ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 58668.11— 2019 (ИСО/МЭК 19794-13: 2018)

Информационные технологии

БИОМЕТРИЯ

Форматы обмена биометрическими данными

Часть 11

Данные голоса

(ISO/IEC 19794-13:2018, Information technology — Biometric data interchange formats — Part 13: Voice data, MOD)

Издание официальное

Москва
Стандартинформ
2019

gost.e-ecolog.ru

Предисловие

1 ПОДГОТОВЛЕН Акционерным обществом «Всероссийский научно-исследовательский институт сертификации» (АО «ВНИИС»), Некоммерческим партнерством «Русское общество содействия развитию биометрических технологий, систем и коммуникаций» (Некоммерческое партнерство «Русское биометрическое общество») на основе собственного перевода на русский язык англоязычной версии стандарта, указанного в пункте 4, при консультативной поддержке Федерального государственного бюджетного образовательного учреждения высшего образования «Московский государственный технический университет имени Н.Э. Баумана (национальный исследовательский университет)» (МГТУ им. Н.Э. Баумана)

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 098 «Биометрия и биомониторинг»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 19 ноября 2019 г. № 1188-ст

4 Настоящий стандарт является модифицированным по отношению к международному стандарту ИСО/МЭК 19794-13:2018 «Информационные технологии. Форматы обмена биометрическими данными. Часть 13. Данные голоса» (ISO/IEC 19794-13:2018 «Information technology — Biometric data interchange Formats — Part 13: Voice data», MOD) путем изменения отдельных фраз (слов, значений показателей, ссылок), которые выделены в тексте курсивом. Внесение указанных технических отклонений направлено на учет потребностей национальной экономики Российской Федерации.

Наименование настоящего стандарта изменено относительно наименования указанного международного стандарта для приведения в соответствие с ГОСТ Р 1.5—2012 (пункт 3.5).

Сведения о соответствии ссылочных национальных и межгосударственных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте, приведены в дополнительном приложении ДА.

Сопоставление структуры настоящего стандарта со структурой примененного в нем международного стандарта приведено в дополнительном приложении ДБ

5 ВВЕДЕН ВПЕРВЫЕ

6 Некоторые элементы настоящего стандарта могут быть объектами патентных прав. Международная организация по стандартизации (ИСО) и Международная электротехническая комиссия (МЭК) не несут ответственности за установление подлинности каких-либо или всех таких патентных прав

*Правила применения настоящего стандарта установлены в статье 26 Федерального закона от 29 июня 2015 г. № 162-ФЗ «О стандартизации в Российской Федерации». Информация об изменениях к настоящему стандарту публикуется в ежегодном (по состоянию на 1 января текущего года) информационном указателе «Национальные стандарты», а официальный текст изменений и поправок — в ежемесячном информационном указателе «Национальные стандарты». В случае пересмотра (замены) или отмены настоящего стандарта соответствующее уведомление будет опубликовано в ближайшем выпуске ежемесячного информационного указателя «Национальные стандарты». Соответствующая информация, уведомление и тексты размещаются также в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет (*gost.e-ecolog.ru

© ISO, 2018 — Все права сохраняются
© Стандартинформ, оформление, 2019

Настоящий стандарт не может быть полностью или частично воспроизведен, тиражирован и распространен в качестве официального издания без разрешения Федерального агентства по техническому регулированию и метрологии

Содержание

1 Область применения 1
2 Нормативные ссылки 2
3 Термины и определения 2
4 Сокращения и обозначения 5
5 Соответствие 5
6 Процессы и идентификаторы 6
6.1 Процесс сбора биометрических данных 6
6.2 Идентификатор типа зарегистрированного формата 8
7 Общий формат обмена данными голоса 8
7.1 Общие требования 8
7.2 Соглашения 9
7.3 Блок «Общий заголовок» записи голоса 9
7.4 Блок «Заголовок представления» 13
7.5 Данные голосового представления 17
7.6 XML-схема данных записи голоса 17
7.7 Пример 18
Приложение А (обязательное) Методология испытаний на соответствие 19
Приложение ДА (справочное) Сведения о соответствии ссылочных национальных и межгосударственных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте 20
Приложение ДБ (справочное) Сопоставление структуры настоящего стандарта со структурой примененного в нем международного стандарта 21
Библиография 22

Введение

В настоящем стандарте предполагается, что формат обмена данными голоса должен относиться к одному человеку и быть записан за одну сессию. Данные голоса — это временная запись звуковых акустических вибраций, создаваемых человеком при вербальном взаимодействии, как правило, включающая в себя речевые и неречевые голосовые звуки, а также неголосовые звуки, которые в данном контексте считаются шумом. Помимо лингвистической функции передачи семантической информации, данные голоса содержат акустическую и семантическую составляющие, которые могут быть использованы для распознавания диктора. Настоящий стандарт распространяется на сбор, хранение и передачу данных голоса, содержащих речь, с целью распознавания индивида.

Настоящий формат применим для различных приложений автоматического распознавания диктора, в том числе текстозависимой и текстонезависимой идентификаций и верификации диктора (SIV) и регистрации биометрических данных; условия сбора данных голоса или условия окружающей среды базы данных рассмотрены в минимальном объеме. Настоящий стандарт применяют для приложений распознавания диктора не только традиционных SIV, но и, например, связывания фрагментов речи одного и того же неизвестного диктора и определения, что известный диктор не является источником фрагмента речи. Различие между речью для создания шаблона для будущих сравнений (что в некоторых приложениях называется регистрацией) и речью для создания голосовых представлений (VR), запрашиваемых в отношении шаблонов, может возникать только в точке приложения, что требует от каждой сохраняемой записи речи поддерживать либо создание шаблона, либо создание запроса. Кроме того, автоматическое распознавание дикторов может включать в себя смежные технологии, такие как распознавание речи и языка, не только в существующих алгоритмах и приложениях, но и в будущих способах, которые невозможно предвидеть. Поэтому настоящий стандарт разработан с целью поддержки максимально широкого диапазона приложений распознавания дикторов и технических подходов.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

Информационные технологии

БИОМЕТРИЯ

Форматы обмена биометрическими данными

Часть 11

Данные голоса

Information technology. Biometrics. Biometric data interchange formats.
Part 11. Voice data

Дата введения — 2020—06—01

1 Область применения

Настоящий стандарт устанавливает требования к формату обмена биометрическими данными для хранения, записи и передачи оцифрованных акустических данных человеческого голоса (речи), полученных в течение одной сессии от одного диктора. Настоящий формат применим:

- для различных приложений идентификации и верификации диктора;
- текстонезависимого и текстозависимого распознаваний диктора.

Инкапсулированные данные в данном формате также могут быть использованы для автоматического распознавания речи, но данные вопросы не рассматриваются в настоящем стандарте.

Настоящий стандарт не устанавливает требования:

- к обработке данных, которые были обработаны на уровне голосовых моделей или признаков;
- потоковой передаче данных;
- конкретным приложениям и оборудованию.

Настоящий стандарт поддерживает хранение дополнительных нестандартизированных данных. Настоящий стандарт позволяет обмениваться как исходными данными, так и данными, которые прошли постобработку. Описание любой обработки исходного сигнала должно быть включено в метаданные голосового представления.

Вопросы, связанные с использованием временных меток и технологий криптографической защиты информации, которые используются для обеспечения подлинности, целостности и конфиденциальности хранимых и передаваемых биометрических данных, выходят за рамки настоящего стандарта.

Информация, которая представлена в данном формате, может быть записана на машиночитаемые носители информации или передана по каналам передачи данных между системами.

Определение формата обмена данными голоса в общем виде сопровождается определением схемы XML.

Настоящий стандарт включает в себя словарь терминов и определений, широко используемых профессиональным сообществом в области распознавания диктора/речи, а также терминологию из других стандартов.

