ФЕДЕРАЛЬНОЕ АГЕНТСТВО

ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

ГОСТ Р 52873— 2021

КОМПЛЕКСЫ ЭЛЕКТРОННО-ЦИФРОВЫЕ ИНТЕРАКТИВНЫЕ, РАССЧИТАННЫЕ НА ИСПОЛЬЗОВАНИЕ ЛЮДЬМИ С ИНВАЛИДНОСТЬЮ И ДРУГИМИ ОГРАНИЧЕНИЯМИ ЖИЗНЕДЕЯТЕЛЬНОСТИ. СИСТЕМЫ ВЫВОДА ТЕКСТОВОЙ ИНФОРМАЦИИ В ФОРМЕ СИНТЕЗИРОВАННОЙ РЕЧИ

Технические требования

Издание официальное

Москва
Российский институт стандартизации
2021

gost.e-ecolog.ru

Предисловие

1 РАЗРАБОТАН Федеральным государственным унитарным предприятием «Российский научно-технический центр информации по стандартизации, метрологии и оценке соответствия» (ФГУП «СТАНДАРТИНФОРМ»)

2 ВНЕСЕН Техническим комитетом по стандартизации ТК 381 «Технические средства и услуги для инвалидов и других маломобильных групп населения»

3 УТВЕРЖДЕН И ВВЕДЕН В ДЕЙСТВИЕ Приказом Федерального агентства по техническому регулированию и метрологии от 25 августа 2021 г. № 820-ст

4 ВЗАМЕН ГОСТ Р 52873—2017

*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

© Оформление. ФГБУ «РСТ», 2021

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 1
3 Термины, определения и сокращения 2
4 Технические требования 4
Библиография 11

Введение

По важности и объему воспринимаемой информации слух является второй, после зрения, системой восприятия человека, а в случае невозможности по тем или иным причинам воспользоваться зрением, роль слуха и звуковой информации резко возрастает и при взаимодействии между людьми, и при взаимодействии человека с информационными системами.

Все более широкое распространение получают электронно-цифровые интерактивные комплексы, предоставляющие информацию пользователю в виде синтезированной речи. Примерами таких комплексов являются специальное рабочее место для инвалида по зрению (ГОСТ Р 51645); персональный компьютер, смартфон или иное мобильное устройство, оснащенное программой экранного доступа и используемое людьми с нарушениями зрения в повседневной жизни; банковский терминал, оснащенный, помимо дисплея, еще и возможностью подключить аудиовоспроизводящее устройство для прослушивания информации о выполняемой транзакции; любой персональный компьютер или аналогичное по своим возможностям устройство, на котором установлены игровые, обучающие, информационные, навигационные и прочие приложения, использующие синтезированную речь для взаимодействия с пользователем; портативные и стационарные медицинские диагностические приборы с возможностью речевого вывода и многое другое.

Для целей настоящего стандарта все многообразие подобных аппаратно-программных комплексов было сведено к одной обобщенной схеме, в которую вошли аппаратно-техническая платформа, включающая все необходимые компоненты для функционирования операционной среды и прикладных программ; операционная среда (или операционная система), которая может содержать, а может и не содержать вспомогательные программные интерфейсы для поддержки синтеза речи по тексту; прикладное программное обеспечение, использующее синтез речи для взаимодействия с пользователем; система вывода текстовой информации в форме синтезированной речи или (для целей настоящего стандарта) система речевого вывода, которая предоставляет соответствующий прикладной программный интерфейс для создания синтезированной речи и вывода ее через аудиовоспроизводящее устройство, поддерживаемое аппаратно-технической платформой.

Систему речевого вывода, обозначенную в данной схеме, в свою очередь, также можно представить в виде набора взаимосвязанных функциональных модулей, одним из которых является синтезатор речи (иногда называемый «речевым движком»). Несмотря на ключевое значение синтезатора речи во всей системе речевого вывода, в настоящем стандарте особого внимания именно синтезатору речи не уделяется. Система речевого вывода рассматривается как единое целое, предоставляющее пользователю возможность получать текстовую информацию в виде звучащей синтезированной речи, параметры которой должны соответствовать настоящему стандарту. Внутренняя архитектура системы вывода текстовой информации в форме синтезированной речи, состав и структура ее модулей, принципы функционирования и методы, используемые для анализа входных текстовых данных, для синтеза звуковых данных и тому подобное, оставлено на усмотрение разработчика. Кроме того, разнообразие современных операционных платформ и их возможностей в части поддержки систем синтеза речи по тексту не позволяет ориентироваться лишь на одну из них.

