning of the voice recording. The item can contain up to 11 digits, meaning that the segment can start at any time within the 27 hour voice recording. Because each segment is expected to be dominated by the primary subject of this Type-11 record, it is expected that segments from the same track of the audio object identified in Field 11.003 not will overlap, meaning that the **RST** of a segment is not expected to occur earlier than the end of a previous segment from the same track, although this is not prohibited. In multiple ANSI/NIST-ITL transactions involving multiple speakers using the same voice data record, segments on the same track across the transactions may overlap during periods of voice collision. If the Type-11 record refers to an analog recording, the method of determining the start time shall be given in the comment item of this subfield.
* The fourth item **(relative end time/RET**) is mandatory for every segment and indicates in microseconds the time of the end of the segment relative to the absolute beginning of the voice recording. The item can contain up to 11 digits, meaning that the segment can end at any time within the 27 hour voice recording. As with the **RST**, it is expected that segments from the subject of this Type-11 record will not overlap, although this is not prohibited.
* The fifth item **(comment/COM**) is an optional unrestricted text string of a maximum of 10,000 characters in length that allows for comments of any type to be made on a segment. This comment item could contain word- or phonic level transcriptions, language translations or security classification markings, as specified in exchange agreements.
1. **Field 11.027-030: Reserved Fields**

These fields are reserved for future use by ANSI/NIST-ITL.

1. **Field 11.031: Time of Segment Recording /TME**

This optional field (**Time of Segment Recording/TME**) contains subfields, each referring to a segment identified in either the snip diary **SPD** of **Field 11.024** or the segment diary **SGD** of **Field 11.026** and gives the date, start, and end times of the original transduction of the contemporaneous vocalizations in the identified segment. This field is only present if **Field 11.024** or **Field 11.026** is present in this record. This field also accommodates circumstances in which the original voice recording was tagged with a time and date field. There is no requirement that the date and times for the original recording match the dates and times of the tags, if the tags have been determined to be inaccurate.

* The first item (**diary identifier/DIA**) is a mandatory in each subfield and is a binary value that indicates the diary to which this subfield refers. If this item refers to a segment in the **SPD** of **Field 11.024**, the value is 0. If this item refers to a segment in the **SGD** of Field 11.026, the value is 1.
* The second item **(segment identifier/SID**) is mandatory and gives the segment identifier from the diary given in **DIA** to which the values in this subfield pertain. Together, the first and second items of each subfield uniquely identify the segment to which the following items apply.
* The third item **(date of original recording/DOR**) is optional and gives the date of the original, contemporaneous capture of the voice data in the segment identified. See **Section 7.7.2.3.**
* The fourth item **(tagged date/TDT**) is optional and gives the date tagged on the original, contemporaneous capture of the voice data in the segment identified. This item may be different from the value of the **DOR** above, if the tag is determined to be inaccurate. See **Section 7.7.2.3.**
* The fifth item (**start time of segment recording/SRT**) is optional and gives the local start time of the original, contemporaneous capture of the voice data in the segment identified. See **Section 7.7.2.4** **Local date and time** for details.
* The sixth item (**tagged** **start time/TST**) is optional and gives the time tagged on original, contemporaneous capture of the voice data at the start of the segment identified. This item may be different from the value of the **SRT** above, if the tag is determined to be inaccurate See **Section 7.7.2.4** **Local date and time** for details.
* The seventh item (**end time of recording/END**) is optional and gives the local end time of the original, contemporaneous capture of the voice data in the segment identified. See **Section 7.7.2.4** **Local date and time** for details.
* The eight item (**tagged** **end time/TET**) is optional and gives the time tagged on original, contemporaneous capture of the voice data at the end of the segment identified. This item may be different from the value of the **END** above, if the tag is determined to be inaccurate. See **Section 7.7.2.4** **Local date and time** for details.
* The ninth item (**Source of the time/STM**) is an optional string of up to 300 characters that gives the reference for the values used for **DOR**, **SRT** and **END**.
* The tenth item (**comment/COM**) is an unrestricted text string of up to 4000 characters in length that allows for comments of any type to be made on the timings of the segment recording, including the perceived accuracy of the values of **DOR**, **SRT** and **END**.
1. **Field 11.032: Segment Geographical Information/GEO**

This field (**Segment Geographical Information/GEO**) contains repeating subfields, each referring to a segment identified in either the snip diary **SPD** of **Field 11.024** or the segment diary **SGD** of **Field 11.026** and giving geographical location of the primary subject of the Type-11 record at the beginning of that segment. This field is only present if **Field 11.024** or **Field 11.026** is present in this record.