Издание официальное

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ 7.67 (ИСО 3166-1:1997) Система стандартов по информации, библиотечному и издательскому делу. Коды названий стран

ГОСТ ИСО 8601 Система стандартов по информации, библиотечному и издательскому делу. Представление дат и времени. Общие требования

ГОСТ ISO/IEC 2382-37 Информационные технологии. Словарь. Часть 37. Биометрия

ГОСТ ISO/IEC 19794-1 Информационные технологии. Биометрия. Форматы обмена биометрическими данными. Часть 1. Структура

Примечание — При пользовании настоящим стандартом целесообразно проверить действие ссылочных стандартов в информационной системе общего пользования — на официальном сайте Федерального агентства по техническому регулированию и метрологии в сети Интернет или по ежегодному информационному указателю «Национальные стандарты», который опубликован по состоянию на 1 января текущего года, и по выпускам ежемесячного информационного указателя «Национальные стандарты» за текущий год. Если заменен ссылочный стандарт, на который дана недатированная ссылка, то рекомендуется использовать действующую ссылку этого стандарта с учетом всех внесенных в данную версию изменений. Если изменен ссылочный стандарт, на который дана датированная ссылка, то рекомендуется использовать версию этого стандарта с указанным выше годом утверждения (принятия). Если после утверждения настоящего стандарта в ссылочный стандарт, на который дана датированная ссылка, внесено изменение, затрагивающее положение, на которое дана ссылка, то это положение рекомендуется применять без учета данного изменения. Если ссылочный стандарт отменен без замены, то положение, в котором дана ссылка на него, рекомендуется применять в части, не затрагивающей эту ссылку.

3 Термины и определения

В настоящем стандарте применены термины по *ГОСТ ISO/IEC 2382-37* и *ГОСТ ISO/IEC 19794-1*, а также следующие термины с соответствующими определениями:

3.1 **разрешение аналого-цифрового преобразователя;** АЦП (analog-to-digital converter (ADC) resolution): Показатель основания 2 (число битов) числа дискретных уровней, которые АЦП может выдать на выходе.

Примечание — Типовые значения разрешения АЦП для звуковых карт: 8, 16, 20 и 24.

3.2 **продолжительность звукозаписи** (audio duration): Продолжительность полной звукозаписи, содержащей все фрагменты речи голосовых представлений, например полные записи звонков.

3.3 **кодирование звукозаписи** (audio encoding): Кодирование, используемое подсистемой сбора биометрических данных, например сотовым телефоном.

Примечания

1 Голосовой сигнал кодируется до передачи в канал. На сегодняшний день существует достаточное число форматов кодирования звукозаписи, и их число будет расти по мере развития телефонов и каналов передачи. Для кодирования звуковых волн используют форматы PCM (ITU-T G.711) и ADPCM (ITU-T G.726), для кодирования AbS используют форматы ACELP (ITU-T G.723.1) и CS-ACELP (ITU-T G.729, приложение A). Импульсно-кодовая модуляция с компандированием по A-характеристике (A-law PCM) и μ-характеристике приведены (mu-law PCM) в ITU-T G.711.

2 Полный список форматов кодирования звукозаписи представлен в 7.4.3.2.

3.4 **сжатие** (compression): Процесс, уменьшающий размер цифрового файла и, соответственно, требуемую скорость передачи данных.

Примечание — Кодирование звукозаписи может включать или не включать сжатие. Сжатие практически всегда происходит с потерями, поэтому оно влияет на исходный голосовой сигнал.

3.5 **граничная частота (нижняя/верхняя)** [cut-off frequency (lower/upper)]: Частота (нижняя/верхняя), на которой звуковая (акустическая) энергия падает на 3 дБ ниже средней энергии в полосе пропускания.

3.6 **дальнее звуковое поле** (far-field): Область, достаточно удаленная от источника, где распределение углового поля не зависит от расстояния до источника.

3.7 **интерактивная голосовая система/системы автоматического доступа к информации, минуя оператора** (interactive voice response, IVR): Компьютер на основе телефонии, который используется для управления потоком телефонных звонков и обеспечения самообслуживания на основе голоса.

Примечания

1 Технология, позволяющая компьютеру определить голос и нажатия клавиш.

2 Интерактивные голосовые системы обрабатывают эффекты реального мира и неестественного содержания [эмоциональные голоса; различные шумы окружающей среды; запись свободной речи с очень распространенными словами (например, «да», «нет», цифры, ключевые слова)].

3 Интерактивные голосовые системы применяют технологию автоматического распознавания речи (ASR) для навигации пользователя вместе с технологией идентификации и верификации диктора (SIV) в различных приложениях для обеспечения безопасности. Например, при осуществлении финансовых операций с помощью телефона. Интерактивные голосовые системы могут объединять технологию ASR и технологию SIV для обнаружения аудио повторов и витальности пользователя с использованием информации, которая должна быть произнесена и генерируется в установленный момент времени.

3.8 **микрофон** (microphone): Подсистема сбора биометрических данных, которая преобразует акустическую волну, излучаемую голосом, в электрический сигнал.

3.9 **среднее звуковое поле** (mid-field): Область между ближним и дальним звуковыми полями, которая имеет сочетание характеристик ближнего и дальнего звуковых полей.

3.10 **ближнее звуковое поле/поле в зоне индукции** (near-field): Область, в которой прямая энергия на микрофоне от первичного источника больше, чем отраженная энергия от этого источника.

Примечание — В поле произвольных размеров ближнее звуковое поле — область, достаточно близкая к источнику, распределение угловой энергии которой зависит от расстояния от источника.

3.11 **коммутируемая телефонная сеть общего пользования** (public switched telephone network): Технология на базе канала, используемая для коммутации аналогового сигнала, обычно телефонные звонки через сеть от источника, такого как телефон, к пункту назначения, такому как другой телефон.

Примечание — Информация о канале, касающаяся места происхождения телефонного звонка, является полезной, так как исторически шум и другие характеристики канала различаются в разных странах. Появление и развитие VoIP и других цифровых телефонных сетей ослабило влияние национальных телекоммуникационных сетей, так как новые технологии не ограничены национальными границами.

3.12 **продолжительность представления** (representation duration): Продолжительность фрагмента речи одного голосового представления.

3.13 **частота дискретизации** (sampling rate): Число отсчетов в секунду (или в другую единицу измерения), полученных из непрерывного сигнала и используемых для получения дискретного сигнала.

Примечания

1 Единица частоты дискретизации — герц (Гц), для частоты дискретизации в секунду.

2 Частота дискретизации равна частоте квантования.

3 Частота дискретизации должна удовлетворять частотному критерию устойчивости Найквиста.

3.14 **сессия** (session): Единый процесс сбора биометрических данных, который выполняется в течение одного непрерывного периода времени.

Примечание — При сборе баз данных 2 сессии должны быть разнесены по времени от 3 до 6 недель, так чтобы могла быть собрана неактуальная речь. Однако в области биометрических систем сессия может интерпретироваться как время записи одного или нескольких образцов без покидания субъектом сбора биометрических данных места расположения устройства сбора биометрических данных, то есть сессия заканчивается при прохождении контрольного этапа/границы, в то время как в течение одной сессии могут происходить множественные отказы.

3.15 **отношение «сигнал-шум»** (signal-to-encoding noise ratio, SNR): Отношение полезного сигнала к шуму, возникающему в результате возможных электронных источников шума.

Примечания

1 Отношение «сигнал-шум» SNR, дБ, вычисляют по формуле

$$\mathrm{SNR} = 10 \lg \left(\frac{P_s}{P_n} \right), \qquad (1)$$

где P_s — средняя мощность сигнала, вычисляемая по формуле

$$P_s = \frac{1}{N}\sum_{i=1}^{N} s(i)^2, \quad (2)$$

P_n — средняя мощность шума, вычисляемая по формуле

$$P_n = \frac{1}{N}\sum_{i=1}^{N} n(i)^2, \quad (3)$$

где N — общее число цифровых отсчетов.

Пример — В кодово-импульсной модуляции (PCM) шум вызывается квантованием и приблизительно рассчитывается в соответствии с [1]:

$$SNR = 6B - 7{,}2,$$

где B — биты квантования.

3.16 **идентификация диктора** (speaker identification): Форма распознавания диктора, при которой голосовой образец сравнивается с набором биометрических (голосовых) контрольных шаблонов, принадлежащих разным людям, для определения одного, кто говорил.

3.17 **распознавание диктора** (speaker recognition): Процесс определения, были ли два речевых сегмента произведены голосовым аппаратом одного и того же субъекта биометрических данных.

3.18 **верификация диктора** (speaker verification): Форма распознавания диктора при которой принимается решение о принадлежности голосового образца индивиду, чья личность была заявлена.

Примечание — Верификация диктора в основном используется для ограничения доступа к информации, объектам или помещениям.