Именно такой интегрированный подход и обеспечил ключевую роль понятия «система речевого вывода» вместо понятия «синтезатор речи» в настоящем стандарте.

Несмотря на то, что настоящий стандарт в первую очередь ориентирован на специально оборудованное компьютерное рабочее место, используемое людьми с инвалидностью по зрению, в том числе и как средство реабилитации, область его применения охватывает также любое прикладное программное обеспечение, взаимодействующее с пользователем с помощью системы вывода текстовой информации в форме синтезированной речи, причем речевой вывод может сочетаться с визуальным, тактильным и т. п. Примерами такого программного обеспечения могут служить программы экранного доступа, используемые людьми с нарушениями зрения на рабочем месте, в бытовых условиях и иных местах пребывания; игровые, обучающие, информационные, навигационные и прочие программы для широкого круга пользователей.

Любое торговое наименование, использованное в настоящем стандарте, является информацией, приводимой для удобства пользователей, и не является свидетельством в пользу того или иного товара.

Стандарт разработан авторским коллективом в следующем составе: юриста, сертифицированного тренера и консультанта по адаптивным информационным технологиям А.В. Зеленова, сооснователя и администратора Портала Tiflocomp (gost.e-ecolog.ru зработчика адаптивных решений А.Н. Камынина, начальника отдела социокультурных проектов и программ ГМКЦ «Интеграция» им. Н.А. Островского, представителя Российской Федерации в глобальной инициативе за инклюзивные ИКТ (G3ICT/ Smart cities for all) А.Д. Попко.

НАЦИОНАЛЬНЫЙ СТАНДАРТ РОССИЙСКОЙ ФЕДЕРАЦИИ

КОМПЛЕКСЫ ЭЛЕКТРОННО-ЦИФРОВЫЕ ИНТЕРАКТИВНЫЕ, РАССЧИТАННЫЕ НА ИСПОЛЬЗОВАНИЕ ЛЮДЬМИ С ИНВАЛИДНОСТЬЮ И ДРУГИМИ ОГРАНИЧЕНИЯМИ ЖИЗНЕДЕЯТЕЛЬНОСТИ. СИСТЕМЫ ВЫВОДА ТЕКСТОВОЙ ИНФОРМАЦИИ В ФОРМЕ СИНТЕЗИРОВАННОЙ РЕЧИ

Технические требования

Interactive digital systems usable by persons with disabilities. Digital information output systems in the form of syntheized speech.Technical requirements

Дата введения — 2022—04—01

1 Область применения

Настоящий стандарт распространяется на системы речевого вывода, предназначенные для представления произвольной текстовой информации в форме синтезированной речи для широкого круга лиц, устанавливает требования и/или особенности деятельности которых предполагают постоянное или ситуационное использование звучащей (устной) речи как способа взаимодействия с человеко-ориентированной информационной системой или контроля ее работы. Примерами подобного взаимодействия могут служить:

- усиление воздействия визуальной информации при помощи речевых сообщений;
- речевое представление информации, когда отсутствуют аппаратно-программные возможности для их визуального представления;
- обеспечение пользователя информацией в ситуациях, когда невозможно использовать зрение (отсутствует освещение или зрение пользователя ослаблено в силу возраста, травмы или болезни);
- выполнение пользователем сложных манипуляций, при которых визуальный канал максимально используется для решения приоритетных задач, но другая необходимая информация воспринимается через слуховой канал (например, управление автомобилем и одновременное прослушивание сообщений навигационного устройства).

