
**Information technology — Biometric
data interchange formats —**

**Part 13:
Voice data**

*Technologies de l'information — Formats d'échanges de données
biométriques —*

Partie 13: Données relatives à la voix





COPYRIGHT PROTECTED DOCUMENT

© ISO/IEC 2018

All rights reserved. Unless otherwise specified, or required in the context of its implementation, no part of this publication may be reproduced or utilized otherwise in any form or by any means, electronic or mechanical, including photocopying, or posting on the internet or an intranet, without prior written permission. Permission can be requested from either ISO at the address below or ISO's member body in the country of the requester.

ISO copyright office
CP 401 • Ch. de Blandonnet 8
CH-1214 Vernier, Geneva
Phone: +41 22 749 01 11
Fax: +41 22 749 09 47
Email: copyright@iso.org
Website: www.iso.org

Published in Switzerland

Contents

| | Page |
|--|-----------|
| Foreword | iv |
| Introduction | v |
| 1 Scope | 1 |
| 2 Normative references | 1 |
| 3 Terms and definitions | 1 |
| 4 Abbreviated terms | 5 |
| 5 Conformance | 6 |
| 6 Processes and identifiers | 7 |
| 6.1 Capture processes and utterances | 7 |
| 6.1.1 Introduction | 7 |
| 6.1.2 Voice utterance | 7 |
| 6.1.3 Structure of a capture process | 7 |
| 6.2 Registered format type identifiers | 9 |
| 7 General voice data interchange format (BDB) | 9 |
| 7.1 Overview | 9 |
| 7.2 Conventions | 10 |
| 7.3 Voice record general header | 10 |
| 7.3.1 Overview | 10 |
| 7.3.2 Version | 11 |
| 7.3.3 Session ID | 11 |
| 7.3.4 Channel | 11 |
| 7.3.5 Capture device | 12 |
| 7.3.6 Transducer | 12 |
| 7.3.7 Audio meta information | 13 |
| 7.3.8 Capture process protocol | 14 |
| 7.3.9 Extended vendor data | 14 |
| 7.4 Voice representation header | 14 |
| 7.4.1 Overview | 14 |
| 7.4.2 Date and time | 14 |
| 7.4.3 Audio content | 15 |
| 7.4.4 Quality information | 17 |
| 7.4.5 Signal enhancement | 18 |
| 7.4.6 Extended vendor data | 19 |
| 7.5 Voice representation data | 19 |
| 7.6 Schema | 19 |
| 7.7 Example | 23 |
| Annex A (normative) Conformance testing methodology | 25 |
| Bibliography | 26 |

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work. In the field of information technology, ISO and IEC have established a joint technical committee, ISO/IEC JTC 1.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation on the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT) see the following URL: www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 37, *Biometrics*.

A list of all the parts in the ISO/IEC 19794 series, can be found on the ISO website.

Introduction

This document assumes that the voice data interchange record is to be attributed to a single individual and recorded in a single session. Voice data is a time record of audible, acoustic vibrations produced by a human in the course of a verbal interaction and will generally contain both speech and non-speech vocal sounds, as well as non-vocal sounds to be considered “noise” in this context. In addition to serving the linguistic function of semantic information transfer, voice data contains both acoustic and semantic information that can be used to recognize speakers. It is the collection, storage and transmission of voice data containing speech for the purpose of recognizing individuals that is the focus of this document.

This format is designed specifically to support a wide variety of automatic speaker recognition applications, including both text-dependent and text-independent Speaker Identification and Verification (SIV) and enrolment, with minimal assumptions made regarding the voice data capture conditions or the collection environment. This document is intended to be sufficiently general that speaker recognition applications beyond traditional SIV could also be supported, such as linking utterances to the same unknown speaker, and determining that a known speaker is not the source of an utterance. The differentiation between speech used to create the reference for future comparisons (which in some applications is called “enrolment”), and that used to create voice representations (VRs) queried against the references, might occur only at the point of application, thus requiring each stored speech record to potentially support either reference or query creation. Further, automated speaker recognition might incorporate related technologies, such as speech and language recognition, not only in current algorithms and applications, but in future ways that cannot be anticipated. Therefore, this document is written from a very broad perspective with the intent of supporting the broadest possible range of speaker recognition applications and technical approaches.

Information technology — Biometric data interchange formats —

Part 13: Voice data

1 Scope

This document specifies a data interchange format that can be used for storing, recording, and transmitting digitized acoustic human voice data (speech) assumed to be from a single speaker recorded in a single session. This format is designed specifically to support a wide variety of Speaker Identification and Verification (SIV) applications, both text-dependent and text-independent, with minimal assumptions made regarding the voice data capture conditions or the collection environment. Other uses for the data encapsulated in this format, such as automated speech recognition (ASR), may be possible, but are not addressed in this document. This document also does not address handling of data that has been processed to the feature or voice model levels. No application-specific requirements, equipment, or features are addressed in this document. This document supports the optional inclusion of non-standardized extended data. This document allows both the original data captured and digitally-processed (enhanced) voice data to be exchanged. A description of any processing of the original source input is intended to be included in the metadata associated with the voice representations (VRs). This document does not address data streaming.

Provisions that stored and transmitted biometric data be time-stamped and that cryptographic techniques be used to protect their authenticity, integrity and confidentiality are out of the scope of this document.

Information formatted in accordance with this document can be recorded on machine-readable media or can be transmitted by data communication between systems.

A general content-oriented subclause describing the voice data interchange format is followed by a subclause addressing an XML schema definition.

This document includes vocabulary in common use by the speech and speaker recognition community, as well as terminology from other ISO standards.

2 Normative references

The following documents are referred to in the text in such a way that some or all of their content constitutes requirements of this document. For dated references, only the edition cited applies. For undated references, the latest edition of the referenced document (including any amendments) applies.

ISO/IEC 19794-1, *Information technology — Biometric data interchange formats — Part 1: Framework*

ISO/IEC 19785-1, *Information technology — Common Biometric Exchange Formats Framework — Part 1: Data element specification*

ISO/IEC 2382-37, *Information technology — Vocabulary — Part 37: Biometrics*