3.19 **идентификация и верификация диктора** (speaker identification and verification, SIV): Процесс автоматического распознавания индивидов с помощью голоса.

Примечание — Сам формат данных не зависит от типа приложения (активный/пассивный SIV).

3.20 **голос/речь** (voice/speech): Звук, производимый голосовым аппаратом во время разговора.

Примечания

1 Обычно определяется фонетистами как звук, который исходит от губ и ноздрей, который включает в себя «звонкие» и «глухие» звуки, образовываемые вибрацией голосовых складок и в сужениях вокального тракта и изменяемые динамическими характеристиками акустической передачи вокального тракта.

2 В настоящем стандарте термины «голос» (voice) и «речь» (speech) имеют равнозначное значение.

3.21 **ширина полосы частот речевого сигнала** (speech signal bandwidth): Диапазон речевых частот между верхней и нижней частотами среза, передаваемых или записываемых системой.

3.22 **распознавание речи/автоматическое распознавание речи** (speech recognition/automatic speech recognition): Преобразование речевого сигнала с помощью функциональной единицы в представление содержания речи.

Примечание — Распознаваемое содержимое может быть выражено в виде корректной последовательности слов или фонем.

3.23 **потоковые данные** (streaming data): Последовательность закодированных когерентных сигналов (пакетов данных), используемых для передачи или получения информации.

3.24 **текстонезависимое распознавание диктора/текстонезависимая система распознавания** (text-independent recognizer/text-independent recognition system): Распознавание речи, которое работает надежно вне зависимости от того, соответствует ли получаемый образец речи преопределенному сообщению.

3.25 **текстозависимое распознавание диктора/текстозависимая система распознавания** (text-dependent recognizer/text-dependent recognition system): Распознавание речи, которое работает надежно, только когда получаемый образец речи соответствует преопределенному сообщению.

3.26 **«запрос текста»/текстовая подсказка** (text prompted): Технология идентификации и верификации диктора (SIV), которая делает запрос субъекту сбору биометрических данных повторить последовательность, представленную системой SIV, или ответить на вопрос.

Примечания

1 Синонимом термина является «вопрос-ответ».

2 «Запрос текста» часто рассматривается как разновидность взаимодействия текстонезависимого распознавания диктора.

3.27 **фрагмент речи** (utterance): Последовательность единиц непрерывной речи (например, фонем, слогов, слов), ограниченных молчанием.

3.28 **голосовая связь по IP-протоколу** (voice over IP): Оцифрованная потоковая передача речи по каналам передачи данных в виде пакетов Интернет-протокола.

3.29 **голосовой ответ/голосовая подсказка** (voice prompt/voice-response prompt): Голосовое сообщение для инструктажа пользователя через диалог в системе голосового ответа.

3.30 **голосовое представление** (voice representation, VR): Один или более фрагментов речи.

3.31 **уровень громкости** (volume): Вычисление громкости входного сигнала (включая речь).

Примечания

1 Когда известно, уровень громкости выражается в терминах алгоритма P.56 Международного союза электросвязи [*2*].

2 Уровень громкости является фактором качества входных фрагментов речи.

4 Сокращения и обозначения

В настоящем стандарте применены следующие сокращения:

ADPCM — адаптивная дифференциальная импульсно-кодовая модуляция (adaptive differential pulse code modulation);

ASR — автоматическое распознавание речи (automatic speech recognition);

ЗОБД — запись для обмена биометрическими данными (biometric data interchange record);

CS-ACELP — алгебраическое линейное предсказание с кодовым возбуждением и сопряженной структурой (conjugate structure algebraic code excited linear prediction);

ILBC — кодек низкоскоростной передачи данных через интернет (internet low bitrate codec);

IP — интернет протокол (internet protocol);

IVR — интерактивная голосовая система/система автоматического доступа к информации, минуя оператора (interactive voice response);

PCM — кодово-импульсная модуляция (pulse code modulation);

SIV — идентификация и верификация диктора (speaker identification and verification);

TTS — преобразование текста в речь (text-to-speech);

URL — универсальный идентификатор ресурса (uniform resource locator);

VR — голосовое представление (voice representation);

VoIP — голосовая связь по IP-протоколу (voice over IP);

XML — расширяемый язык разметки (eXtensible Markup Language).

5 Соответствие

Запись биометрических данных соответствует настоящему стандарту в том случае, если она удовлетворяет всем обязательным требованиям, относящимся:

- к структуре, значениям данных, взаимосвязям между элементами в XML, как определено в *ГОСТ ISO/IEC 19794-1* и разделе 7;
- связи между значениями данных и соответствующими входными биометрическими данными, как определено в разделе 6.

Система, создающая записи биометрических данных, соответствует настоящему стандарту в том случае, если все производимые ею записи биометрических данных, содержащиеся в заявлении о соответствии реализации (ЗСР) данной системы, соответствуют настоящему стандарту (как определено выше). При этом записи биометрических данных, создаваемые системой, могут охватывать не все аспекты настоящего стандарта, а только те, которые должны поддерживаться системой согласно ЗСР.

Система, использующая записи биометрических данных, соответствует настоящему стандарту в том случае, если она способна прочитать и применить по назначению все записи биометрических данных, содержащиеся в ЗСР данной системы, которые соответствуют настоящему стандарту (как определено выше). При этом записи биометрических данных, используемые системой, могут охватывать не все аспекты настоящего стандарта, а только те, которые должны поддерживаться системой согласно ЗСР.

Примечание — Методология проведения испытаний на соответствие приведена в приложении А.

6 Процессы и идентификаторы

6.1 Процесс сбора биометрических данных

6.1.1 Общие положения

Настоящий раздел устанавливает требования к основному элементу взаимодействия SIV — процессу сбора биометрических данных, в соответствии с определением, приведенном в ГОСТ *ISO/IEC 2382-37*, а также к VR речи субъекта сбора биометрических данных, собранной во время указанного взаимодействия или сессии.

Во время процесса сбора биометрических данных голосовые звуки, возникающие не от регистрируемого диктора, могут быть непреднамеренно записаны перекрывающимися или неперекрывающими целевыми речевыми последовательностями; эту речь следует рассматривать как шум. Для обеспечения совместимости между различными алгоритмами SIV необходимо обеспечить совместимую структуру процесса сбора биометрических данных и описания звуковых сигналов.

6.1.2 Голосовой фрагмент речи (voice utterance)

Предполагается, что голосовой фрагмент речи исходит от одного диктора с целью распознавания (или используется для создания биометрического (голосового) контрольного шаблона для будущих сравнений). Если другие голоса от разных людей включены в голосовой фрагмент речи, такая информация должна рассматриваться как шум, который может повлиять на систему SIV. Целью настоящего стандарта не является определений граничных условий для голосового фрагмента речи, но голосовые фрагменты речи должны быть разделены следующим образом:

a) изменение или повторение подсказки; или

b) пауза гораздо большей продолжительности, чем частота следования слогов.

Минимальная и максимальная длина голосового фрагмента речи не установлена.

6.1.3 Структура процесса сбора биометрических данных

Процесс сбора биометрических данных при SIV является вербальным взаимодействием, которое может быть использовано для биометрической регистрации, верификации или идентификации субъекта сбора биометрических данных с помощью автоматизированной системы или другого человека. В общем случае процесс сбора биометрических данных может сопровождаться фоновым шумом, возможно, от самого источника информации.

Процесс сбора биометрических данных при SIV может быть активным (пользователь осведомлен о сборе биометрических данных) или пассивным (пользователь не осведомлен о сборе биометрических данных), с поведенческой адаптацией пользователей или без поведенческой адаптацией пользователей (лояльные/частые пользователи, намеренные адаптироваться для целей производительности), а также с кооперативными (лояльными) или некооперативными пользователями.

Процесс сбора биометрических данных SIV также известен как сессия. На примере, приведенном на рисунке 3, видно, что образец записи может включать как фрагмент речи всего вызова звонка регистрации, так и фрагменты речи однократных подсказок. Фрагментом речи является непрерывный поток вокализации от одного говорящего; он может содержать междуслоговое или междусловное молчание и ограничивается паузами. Паузы — это приостановка вокализации различимой продолжительности, которая длиннее междуслогового или междусловного молчания, то есть тишина, ощущаемая человеком.