2 Нормативные ссылки

В настоящем стандарте использованы нормативные ссылки на следующие стандарты:

ГОСТ Р 50840—95 Передача речи по трактам связи. Методы оценки качества, разборчивости и узнаваемости

ГОСТ Р 51645 Рабочее место для инвалида по зрению типовое специальное компьютерное. Технические требования к оборудованию и производственной среде

ГОСТ Р 52872 Интернет-ресурсы и другая информация, представленная в электронно-цифровой форме. Приложения для стационарных и мобильных устройств, иные пользовательские интерфейсы. Требования доступности для людей с инвалидностью и других лиц с ограничениями жизнедеятельности

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,

Издание официальное

на который дана недатированная ссылка, то рекомендуется использовать действующую версию этого стандарта с учетом всех внесенных в данную версию изменений. Если заменен 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, определения и сокращения

3.1 В настоящем стандарте применены следующие термины с соответствующими определениями:

3.1.1 **вспомогательные технологии:** Аппаратное и программное обеспечение, применяющееся пользователем с ограниченными возможностями здоровья отдельно или совместно с основным аппаратно-программным комплексом для обеспечения функциональности, недостижимой с помощью обычных аппаратных и программных средств.

3.1.2 **высота тона:** Параметр речи, соответствующий высоте основного тона синтезированной речи.

3.1.3 **голос:** Совокупность акустических и фонетических характеристик синтезированной речи, зависящая от используемого метода синтеза, фонетической базы синтезатора речи и воспринимаемое слушателем как уникальное звучание.

3.1.4 **графическая среда:** Комплекс программных средств пользовательского устройства, обеспечивающий визуальное представление элементов графического пользовательского интерфейса на видеодисплее и взаимодействие пользователя с этими элементами посредством различных устройств ввода (клавиатуры, мыши, переключателей, сенсорных панелей и других).

Примечание — Графическая среда может быть как частью конкретной операционной среды, так и кроссплатформенным программным обеспечением, способным функционировать в различных операционных средах.

3.1.5 **громкость речи:** Параметр речи, соответствующий уровню громкости выходного звукового сигнала системы речевого вывода.

3.1.6 **импульсно-кодовая модуляция:** Способ представления цифрового сигнала в виде числовых значений амплитуды (отсчетов), следующих через фиксированные интервалы времени.

Примечание — Период следования отсчетов определяется частотой дискретизации, а диапазон числовых значений — количеством двоичных разрядов (бит) в одном отсчете.

3.1.7 **интерфейс прикладного программирования:** Описание возможностей, которые операционная среда или прикладная программа через средства программирования предоставляет другим программам, а также требования к программному коду, использующему эти возможности.

Примечание — Интерфейс прикладного программирования часто реализуется в виде отдельной программной библиотеки или службы операционной среды.

3.1.8 **интонация:** Параметр речи, соответствующий просодическим характеристикам синтезированной речи.

3.1.9 **качество речи:** Субъективная оценка слушателем звучания синтезированной речи.

3.1.10 **Международный фонетический алфавит:** Фонетический алфавит для транскрибирования любого языка.

Примечание — Международный фонетический алфавит (см.[1]) разработан и поддерживается Международной фонетической ассоциацией.

3.1.11 **метаданные:** Вспомогательная информация, поступающая в систему речевого вывода совместно с преобразуемым текстом.

Примечание — Метаданные, как правило, представляют собой цифровой текст, спецификация которого определена либо международным стандартом, либо разработчиком синтезатора, и содержат информацию, управляющую процессом и/или изменяющую параметры синтеза речи.

3.1.12 **операционная среда:** Комплекс программных средств пользовательского устройства, обеспечивающий функционирование на нем прикладных программ.

3.1.13 **параметр речи:** Характеристика синтезированной речи, изменяемая системой речевого вывода в зависимости от исходного текста и метаданных, а также через интерфейс прикладного программирования, предоставляемого системой речевого вывода для взаимодействия с пользовательским приложением.

3.1.14 **пауза:** Интервал времени между определенными частями синтезированной речи, не содержащий звука.

3.1.15 **передискретизация:** Обработка отсчетов исходного цифрового сигнала для получения цифрового сигнала с иной частотой дискретизации.

Примечание — Отсчеты сигнала, соответствующие новой частоте дискретизации, вычисляются по уже имеющимся отсчетам и не содержат новой информации.