* The first item (**diary identifier/DIA**) is a mandatory in each subfield and is a binary indicator of the diary to which this subfield refers. If this item refers to a segment in the **SPD** of **Field 11.024**, the value is 0. If this item refers to a segment in the **SGD** of Field **11.026**, the value is 1.
* The second item (**segment identifier/SID**) is mandatory in each subfield and gives the segment identifiers from diary to which the values in this subfield pertain. The number of segment identifiers listed is limited to 600,000. A value of 0 in this subfield indicates the segment geographical information in this subfield shall be considered the default value for all segments not specifically identified in other occurrences of this subfield. If multiple segments are identified, they are designated as integers separated by commas.
* The third item (**segment cell phone tower code/SCT)** is optional and identifies the cell phone tower, if any, that relayed the audio data at the start of the segment or segments referred to in this subfield. It is a text field of up to 100 unrestricted characters.
* The next six items are latitude and longitude values.  **See Section 7.7.3**
* The tenth information item (**elevation / ELE)** is optional. It is expressed in meters. **See Section 7.7.3.** Permitted values are in the range of -442 to 8848 meters. For elevations outside of this range, the lowest or highest values shall be used, as appropriate.
* The eleventh information item (**geodetic datum code / GDC)** is optional. **See Section 7.7.3.**
* The twelfth, thirteenth and fourteenth information items (**GCM**/**GCE**/**GCN**) are treated as a group and are optional. These three information items together are a coordinate which represents a location with a Universal Transverse Mercator (**UTM**) coordinate. If any of these three information items is present, all shall be present. **See Section 7.7.3**
* The fifteenth information item (**geographic reference text /GRT)** is optional. **See Section 7.7.3.**
* A sixteenth information item (**geographic coordinate other system identifier / OSI)** is optional and allows for other coordinate systems and the inclusion of geographic landmarks. **See Section 7.7.3.**

* A seventeenth information item (**geographic coordinate other system value / OCV)** is optional andshall only be present if **OSI** is present in the record**. See Section 7.7.3**

The Geographic entry may be modified slightly based upon some issues related to NIEM in the XML encoding.

1. **Field 11.033: Segment Quality Values/SQV**

This field (**Segment Quality Values/SQV**) contains repeating subfields, each referring to a list of segments identified in either the snip diary **SPD** of **Field 11.024** or the segment diary **SGD** of **Field 11.026.**  The items in each subfield give an assessment of the quality of the voice data within the segments identified in the subfield. This field is present only if Field 11.024 or Field 11.026 exists in the record. This contrasts with **Field 11.012** that gives the general quality across the entire audio recording. Values in this field dominate any values given in **Field 11.012**. It is possible for each segment given in the associated diary to have different quality. The subfields accommodate only a single quality value. If segments have multiple quality values based on different types of quality assessments, then multiple subfields are entered for those segments.

* The first item (**diary identifier/DIA**) is mandatory and is a binary indicator of the diary to which this subfield refers. If this item refers to a segment in the **SPD** of **Field 11.024**, the value is 0. If this item refers to a segment in the **SGD** of **Field 11.026**, the value is 1.
* The second item (**segment identifiers/SID**) is a mandatory list of integers and gives the segment identifiers from the diary to which the values in this subfield pertain. The number of segment identifiers listed is limited to 600,000. A value of 0 in this subfield indicates the segment quality information in this subfield shall be considered the default value for all segments not specifically identified in other subfields of this field. If multiple segments are entered, they are listed as integers separated by commas.
* The third **information item (quality value/QVU**) is mandatory and shall indicate the segment quality value between 0 (low quality) and 100 (high quality). A value of 255 indicates that quality was not assessed. An example would be the *Speech Intelligibillity Index, ANSI 3.5 1997.*
* A fourth information item is mandatory and shall specify the ID of the vendor of the quality assessment algorithm used to calculate the quality score, which is an **algorithm vendor identification / QAV**. This 4-digit hex value (See **Section 5.5 Character types )** is assigned by IBIA and expressed as four characters. The IBIA maintains the Vendor Registry of CBEFF Biometric Organizations that map the value in this subfield to a registered organization. A value of 0x00 indicates a vendor without a designation by IBIA. In such case, an entry shall be made in COM of this subfield describing the algorithm and its owner / vendor.
* A fifth information item is mandatory and shall specify a numeric product code assigned by the vendor of the quality assessment algorithm, which may be registered with the IBIA, but registration is not required. This is the **algorithm product identification / QAP** that indicates which of the vendor’s algorithms was used in the calculation of the quality score. This information item contains the integer product code and should be within the range 0 to 65,534. A value of 0 indicates a vendor without a designation by IBIA. In such case, an entry shall be made in COM of this subfield describing the algorithm and its owner / vendor.
* The sixth information item (**comment/COM**) is optional but shall be used to provide information about the quality assessment process, including a description of any unregistered quality assessment algorithms used. (if QAV= 0000 or QAP = 0)
1. **Field 11.034: Vocal Collision Indicator/VCI**