Примечания

1 Для целей SIV могут быть использованы речевые и неречевые звуки, произносимые субъектом сбора биометрических данных. Обычно фрагмент речи демаркируется как непрерывная последовательность речи, однако приложения также могут использовать использование субфрагментов речи для VR.

2 Неречевые звуки не указывают на приостановку вокализации.

3 Фрагменты речи могут включать временные остановки в речи, такие как временные прерывания, поскольку человеческое восприятие может все еще «слышать» и не воспринимать приостановку вокализации.

Один процесс сбора биометрических данных обычно выполняется в течение одного непрерывного периода времени (или сессии) и содержит одно или несколько фрагментов речи, известных как VR. В основном, VR содержит голос одного диктора и может быть инициирован подсказкой субъекту сбора биометрических данных, запрашивающей ответ. На рисунке 1 показан простой процесс сбора биометрических данных при верификации с голосовым фрагментом речи, инициированным подсказкой интерактивной голосовой системы (IVR).

Запрос от IVR: Добро пожаловать в интерактивную голосовую систему банка. Пожалуйста, назовите свой номер счета

Фрагмент речи диктора А: 3571289999

Запрос от IVR: Спасибо. Пожалуйста, скажите парольную фразу

Фрагмент речи диктора А: 3456895612

Запрос от IVR: Спасибо

Рисунок 1 — Процесс сбора биометрических данных 1.
Основной сбор биометрических данных при базовой верификации диктора в технологии голосовых подсказок

Процесс сбора биометрических данных на рисунке 1 представляет собой одну сессию, которая может содержать один или два фрагмента речи диктора А. Варианты разбивки показаны на рисунке 2, когда могут быть выделены одно или два голосовых представления.

VR 1

Фрагмент речи 1: 3571289999

Фрагмент речи 2: 3456895612

a) Как одно представление

VR 1

Фрагмент речи 1: 3571289999

VR 2

Фрагмент речи 2: 3456895612

b) Как два представления

Рисунок 2 — Голосовые представления из голосовых фрагментов речи процесса сбора биометрических данных 1

Указанный пример относится к приложению контроля доступа. В этом примере первый голосовой фрагмент речи является заявленным указателем на биометрический (голосовой) контрольный шаблон субъектом биометрических данных «диктор А». Для определения указателя на биометрический (голосовой) контрольный шаблон путем извлечения содержимого из первого фрагмента речи может быть использована система автоматического распознавания речи (ASR), независимая от диктора. Второй фрагмент речи является текстозависимой кодовой фразой, требуемой для проверки утверждения с использованием сохраненной модели голоса указателя на биометрический (голосовой) контрольный шаблон. Не требуется изменять процесс сбора биометрических данных, приведенный на рисунке 1, при взаимодействии субъектов сбора биометрических данных с людьми (например, оператором центра обработки вызовов). Варианты процесса сбора биометрических данных 1 включают запрос или предоставление субъектам сбора биометрических данных ввода указателя на биометрический (голосовой) контрольный шаблон (учетный номер/идентификатор) вручную (например, с помощью сенсорной клавиатуры телефона). Запросы могут быть представлены в виде аудио с воспроизведением одного или нескольких звуковых файлов или генерации вывода TTS для внутренней строки. Запросы могут быть представлены в виде текстовых дисплеев (например, на КПК*, мобильных устройствах или смарт-устройствах).

С точки зрения субъектов сбора биометрических данных наиболее простой активный процесс сбора биометрических данных должен содержать только один фрагмент речи. В процессе сбора биометрических данных 1 это может быть выполнено двумя способами. В некоторых приложениях используется идентификатор вызывающего абонента и/или другие методы для неявного установления указателя на биометрический (голосовой) контрольный шаблон. Результатом является процесс сбора биометриче-

* Карманный персональный компьютер (personal digital assistant, PDA).

ских данных одного фрагмента речи (только фрагмент 2). Процесс сбора биометрических данных также может быть сокращен до одного фрагмента речи (только фрагмент 1) при использовании ASR. В этом случае IVR делает запрос «диктору А» назвать учетный номер/идентификатор. ASR декодирует цифры и использует их для получения указателя на биометрический (голосовой) контрольный шаблон. Затем эти же входные данные поступают в компонент SIV для верификации диктора.

Примечание — Как показано на рисунке 3, тот же процесс сбора биометрических данных и структура фрагментов речи могут быть использованы для биометрической регистрации.

Запрос от IVR: Добро пожаловать в интерактивную голосовую систему банка. Пожалуйста, назовите свой номер счета

Фрагмент речи 1 диктора А: 3571289999

Запрос от IVR: Спасибо. Теперь Вам необходимо произнести парольную фразу 4 раза. Пожалуйста, произнесите парольную фразу после звукового сигнала

Фрагмент речи 2 диктора А: 3456895612

Запрос от IVR: Пожалуйста, произнесите парольную фразу после звукового сигнала

Фрагмент речи 3 диктора А: 3456895612

Запрос от IVR: Пожалуйста, произнесите парольную фразу после звукового сигнала

Фрагмент речи 4 диктора А: 3456895612

Запрос от IVR: Пожалуйста, произнесите парольную фразу после звукового сигнала

Фрагмент речи 5 диктора А: 3456895612

Запрос от IVR: Спасибо. Регистрация прошла успешно

Рисунок 3 — Процесс сбора биометрических данных 2. Биометрическая регистрация

Процесс сбора биометрических данных на рисунке 3 содержит пять фрагментов речи «диктора А». Сначала диктор определяет указатель на биометрический (голосовой) контрольный шаблон, затем следуют четыре повторения кодовой фразы, запрашиваемые звуковым сигналом. Голосовые данные, полученные в этих фрагментах речи, составляют VR, которые являются основными элементами XML в блоке биометрических данных (ББД) голоса.

6.2 Идентификатор типа зарегистрированного формата

Запись в таблице 1 сделана регистрационным органом ЕСФОБД для идентичности формату записи данных голоса. Владельцем формата является ИСО/МЭК СТК1/ПК37, зарегистрированный идентификатор владельца формата — 31 (0x001F).

Таблица 1 — Идентификаторы типа формата

Идентификатора типа формата ББД ЕСФОБД	Короткое имя	Полный идентификатор объекта
257 (0x0101)	voice-data	{iso(1) registration-authority(1) cbeff(19785) biometricorganization(0) jtc1-sc37(257) bdbs(0) voice-data(31)}

7 Общий формат обмена данными голоса

7.1 Общие требования

Настоящий стандарт будет реализован только в формате XML. В данном разделе описан заголовок данных голоса в ЗОБД, который содержит информацию о условиях сбора данных VR и информацию о постобработке. Определение того, какие условия окружающей среды, методы сбора данных, а

также различная предварительная обработка (например, обнаружение/сегментация, предварительная фильтрация) сделаны на фрагментах речи данных голоса, составляющих процесс сбора данных, не является целью настоящего стандарта.

Структура XML-элементов представлена на рисунке 4. Формат записи включает:

- общий заголовок записи голоса, содержащий информацию о всей записи (7.3);
- элемент представления для каждого VR (7.4).

Каждый VR должен состоять:

- из заголовка VR, содержащего информацию о данных для одного представления;
- данных VR, где каждый заголовок содержит элемент для дополнительных данных изготовителя (см. таблицы 2 и 6).

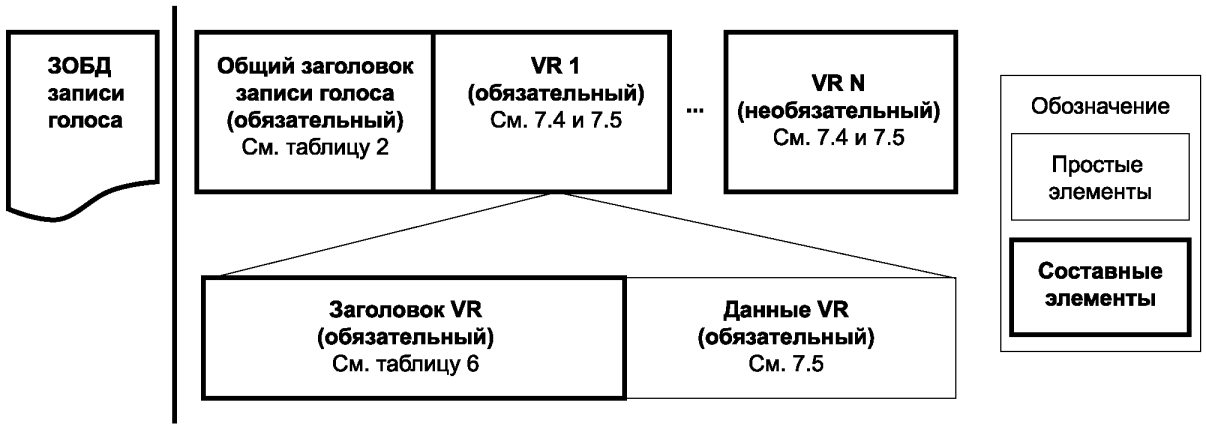

Рисунок 4 — Структура XML-элементов

7.2 Соглашения

Элементы могут быть простыми или составными. Составные элементы содержат другие элементы.

Элементы могут быть обязательными или необязательными. Необязательные составные элементы могут содержать как необязательные, так и обязательные элементы и характеристики.

Соглашение об именовании XML-элементов и характеристик, используемых в данном формате, должно состоять из прописных и строчных букв, таких как номер VR, без дефисов и пробелов. Соглашение о печати для допустимых строковых значений заключает каждое допустимое значение в кавычки.

7.3 Блок «Общий заголовок» записи голоса

7.3.1 Общие требования

Блок «Общий заголовок» записи голоса состоит из восьми элементов. Структура блока «Общий заголовок» приведена в таблице 2.

Таблица 2 — Структура блока «Общий заголовок» для записи голоса

Элемент	Пункт	Тип данных	Допустимое значение	Обязательное/ необязательное
Version (номер версии)	7.3.2	VersionType	См. *ГОСТ ISO/IEC 19794-1*	Обязательное
SessionID (идентификатор сессии)	7.3.3	string	Длина строки не ограничена	Необязательное
Channel (канал)	7.3.4	ChannelType	См. таблицу 3	Обязательное
CaptureDevice (устройство сбора биометрических данных)	7.3.5	CaptureDeviceModelID	См. *ГОСТ ISO/IEC 19794-1*	Необязательное

Окончание таблицы 2

Элемент	Пункт	Тип данных	Допустимое значение	Обязательное/ необязательное
Transducer (преобразователь)	7.3.6	TransducerType	См. таблицу 4	Необязательное
AudioMetaInformation (аудио-метаданные)	7.3.7	AudioMeta-InformationType	См. таблицу 5	Обязательное
CaptureProcessProtocol (протокол сбора биометрических данных)	7.3.8	CapturePro-cessProtocolType	Длина строки не ограничена	Необязательное
ExtendedVendorData (дополнительные данные разработчика)	7.3.9	VendorSpeci-ficDataType	См. *ГОСТ ISO/IEC 19794-1*, максимальное значение — 256	Необязательное

7.3.2 Элемент Version (номер версии)

Элемент Version содержит информацию о номере версии связанной сущности (например, номер версии ЕСФОБД, спецификации ведущей организации, спецификации формата данных).

7.3.3 Элемент SessionID (идентификатор сессии)

Элемент SessionID должен содержать уникальный идентификатор сессии, присвоенный голосовым приложением.

7.3.4 Элемент Channel (канал)

7.3.4.1 Общие требования

Элемент Channel описывает характеристики используемого канала, с помощью которого были получены данные. Элемент Channel является составным и состоит из четырех элементов. Структура элемента Channel приведена в таблице 3.

Таблица 3 — Структура элемента Channel

Элемент	Подпункт	Тип данных	Допустимые значения	Обязательное/ необязательное
Type (тип)	7.3.4.2	string	«Unknown», «Analog», «DigitalNonVoIP», «DigitalVoIP», «Mixed»	Обязательное
CutoffUpperFrequency (верхняя частота отсечения)	7.3.4.3	numeric	От 0 до 65535	Необязательное
CutoffLowerFrequency (нижняя частота отсечения)		numeric	От 0 до 65535	
Country of Origin (страна происхождения канала)	7.3.4.4	string	Трехсимвольная строка	Необязательное

7.3.4.2 Элемент Type (тип)

Элемент Type должен содержать информацию о виде канала, с помощью которого были собраны данные. Допустимые значения: «Analog», «DigitalNon-VoIP», «DigitalVoIP», «Mixed» и «Unknown». По умолчанию для элемента Type должно быть установлено значение «Unknown».

7.3.4.3 Элемент CutoffUpperFrequency (верхняя частота отсечения) и элемент CutoffLowerFrequency (нижняя частота отсечения)

Верхняя и нижняя частота отсечения должны быть представлены в виде целых чисел, которые наилучшим образом представляют частоты на верхней и нижней границах звуковой полосы, при которых энергия упала на 3 дБ ниже средней энергии полосы. Значения по умолчанию для элементов CutoffUpperFrequency и CutoffLowerFrequency не устанавливаются. Если значение частоты не известно, то должно быть установлено значение 0.

7.3.4.4 Элемент CountryofOrigin (Страна происхождения канала)

Элемент CountryofOrigin должен идентифицировать страну происхождения канала, если она известна.

Код страны происхождения должен быть представлен альфа-кодом, соответствующем двухбуквенному коду страны по *ГОСТ 7.67*, в котором определены три вида кодов стран: двухбуквенные, трехбуквенные и числовые.

7.3.5 Элемент CaptureDevice (устройство сбора биометрических данных)

Элемент CaptureDevice должен включать зарегистрированный идентификатор типа используемого устройства сбора биометрических данных (ЗОБД)*.

7.3.6 Элемент Transducer (преобразователь)

7.3.6.1 Общие требования

Элемент Transducer должен определять используемое устройство сбора биометрических данных. Комплексные системы сбора биометрических данных могут состоять из нескольких преобразователей, в этом случае требования данного пункта не обязательны. В подобных случаях должно быть использовано значение по умолчанию «unknown». Структура элемента Transducer приведена в таблице 4.

Примечание — Элемент Transducer предназначен, главным образом, для поддержки научно-исследовательских и опытно-конструкторских работ и компонентов, требующих регистрации устройства.

Таблица 4 — Структура элемента Transducer

Элемент	Подпункт	Тип данных	Допустимые значения	Обязательное/ необязательное
CaptureTechnologyID (тип устройства сбора биометрических данных)	7.3.6.2	string	«Telephone» (телефон), «Microphone» (микрофон), «Handheld» (коммуникатор), «Mobile phone» (мобильный телефон), «Stethoscope» (стетоскоп), «Other» (другое), «Unknown» (неизвестен)	Необязательное
MicrophoneType (тип микрофона)	7.3.6.3	string	«Carbon» (карбоновый), «Electret» (электретный), «Other» (другое), «Unknown» (неизвестен)	Необязательное
Manufacturer (изготовитель)	7.3.6.4	string	Длина строки не ограничена	Необязательное
Model (модель)	7.3.6.5	string	Длина строки не ограничена	Необязательное
MicCutoffUpper (верхняя граница частоты микрофона)	7.3.6.6	numeric	От 0 до 65535	Необязательное
MicCutoffLower (нижняя граница частоты микрофона)		numeric	От 0 до 65535	Необязательное
DeviceInfo (информация об устройстве)	7.3.6.7	string	Длина строки не ограничена	Необязательное

* Деятельность по присвоению уникальных идентификаторов биометрическим организациям, осуществляющим деятельность в Российской Федерации, и биометрическим продуктам, разрабатываемым и/или серийно выпускаемым, и/или реализуемым в Российской Федерации, а также ведение соответствующих реестров осуществляет Некоммерческое партнерство «Русское биометрическое общество», официально зарегистрированное Международной ассоциацией биометрии и идентификации (МАБИ) [The International Biometrics & Identification Association (IBIA)] в качестве ведущей организации ЕСФОБД.

7.3.6.2 Элемент CaptureTechnologyID (тип устройства сбора биометрических данных)

Элемент CaptureTechnologyID должен содержать информацию о типе используемого устройства сбора биометрических данных, если она известна. По умолчанию для элемента CaptureTechnologyID должно быть установлено значение «Telephone».

7.3.6.3 Элемент MicrophoneType (тип микрофона)

Элемент MicrophoneType должен содержать информацию о типе используемого микрофона в устройстве сбора биометрических данных, если она известна. Допустимыми значениями являются «Carbon», «Electret», «Other» и «Unknown».