This optional field (**Vocal Collision Indicator/VCI**) contains 2 mandatory information items, each referring to a list of segments identified in either the snip diary **SPD** of **Field 11.024** or the segment diary **SGD** of **Field 11.026** and indicating that a vocal collision (two or more persons talking at once) occurs within the segment. This field shall only appear if **Field 11.024** or **Field 11.026** exists in this record.

* The first item (**diary identifier/DIA**) is mandatory and is a binary indicator of the diary to which this subfield refers. If this item refers to a segment in the **SPD** of **Field 11.024**, the value is 0. If this item refers to a segment in the **SGD** of **Field 11.0**26, the value is 1.
* The second item (**segment identifier/SID**) is a mandatory list of integers separated by commas and gives the segment identifiers from the diary named in the item above in which vocal collisions occur. There may be up to 600,000 segments identified in this subfield.
1. **Field 11.035: Processing Priority /PPY**

This optional field (**Processing Priority/PPY**) contains repeating subfields, each referring to a list of segments identified in either the snip diary **SPD** of **Field 11.024** or the segment diary **SGD** of **Field 11.026** and indicating the priority with which the segments named in those diaries should be processed. If this field exists, segments not identified should be given the lowest priority. This field is distinct from **Field 1.006**, which gives a priority for processing the entire transaction.

* The first item (**diary identifier/DIA**) is mandatory and is a binary indicator of the diary to which this subfield refers. If this item refers to a segment in the **SPD** of **Field 11.024**, the value is 0. If this item refers to a segment in the **SGD** of **Field 11.026**, the value is 1.
* The second item (**segment identifier/SID**) is a mandatory list of integers, separated by commas, and gives the segment identifiers from diary named in the first item above to which the values in this subfield pertain. There may be up to 600,000 values of this field, one for each segment identified in the diaries of **Field 11.024** or **Field 11.026**. A value of 0 in this item indicates the segment content information in this field shall be considered the default value for all segments not specifically identified in other subfields of this field.
* The third information item (**processing priority/ PTY**) is optional and indicates the priority with which the segments identified in this subfield should be processed. Priority values shall be between 1 and 9 inclusive. As with **Field 1.006**, 1 will indicate the highest priority and 9 the lowest.
1. **Field 11.036: Segment Content/SCN**

This optional field (**Segment Content/SCN**) contains subfields, each referring to a segment identified in either the snip diary **SPD** of **Field 11.024** or the segment diary **SGD** of **Field 11.026.** Each subfield gives an assessment of the content of the voice data within the identified segment and includes provision for semantic transcripts, phonetic transcriptions and translations of the segment. It may only appear if Field 11.024 or Field 11.026 is present in this record.

* The first item (**diary identifier/DIA**) is mandatory and is a binary indicator of the diary to which this subfield refers. If this item refers to a segment in the **SPD** of **Field 11.024**, the value is 0. If this item refers to a segment in the **SGD** of **Field 11.026**, the value is 1.
* The second item (**segment identifier/SID**) is a mandatory list of integers separated by commas and gives the segment identifiers from diary to which the values in this subfield pertain. There may be 600,000 values of this item, one for each segment identified in related diary. A value of 0 of this item indicates the segment content information in this subfield shall be considered the default value for all segments not specifically identified in other subfields of this field.
* The third information item (**transcript/TRN**) is an optional text field of up to 100,000 characters and may contain a semantic transcription, a phonetic transcription, translation, or comments on the segment.
* The fourth information item (**transcript authority/TRA**) is an optional text field of up to 10,000 characters and shall state the authority providing the transcription, translation or comments if **TRN** is used. If an automated process was used to develop the transcript, information about the process (i.e., the automated algorithm used) should be included in this text

1. **Field 11.037: Segment Speaker Characteristics/SCC**

This optional field (**Segment Speech Characteristics/SCC**) contains subfields, each referring to a segment identified in either the snip diary **SPD** of **Field 11.024** or the segment diary **SGD** of **Field 11.026**. Each subfield gives an assessment of the characteristics of the voice within the segment, including intelligibility, emotional state and impairment. This field shall only appear if Field 11.024 or Field 11.026 exists in the record.