7.3.6.4 Элемент Manufacturer (изготовитель)

Элемент Manufacturer должен содержать идентификатор изготовителя устройства сбора биометрических данных* в формате строки.

7.3.6.5 Элемент Model (модель)

Элемент Model должен содержать идентификатор модели устройства сбора биометрических данных в формате строки.

7.3.6.6 Элемент MicCutoffUpper (верхняя граница частоты микрофона) и элемент MicCutoffLower (нижняя граница частоты микрофона)

Необязательные значения верхней и нижней частот микрофона, на которых мощность преобразования энергии микрофона упала на 3 дБ ниже средней энергии полосы, должны быть целыми числами. Значение по умолчанию не устанавливаются, при неизвестной информации должно быть установлено значение 0.

7.3.6.7 Элемент DeviceInfo (информация об устройстве)

Элемент DeviceInfo содержит дополнительную информацию об устройстве, за исключением процесса сбора биометрических данных или данных субъекта сбора биометрических данных. Размер элемента DeviceInfo должен быть ограничен данными, которые может различать и использовать механизм или приложение получателя SIV.

7.3.7 Элемент AudioMetaInformation (аудиометаданные)

7.3.7.1 Общие требования

В данном пункте приведены технические характеристики процесса сигнала для сбора всех VR в записи голоса.

Структура элемента AudioMetaInformation приведена в таблице 5.

Таблица 5 — Структура элемента AudioMetaInformation

Элемент	Пункт	Тип данных	Допустимые значения	Обязательное/ необязательное
ChannelCount (число каналов)	7.3.7.2	numeric	От 1 до 15	Обязательное
SamplingRate (частота дискретизации)	7.3.7.3	numeric	От 0 до 128000	Обязательное
BitsPerSample (разрядность)	7.3.7.4	numeric	От 0 до 255	Обязательное
AudioDuration (длительность аудиозаписи)	7.3.7.5	numeric	Встроенный тип	Обязательное

7.3.7.2 Элемент ChannelCount (число каналов)

Обязательный элемент ChannelCount должен содержать информацию о числе каналов во входном потоке в виде целого числа. По умолчанию должно быть установлено значение 1.

7.3.7.3 Элемент SamplingRate (частота дискретизации)

Обязательный элемент SamplingRate должен содержать информацию о количестве отсчетов сигнала в секунду, с которым дискретизирован входной звуковой поток. Значения элемента SamplingRate должны быть представлены в виде целого числа.

* Деятельность по присвоению уникальных идентификаторов биометрическим организациям, осуществляющим деятельность в Российской Федерации, и биометрическим продуктам, разрабатываемым и/или серийно выпускаемым, и/или реализуемым в Российской Федерации, а также ведение соответствующих реестров осуществляет Некоммерческое партнерство «Русское биометрическое общество», официально зарегистрированное Международной ассоциацией биометрии и идентификации (МАБИ) [The International Biometrics & Identification Association (IBIA)] в качестве ведущей организации ЕСФОБД.

7.3.7.4 Элемент BitsPerSample (разрядность)

Обязательный элемент BitsPerSample должен содержать информацию о битовой глубине одного отсчета звукового сигнала. Значения элемента BitsPerSample должны быть представлены в виде целого числа. Если используется формат с переменной битовой глубиной, то для элемента BitsPerSample должно быть установлено значение 0.

7.3.7.5 Элемент AudioDuration (длительность аудиозаписи)

Элемент AudioDuration должен содержать информацию о длительности фрагмента речи в миллисекундах. Значения элемента AudioDuration должны быть представлены в виде целого числа.

7.3.8 Элемент CaptureProcessProtocol (протокол процесса сбора биометрических данных)

Элемент CaptureProcessProtocol является необязательным. Используется для описания дополнительной информации о процессе сбора биометрических данных, за исключением информации о субъекте сбора биометрических данных или устройстве сбора биометрических данных. Размер элемента должен быть ограничен данными, которые может различать и использовать механизм или приложение получателя SIV.

7.3.9 Элемент ExtendedVendorData (дополнительные данные разработчика)

Элемент ExtendedVendorData является необязательным и используется в случае необходимости включения не стандартизированных данных, проприетарных разработчику/программному продукту.

7.4 Блок «Заголовок представления»

7.4.1 Общие требования

Заголовок представления является дочерним блоком элемента процесса сбора биометрических данных, который содержит элементы и поля, которые могут меняться в процессе сбора биометрических данных. Для каждого процесса сбора биометрических данных должно быть не менее одного представления. Структура блока «Заголовок представления» приведена в таблице 6.

Примечание — В элементах VR не рассматривается информация об произносимом тексте, языке, диалектах или поле субъекта сбора биометрических данных. Если эта или другая информация может способствовать процессу распознавания, аналитики могут использовать программное обеспечение ASR, автоматическое распознавание языка (ALR*) или автоматическое определение пола (AGD**).

Таблица 6 — Структура блока «Заголовок представления»

Элемент	Пункт	Тип данных	Допустимые значения	Обязательное/ необязательное
DateAndTime (дата и время регистрации)	7.4.2	DateAndTimeType	См. таблицу 7	Необязательное
AudioContent (аудиоматериал)	7.4.3	AudioContentType	См. таблицу 8	Обязательное
Quality (качество)	7.4.4	VRQualityType	См. таблицу 10	Необязательное
SignalEnhancement (усиление уровня сигнала)	7.4.5	string	Длина строки не ограничена	Необязательное
ExtendedVendorData (дополнительные данные разработчика)	7.4.6	VendorSpecific DataType	См. *ГОСТ ISO/IEC 19794-1*, максимальное значение — 256	Необязательное

7.4.2 Элемент DateAndTime (дата и время регистрации)

7.4.2.1 Общие требования

Элемент DateAndTime должен содержать информацию о дате и времени начала и завершения записи голоса. Время начала записи голоса считается «временем сбора данных». Спецификация времени должна соответствовать WC3 — XML 1.0. Структура данного элемента приведена в таблице 7.

---

* Automatic Language Recognition.

** Automatic Gender Detection.

Таблица 7 — Структура элемента DateAndTime (дата и время)

Элемент	Подпункт	Тип данных	Допустимые значения	Обязательное/ необязательное
Start (дата и время начала записи голоса)	7.4.2.2	dateTime	См. *ГОСТ ISO/ IEC 19794-1*	Необязательное
End (дата и время завершения записи голоса)	7.4.2.3	dateTime	См. *ГОСТ ISO/ IEC 19794-1*	Необязательное
VoiceStartTime (время начала голосового фрагмента)	7.4.2.4	dateTime	Встроенный тип	Необязательное
VoiceEndTime (время завершения голосового фрагмента)	7.4.2.5	dateTime	Встроенный тип	Необязательное
VoiceElapsedTime (длительность звучащей речи)	7.4.2.6	time	Встроенный тип	Необязательное

7.4.2.2 Элемент Start (дата и время начала записи голоса)

Элемент Start должен содержать дату и время, когда была начата запись голоса, в соответствии с ГОСТ ИСО 8601. Время начала считается временем сбора записи голоса согласно 7.5.

7.4.2.3 Элемент End (дата и время завершения записи голоса)

Элемент End должен содержать дата и время, когда была завершена запись голоса. Из-за возможного использования программного обеспечения для обнаружения активности длина аудиоданных в представлении может быть короче разницы между временем начала и окончания.

7.4.2.4 Элемент VoiceStartTime (время начала голосового фрагмента)

Элемент VoiceStartTime должен содержать информацию о времени начала голосового фрагмента на записи.

7.4.2.5 Элемент VoiceEndTime (время завершения голосового фрагмента)

Элемент VoiceEndTime должен содержать информацию о времени завершения голосового фрагмента на записи.

7.4.2.6 Элемент VoiceElapsedTime (длительность звучащей речи)

Элемент VoiceElapsedTime должен содержать информацию о длительности звучащей речи. В случае спонтанной/свободной или разговорной речи или характеристиками являются время начала и окончания разговора.

7.4.3 Элемент AudioContent (аудиоматериал)

7.4.3.1 Общие требования

Элемент AudioContent содержит подробную информацию о аудиоматериале, содержащемся на записи, включая обязательную информацию о формате аудиокодирования, продолжительности записи в секундах и информации о типе разговора на записи (если известно), уровне громкости и оценке SNR. Элемент AudioContent является составным. Структура элемента AudioContent приведена в таблице 8.

Примечание — Вербализованный текст субъекта сбора биометрических данных не включается в формат данных, поскольку он может содержать конфиденциальные данные. Таким образом, не учитываются ни результаты анализа систем ASR, ни фактический вербализованный текст. Однако система ASR может быть использована в качестве подсистемы для SIV.

Таблица 8 — Структура элемента AudioContent

Элемент	Подпункт	Тип данных	Допустимые значения	Обязательное/ необязательное
AudioEncoding (формат аудиокодирования)	7.4.3.2	string	«Linear PCM», «Mu-Law», «A-Law», «Non-streaming OGG Vorbis», «Speex», «ADPCM», «CS-ACELP», «PCM», «AMR», «ILBC», «MPEG», «AC3», «AAC», «AMR», «APE», «FLAC», «MMF», «M4A», «MP2», «MP3», «MP4», «RA», «Full-HD Voice», «other», «unknown»	Обязательное
Duration (продолжительность)	7.4.3.3	numeric	Встроенный тип	Обязательное
Conversation (тип разговора)	7.4.3.4	complex	«Unknown», «Spontaneous/Free», «Reading», «Prompt», «Conversational», «Other», Если «Prompt», см. таблицу 9	Необязательное
DominantLanguage (доминирующий язык)	7.4.3.5	string	Трехсимвольная строка	Необязательное

7.4.3.2 Элемент AudioEncoding (формат аудиокодирования)

Каждое VR должно содержать информацию о формате аудиокодирования для хранимых данных из множества допустимых значений.

Допустимые значения элемента AudioEncoding: Linear PCM (линейная PCM), Mu-Law, A-Law, non-streaming OGG Vorbis (непотоковый OGG Vorbis), Speex, ADPCM, CS-ACELP, PCM, AMR, ILBC, MPEG, AC3, AAC, AMR, APE, FLAC, MMF, M4A, MP2, MP3, MP4, RA, Full-HD Voice, «other» (другой) или «unknown» (неизвестно).

7.4.3.3 Элемент Duration (продолжительность)

Элемент Duration содержит информацию об общем времени представления в виде целого числа в миллисекундах. Конечный результат должен позволять обратно вычислять частоту дискретизации.

7.4.3.4 Элемент Conversation (тип разговора)

Элемент Conversation содержит информацию о типе разговора. Допустимыми значениями являются «Unknown» (неизвестно), «Spontaneous/Free» (спонтанный/свободный), «Reading» (чтение), «Prompt» (подсказка), «Conversational» (разговорный) и «Other» (другой).

Если указано значение «Prompt», то элемент становится составным, структура в этом случае приведена в таблице 9. Составная структура включает в себя допустимые значения элементов StringPromptContent (содержимое запроса строк) и/или AudioPromptContent (содержимое звуковых подсказок). Элемент StringPromptContent должен содержать текст подсказки, если он известен. Если была использована звуковая подсказка, и доступен URL-адрес, содержащий эту звуковую подсказку, то элемент AudioPromptContent должен указывать URL-адрес звуковой подсказки или зависимого идентификатора, если это известно. Возможно, что будут доступны и файл, и транскрипция звуковой подсказки. В этом случае оба поля имеют контент.

Таблица 9 — Структура элемента ConversationRoot, если элемент Conversation имеет значение «Prompt» (подсказка)

Элемент	Подпункт	Тип данных	Допустимые значения	Обязательное/ необязательное
StringPromptContent (содержимое запроса строк)	7.4.3.4	string	Длина строки не ограничена	Необязательное
AudioPromptContent (содержимое звуковых подсказок)		URL	Длина URL не ограничена	Необязательное

7.4.3.5 Элемент DominantLanguage (доминирующий язык)

Элемент DominantLanguage содержит информацию о доминирующем языке в VR. Идентификатор доминирующего языка должен быть представлен в виде строки в соответствии с кодами [*3*] и [*4*].

7.4.4 Элемент QualityInformation (информация о качестве)

7.4.4.1 Общие требования

Данный элемент содержит подробную информацию о качестве аудиозаписи VR, включая обязательную информацию о условиях сбора биометрических данных и о проведенной постобработке сигнала. Структура элемента QualityInformation приведена в таблице 10.

Таблица 10 — Структура элемента QualityInformation (Информация о качестве)

Элемент	Подпункт	Тип данных	Допустимые значения	Обязательное/ необязательное
Quality (качество)	7.4.4.2	QualityType	См. *ГОСТ ISO/IEC 19794-1*	Обязательное
Field (звуковое поле)	7.4.4.3	string	«Near-field» (ближнее звуковое поле), «Mid-field» (среднее звуковое поле), «Far-field» (дальнее звуковое поле), «Other» (другое), «Unknown» (неизвестно)	Необязательное
MicrophoneDistance (расстояние до микрофона)	7.4.4.4	string	«Close» (близкое), «Mid-range» (среднее), «Far» (дальнее)	Необязательное
Volume (уровень громкости)	7.4.4.5	float	Встроенный тип	Необязательное
SNR (отношение «сигнал-шум»)	7.4.4.6	float	Встроенный тип	Необязательное

7.4.4.2 Элемент Quality (качество)

Данный элемент содержит информацию о качестве биометрических данных и о алгоритме оценки качества, который был использован.

Если оценка качества не проводилась, то элемент Quality отсутствует. Минимальное значение элемента Quality — 0 (minOccurs=0).

7.4.4.3 Элемент Field (звуковое поле)

Элемент Field должен содержать информацию о звуковом поле. Допустимые значения:

- Near-field (ближнее звуковое поле),
- Mid-field (среднее звуковое поле),
- Far-field (дальнее звуковое поле),
- Other (другое),
- Unknown (неизвестно).

По умолчанию должно быть установлено значение «Near-field».

Примечание — «Ближнее звуковое поле» и другие указанные термины являются общими характеристиками звукового поля (см. раздел 3). «Среднее звуковое поле» используется в беспроводных наушниках и других

микрофонах телефонов. Указанная классификация предпочтительнее указания фактического расстояния в численном виде, поскольку, как правило, невозможно получить такую детальную информацию.

7.4.4.4 Элемент MicrophoneDistance (расстояние до микрофона)

Элемент MicrophoneDistance должен содержать информацию о расстоянии до микрофона.

Допустимые значения:

- «Close» (близкое): случай использования гарнитуры, то есть оборудования со встроенным динамиком/микрофоном, в котором голос становится слышимым при расстоянии между гарнитурой и ухом от 0,05 до 0,15 м.
- «Mid-range» (среднее): случай использования микрофона, телефона (на громкой связи) или планшета, то есть использования микрофона не в пределах диапазона «Near-field» (ближнее звуковое поле), а с руки говорящего на расстоянии от 0,1 до 0,5 м.
- «Far» (дальнее): случай использования микрофона за пределами диапазона руки диктора (более 0,5 м).

7.4.4.5 Элемент Volume (уровень громкости)

Если значение известно, то оно должно быть выражено в терминах алгоритма P.56 Международного союза электросвязи [2]. Если значение неизвестно, то должно быть установлено значение «Unknown» (неизвестно).

7.4.4.6 Элемент SNR (отношение «сигнал-шум»)

Элемент SNR является необязательным. Используется, если шум кодирования был учтен при расчете отношения «сигнал-шум» (SNR) для каждого VR.

7.4.5 Элемент SignalEnhancement (усиление уровня сигнала)

Усиление уровня сигнала является предварительной обработкой речевого сигнала. Элемент SignalEnhancement содержит информацию о проведенной предварительной обработке сигнала. Усиление уровня сигнала может увеличить или уменьшить вероятность распознавания. Должно быть указано, какой тип усиления уровня сигнала, если применялся, был применен к речевому сигналу.

Типовые примеры предварительной обработки, направленные на усиление речевого сигнала:

a) подавление шума;
b) подавление отраженных сигналов;
c) эхоподавление;
d) активная защита от помех;
e) выделение речи (включая результат обработки микрофона);
f) автоматическая регулировка усиления (AGC*);
g) частотная коррекция, фильтрация, предварительное выделение;
h) дереверберация (удаление реверберации);
i) разметка завершения;
j) удаление тишины;
k) другие варианты предварительной обработки.

7.4.6 Дополнительные данные разработчика

Дополнительные данные разработчика добавляются к основной записи в случае, когда необходимо добавить нестандартизированные проприетарные данные разработчика/продукта.