* The first item (**diary identifier/DIA**) is mandatory and is a binary indicator of the diary to which this subfield refers. If this item refers to a segment in the **SPD** of **Field 11.024**, the value is 0. If this item refers to a segment in the **SGD** of **Field 11.026**, the value is 1.
* The second item (**segment identifier/SID**) is a mandatory list of integers separated by commas and gives the segment identifiers from **Field 11.024** to which the values in this subfield pertain. There may be up to 600,000 values in this item, one for each segment identified in **Field 11.026**. A value of 0 indicates the segment content information in this item shall be considered the default value for all segments not specifically identified in other occurrences of this item.
* The third information item **(impairment/IMP**) is optional and shall indicate an observed level of neurological diminishment, whether from fatigue, disease, trauma, or the influence of medication/substances, across the speech segments identified. No attempt is made to differentiate the sources of impairment**.** The valueshall be an integer between 0 (no noticed impairment) and 5 (significant), inclusive.
* The fourth item (**language being spoken/LBS**) is optional and gives the 3 character *ISO 639-3* code for the dominant language in the segments identified in this subfield.
* A fifth information item (**language proficiency/LPF**) is optional and rates the fluency of the language being spoken on a scale of 0 (no proficiency) to 9 (high proficiency).
* The sixth information item **(style of speech/STY**) is optional and shall be an integer as given in **Table V-6**. There may be no more than one value for each of the segments identified in this subfield and will indicate the dominant style of speech within the segments. If attribute code “10” is chosen to indicate “other”, additional explanation should be included in the tenth item (**comment/COM**) below.

**Table V-6**

**Style of Speech**

|  |  |
| --- | --- |
| **Style of Speech** | **Attribute Code** |
| Unknown | 0 |
| Public speech (oratory) | 1 |
| Conversational telephone | 2 |
| Conversation face-to-face | 3 |
| Read | 4 |
| Prompted/repeated | 5 |
| Storytelling/Picture description | 6 |
| Map task and related methods | 7 |
| Interview | 8 |
| Recited/memorized | 9 |
| Other | 10 |
| **RESERVED FOR FUTURE USE only by ANSI/NIST-ITL** | 11-20 |

* The seventh information item (**intelligibility/INT**) is optional and shall be an integer from 0 (unintelligible) to 9 (clear and fully intelligible).
* The eighth information item (**intimacy/ITM**) is optional and indicates the degree of familiarity between the data subject and the interlocutor, with 0 indicating no familiarity and 5 indicating high familiarity/intimacy.
* The ninth information item **(health status/HST**) is optional text noting any observable health issues impacting the data subject during the speech segment, such as symptoms of the common cold (hoarse voice, pitch lowering, increased nasality) and an indicator if the data subject regularly smokes tobacco products.
* The tenth information item **(emotional state/EM**) is an optional integer giving an estimation of the emotional state of the data subject across the segments identified in this subfield. Admissible attribute values are given in **Table V-7**. Only one value for this item is allowed across all of the segments identified in this subfield. If attribute code “8” is chosen to indicate “other”, additional explanation may be included in the tenth item (**comment/COM**) below.

**Table V-7**

**Emotional State**

|  |  |
| --- | --- |
| **Emotional State** | **Attribute Code** |
| Unknown | 0 |
| Calm | 1 |
| Hurried | 2 |
| Happy/joyful | 3 |
| Angry | 4 |
| Fearful | 5 |
| Agitated /Combative | 6 |
| Defensive | 7 |
| Crying | 8 |
| Other | 9 |
| **RESERVED FOR FUTURE USE only by ANSI/NIST-ITL** | 10-20 |

* The eleventh information item **(vocal effort/VEF**) is an optional integer between 0 (very low vocal effort) and 5(screaming/crying) which reports perceived vocal effort across the identified segments. Only one value is allowed for this item in each subfield.
* The twelfth information item (**vocal style/VSY**) is an optional integer assessing the predominant vocal style across the identified segments. The attribute value shall be chosen from **Table V-8**. Only one value is allowed for this item in each subfield.