7.5 Данные голосового представления

Необходимо указать URL-адрес расположения данных VR или большой двоичный объект голосового фрагмента речи в формате кодирования Base64.

7.6 XML-схема данных записи голоса

Электронная версия схемы, доступна по ссылке:

gost.e-ecolog.ru .xsd

Данная схема должна использоваться для валидации (подтверждения) голосовых записей XML, закодированных в формате XML.

Пользователю разрешается использовать схему в ее исходном формате без каких-либо изменений для целей, указанных в настоящем стандарте.

---

* AGC — Automatic gain control.

7.7 Пример

```
<?xml version="1.0" encoding="UTF-8"?>
<vdi:VoiceRecord xmlns:cmn="gost.e-ecolog.ru /ed-
2/amd/2" xmlns:vdi="gost.e-ecolog.ru
xmlns:xsi="http:// gost.e-ecolog.ru "
xsi:schemaLocation="gost.e-ecolog.ru org/iso-iec/19794/-13/ed-1 19794-
13_ed1.xsd gost.e-ecolog.ru org/iso-iec/19794/-
1/ed-2/amd/2 19794-1_ed2_amd2.xsd " cmn:SchemaVersion="1.0">
  <vdi:VoiceRecordGeneralHeader>
    <vdi:Version>
      <cmn:Major>1</cmn:Major>
      <cmn:Minor>0</cmn:Minor>
    </vdi:Version>
    <vdi:Channel>
      <vdi:Type>Unknown</vdi:Type>
    </vdi:Channel>
    <vdi:AudioMetaInformation>
      <vdi:ChannelCount>1</vdi:ChannelCount>
      <vdi:SamplingRate>8000</vdi:SamplingRate>
      <vdi:BitsPerSample>16</vdi:BitsPerSample>
      <vdi:AudioDuration>0</vdi:AudioDuration>
    </vdi:AudioMetaInformation>
  </vdi:VoiceRecordGeneralHeader>
  <vdi:VoiceRepresentation>
    <vdi:VoiceRepresentationHeader>
      <vdi:AudioContent>
        <vdi:AudioEncoding>Linear PCM</vdi:AudioEncoding>
<vdi:Duration>0</vdi:Duration>
      </vdi:AudioContent>
    </vdi:VoiceRepresentationHeader>
    <vdi:VoiceRepresentationData>
  <vdi:BLOB>UklGRiQAAABXQVZFZm10IBAAAAABAAEAQB8AAIA+AAACABAAZGF0YQAAAAA=</
vdi:BLOB>
  </ vdi:VoiceRepresentationData>
  </vdi:VoiceRepresentation>
  </vdi:VoiceRecord>
```

Приложение А
(обязательное)

Методология испытаний на соответствие

А.1 Общие положения

Настоящий стандарт определяет формат обмена биометрическими данными для хранения, записи и передачи одного или нескольких представлений VR. Каждое представление сопровождается метаданными, содержащимися в заголовке записи. Настоящее приложение определяет порядок проведения испытаний для проверки корректности записи.

Цель настоящего стандарта не может быть в полной мере достигнута до тех пор, пока биометрические продукты не пройдут испытания на соответствие требованиям настоящего стандарта. Соответствие реализации требованиям является необходимым условием для достижения совместимости между реализациями, поэтому существует необходимость в стандартизированной методологии испытаний на соответствие, тестовых утверждениях и методиках испытаний применительно к конкретным биометрическим модальностям, рассмотренным в настоящем стандарте. Тестовыми утверждениями проверяются наиболее важные требования настоящего стандарта, и соответствие результатов, полученных с помощью комплектов для проведения испытаний на соответствие, будет показывать степень соответствия реализаций настоящему стандарту. Это является причиной разработки данной методологии испытаний на соответствие.

Настоящее приложение является обязательным и предназначено для определения элементов методологии испытаний на соответствие, тестовых утверждений и методик испытаний применительно к настоящему стандарту. Для текущей версии настоящего стандарта содержание настоящего приложения будет доступно в виде отдельного документа (изменения), дополняющего настоящий стандарт.

А.2 Испытание на соответствие

Каждая часть серии *ГОСТ Р ИСО/МЭК 19794*, которая определяет XML-схему данных, должна требовать проверки соответствия с точки зрения строгой проверки определения схемы XML.

Дополнительно каждая часть серии *ГОСТ Р ИСО/МЭК 19794* должна содержать таблицу с тестовыми утверждениями для будущих требований, которые явно не включены в процесс подтверждения схемы XML.

Каждая часть серии *ГОСТ Р ИСО/МЭК 19794* может содержать обязательное приложение, которое определяет язык преобразования XML-документов (XSLT*) для использования при проведении испытаний уровня 2 корректных документов XML на соответствие данной схеме XML.

* Extensible Stylesheet Language Transformations.

Приложение ДА
(справочное)

Сведения о соответствии ссылочных национальных и межгосударственных стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте

Таблица ДА.1

Обозначение ссылочного национального, межгосударственного стандарта	Степень соответствия	Обозначение и наименование ссылочного международного стандарта
ГОСТ 7.67—2003 (ИСО 3166-1:1997)	MOD	ISO 3166-1:1997 «Коды для представления названий стран и единиц их административно-территориального деления. Часть 1. Коды стран»
ГОСТ ИСО 8601—2001	IDT	ISO 8601:2000 «Элементы данных и форматы для обмена информацией. Обмен информацией. Представление дат и времени»
ГОСТ ISO/IEC 2382-37—2016	IDT	ISO/IEC 2382-37:2012 «Информационные технологии. Словарь. Часть 37. Биометрия»
ГОСТ ISO/IEC 19794-1—2015	IDT	ISO/IEC 19794-1:2011 «Информационные технологии. Форматы обмена биометрическими данными. Часть 1. Структура»
ГОСТ Р ИСО/МЭК 19794 (все части)	IDT	ISO/IEC 19794 (все части) «Информационные технологии. Форматы обмена биометрическими данными»

Примечание — В настоящей таблице использованы следующие условные обозначения степени соответствия стандартов:

- IDT — идентичные стандарты;
- MOD — модифицированные стандарты.

Приложение ДБ
(справочное)

Сопоставление структуры настоящего стандарта со структурой примененного в нем международного стандарта

Таблица ДБ.1

Структура настоящего стандарта	Структура международного стандарта ИСО/МЭК 19794-13:2018
Приложение ДА Сведения о соответствии ссылочных национальных и межгосударственного стандартов международным стандартам, использованным в качестве ссылочных в примененном международном стандарте	—
Приложение ДБ Сопоставление структуры настоящего стандарта со структурой примененного в нем международного стандарта	—
Библиография	Библиография
Примечание — Сопоставление структуры стандартов приведено, начиная с приложения ДА, так как предыдущие разделы стандартов идентичны.	

Библиография

[1] *Furui «Digital Speech Processing, Synthesis and Recognition» (Dekker, 1989)*

[2] ITU-T P.56:1993, Objective measurement of active speech level. Geneva: International Telecommunication Union — Telecommunication Standardization Sector

[3] IETF RFC 5646: 2009, Tags for the Identification of Languages. Edited by Phillips, A. (gost.e-ecolog.ru txt)

[4] ИСО 639-3:2007 Коды для представления названий языков. Часть 3. Код альфа-3 для всестороннего охвата языков (Codes for the representation of names of languages — Part 3: Alpha-3 code for comprehensive coverage of languages)

УДК 004.93'1:006.89:006.354 ОКС 35.040

Ключевые слова: форматы обмена биометрическими данными, голос, биометрия

БЗ 12—2019/85

Редактор *Н.В. Таланова*
Технический редактор *И.Е. Черепкова*
Корректор *И.А. Королева*
Компьютерная верстка *Е.А. Кондрашовой*

Сдано в набор 21.11.2019. Подписано в печать 16.12.2019. Формат 60×84⅛. Гарнитура Ариал.
Усл. печ. л. 3,26. Уч.-изд. л. 2,77.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении во ФГУП «СТАНДАРТИНФОРМ»
для комплектования Федерального информационного фонда стандартов,

117418 Москва, Нахимовский пр-т, д. 31, к. 2.

gost.e-ecolog.ru gost.e-ecolog.ru

gost.e-ecolog.ru