**Table V-8**

**Vocal Style**

|  |  |
| --- | --- |
| **Vocal Style** | **Attribute Code** |
| Unknown | 0 |
| Spoken | 1 |
| Whispered | 2 |
| Sung | 3 |
| Chanted | 4 |
| Rapped | 5 |
| Mantra | 6 |
| Falsetto/Head voice | 7 |
| Spoken with laughter | 8 |
| Megaphone/Public Address System | 9 |
| Shouting/yelling | 10 |
| Other | 11 |
| **RESERVED FOR FUTURE USE only by ANSI/NIST-ITL**  | 12-20 |

* The thirteenth information item (**awareness of the recording process/AWR**) is optional and indicates whether the data subject is aware that a recording is being made. 0 indicates unknown, 1 indicates aware and 2 indicates unaware.

* The fourteenth (**script/SCR**) is optional and may be used to give the script used for read, prompted or repeated speech. This item may have up to 9,999 characters.
* The fifteenth (**comment/COM**) is optional and may be used to give additional information about the quality assessment process, including a description of any unregistered quality assessment algorithms used, notes on any known external stresses applicable to the data subject, such as extreme environmental conditions or heavy physical or cognitive load, and a description of how the values in the items of this subfield were assigned. This item may have up to 4,000 characters.
1. **Field 11.038: Segment Channel/SCH**

This field (**Segment Channel/SCH**) contains subfields, each referring to a segment identified in either the snip diary **SPD** of **Field 11.024** or the segment diary **SGD** of **Field 11.026**. Each subfield describes the transducer and transmission channel within the identified segments. This field shall only be present if **Field 11.024** or **Field 11.026** appears in this record.

* The first item (**diary identifier/DIA**) is mandatory and is a binary indicator of the diary to which this subfield refers. If this item refers to a segment in the **SPD** of **Field 11.024**, the value is 0. If this item refers to a segment in the **SGD** of **Field 11.026**, the value is 1.
* The second item (**segment identifier/SID**) is a mandatory list of integers separated by commas, and gives the segment identifiers from the diary to which the values in this subfield pertain. There may be up to 600,000 values in this item. A value of 0 in this item indicates the segment content information in this subfield shall be considered the default value for all segments not specifically identified in other subfields of this field.
* The third item **(transducer type/TYP)** is an optional integer with attribute values given in **Table V-8.** It is recognized that for most of the acquisition sources in Field 11.006 **REC\_AQS**, as specified by Table 83, the transducer type will not be known.

**Table V-9**

**Transducer Type**

|  |  |
| --- | --- |
| **Transducer Type** | **Attribute Code** |
| Unknown | 0 |
| Array | 1 |
| Multiple style microphones | 2 |
| Earbud | 3 |
| Body Wire  | 4 |
| Microphone | 5 |
| Handset | 6 |
| Headset | 7 |
| Speaker phone | 8 |
| Lapel Microphone | 9 |
| Other | 10 |
| **RESERVED FOR FUTURE USE only by ANSI/NIST-ITL** | 11-99 |

* The fourth item (**transducer/TRN)** is an optional integer that specifies the transducer type as unknown=0, carbon=1, electret=2, or other=3. Transducer arrays using mixed transducer types shall be designated “other”.
* The fifth item (**capture environment/ENV**) is an optional text field of up to 4000 characters to describe the acoustic environment of the recording. Examples of text placed in this item would be “reverberant busy restaurant”, “urban street”, “public park during day”.
* The sixth item (**distance to transducer/DST**) is an optional integer and specifies the approximate distance in centimeters, rounded to the nearest integer number of centimeters, between the speaker in the identified segments and the transducer. A value of 0 will be used if the distance is less than one-centimeter. Some example distances: handheld = 5cm; throat mic = 0cm, mobile telephone = 15cm; Voice-over-internet-protocol (VOIP) with a computer = 80cm, unless other information is available.
* The seventh item (**acquisition source/ACS**) is an optional integer that specifies the source from which the voice in the identified segments was received. Only one value is allowed. Permissible values are given in **Table 83** of the **Type-20** record. Any conflict between this value and **Field 11.006** **REC\_AQS** shall be resolved by taking this item to be correct for all segments identified in the subfield, **SCH\_DIA** and **SCH\_SID,** of this occurrence of **Field 11.038**.
* The eighth item (**alteration/ALT**) is an optional, unrestricted string for a description of any digital masking between transducer and recording, disguisers or other attempts to change the voice quality.
* The ninth information item (**comment/COM**) is an optional, unrestricted string for additional information to identify or describe the transduction and transmission channels of the identified segments.
1. **Field 11.039-050: Reserved Fields**

These fields are reserved for future use by ANSI/NIST-ITL.

1. **Field 11.051: Comments/COM**

This field (**Comments/COM**) is an optional unrestricted text string of up to 4000 characters in length that may contain comments of any type on the **Type 11** record as a whole. Comments on individual segments shall be given in **Field 11.024, SNP\_COM**, or in **Field 11.026, SGD\_COM**. This field should record any intellectual property rights associated with any of the segments in the voice recording, any court orders related to the voice recording and any administrative data not included in other fields.

1. **Fields 11.052-099: Reserved Fields**

These fields are reserved for future use by ANSI/NIST-ITL.

1. **Fields 11.100-900: User-defined fields / UDF**

These fields are user-defined fields. Their size and content shall be defined by the user and be in accordance with the receiving agency

1. **Field 11.901: Reserved field**

This field is reserved for future use by ANSI/NIST-ITL.

1. **Field 11.902: Annotation information / ANN**

This is an optional field, listing the operations performed on the original source in order to prepare it for inclusion in a biometric record type. This field logs information pertaining to this Type-11 record and the voice recording pointed to or included herein. See **Section 7.4.1.** This section is not intended to contain any transcriptions or translations themselves, but may contain information about the source of such fields in the record.

1. **Field 11.903: Device unique identifier/ DUI**

This is an optional field. See **Clause 7.7.1.1** for details. This field will require future development. Which equipment? Is this field needed or is it confusing?

1. **Field 11.904: Make/Model/Serial number / MMS**

This is an optional field. See **Clause 7.7.1.2** for details. This field will require future development. Which equipment? Is this field needed or is it confusing?

1. **Fields 11.905-992: Reserved Fields**

These fields are reserved for future use by ANSI/NIST-ITL.

1. **Field 11.993: Source agency name / SAN**

This is an optional field. It may contain up to 125 Unicode characters. This is the name of the agency referred to in **Field 11.004** using the identifier given by domain administrator.

1. **Field 11.994: External file reference / EFR**

This conditional field shall be used to enter the URL / URI or other unique reference to a storage location for all source representations, if the data is not contained in **Field 11.999.** If this field is used, **Field 11.999** shall not be set. However, one of the two fields shall be present in all instances of this record type. A non-URL reference might be similar to: “Case 2009:1468 AV Tape 5”. It is highly recommended that the user state the format of the external file in **Field 11.051: Comment / COM .**

1. **Field 11.995: Associated context / ASC**

This optional field refers to one or more **Type-21 Records**. See **Section 7.3.3. Record Type-21** contains audio, video and images that are NOT used to derive the biometric data in **Field 11.999: Voice Record / DATA** but that may be relevant to the collection of that data.

1. **Field 11.996: Hash/ HAS**

This optional field shall contain the hash value of the data in **Field 11.999: Voice Data** of this record, calculated using SHA-256. See **Section 7.5.2**. Use of the hash enables the receiver of the data to check that the data has been transmitted correctly, and may also be used for quick searches of large databases to determine if the data already exist in the database. It is not intended as an information assurance check, which is handled by **Record Type-98.**

1. **Field 11.997: Source representation / SOR**

This optional field refers to a representation in **Record Type-20** with the same **SRN**.

1. **Field 11.998: Reserved field**

This field is reserved for future use by ANSI/NIST-ITL.

1. **Field 11.999: Voice record / DATA**

This field contains the voice data. See Section **7.2** for details.

Annex B of ANSI/NIST-ITL 1-2011 Table 97 is updated as follows:

Record Identifier Logical record contents Type of Data

 11 Voice ASCII/Binary

Annex B Section B.2.7 is updated:

There are no special requirements for this record type.

Annex G, Insert table Type-11

DEVELOP TABLE FOR XML REPRESENTATION HERE

1. Available at <http://www.nist.gov/itl/iad/ig/ansi_standard.cfm> [↑](#footnote-ref-2)
2. Although this was an original requirement of the IVBC, Record Type-11 does not have a metric or a means of codification of “accent”, since it is not objectively quantifiable. [↑](#footnote-ref-3)
3. Available at <http://www.nist.gov/itl/iad/ig/ansi_standard.cfm> [↑](#footnote-ref-4)
4. This table is copied from the Type-20 record, but is included here for convenience. [↑](#footnote-ref-5)