3.1.16 **пользовательское приложение:** Прикладное программное обеспечение, взаимодействующее с пользователем, получающее, обрабатывающее входные данные и предоставляющее информацию пользователю, в том числе посредством системы речевого вывода.

3.1.17 **прикладное программное обеспечение:** Совокупность программных средств, выполняющих задачи, решаемые в предметной области пользователя.

3.1.18 **программа экранного доступа:** Пользовательское приложение, позволяющее незрячим пользователям воспринимать текстовую и иную экранную информацию посредством рельефно-точечного шрифта Брайля, синтезированной речи, вибрационных, звуковых и иных сигналов.

3.1.19 **разборчивость речи:** Характеристика синтезированной речи, определяющая семантически верное восприятие пользователем единиц речевой информации.

3.1.20 **расширенный фонетический алфавит методов оценки речи:** Система представления знаков международного фонетического алфавита с помощью набора символов 7-битного компьютерного кода ASCII.

Примечание — Система представления знаков — см. [2].

3.1.21 **синтезатор речи:** Компонент системы речевого вывода, преобразующий цифровой текст в синтезированную речь, передаваемую на аппаратные средства вывода звука.

Примечание — Синтезатор речи может быть реализован как программное обеспечение, выполняемое на пользовательском устройстве, и как внешний модуль, подключаемый к пользовательскому устройству. Подключаемый вариант также называется аппаратным синтезатором и, как правило, использует собственное устройство вывода звука.

3.1.22 **синтезированная речь:** Искусственно созданный звуковой сигнал, воспринимаемый слушателем как аналог человеческой речи.

3.1.23 **система речевого вывода:** Программное обеспечение, преобразующее цифровой текст в синтезированную речь, а также звуковые сигналы, являющиеся смысловыми аналогами речевых сообщений.

Примечания

1 Система речевого вывода взаимодействует с пользовательским приложением, получая от него текст, представленный кодами символов, который необходимо преобразовать в синтезированную речь, а также метаданные, содержащие информацию, управляющую процессом синтеза речи и/или изменяющие значения параметров речи и системы речевого вывода.

2 Системы речевого вывода используют различные архитектурно-программные решения, способы взаимодействия между внутренними компонентами, с внешними приложениями и операционной средой, чтобы реализовать требуемый функционал и обеспечить требуемое качество синтезированной речи. Например, система речевого вывода может быть реализована как неотъемлемая часть пользовательского приложения; как загружаемые модули для работы в рамках одного процесса с пользовательским приложением, предоставляя соответствующий интерфейс прикладного программирования для такого взаимодействия; как самостоятельный процесс (служба) для работы с любым пользовательским приложением по протоколу, определяемому международными, национальными, общественными или корпоративными стандартами и спецификациями.

3 Обобщенная система речевого вывода включает в себя: реализацию интерфейса прикладного программирования для взаимодействия с пользовательским приложением; блок синтеза речи, содержащий один или несколько программных или аппаратных синтезаторов речи; компонент для взаимодействия с аппаратными устройствами вывода звука.

4 В настоящем стандарте система речевого вывода рассматривается исключительно в аспекте ее главного предназначения — обеспечивать пользователя корректной, релевантной информацией в форме синтезированной речи, сформированной на основе информации, предоставляемой в виде цифрового текста пользовательским приложением и/или операционной средой.

3.1.24 **скорость речи:** Параметр речи, соответствующий количеству передаваемых речевых единиц в единицу времени (слогов в секунду или слов в минуту).

3.1.25 **специальное рабочее место:** Категория оснащенных компьютером рабочих мест, специализированных для выполнения трудовых обязанностей человеком с инвалидностью по зрению и соответствующих необходимым требованиям.

Примечание — Требования к рабочим местам — по ГОСТ Р 51645.

3.1.26 **спецификация лексикона произношения:** Язык разметки, основанный на XML и разработанный консорциумом W3C для унифицированного представления информации о произношении слов или фраз в системах распознавания и синтеза речи.

Примечания

1 Использование данного языка разметки позволяет задать соответствие между словами (или короткими фразами), их письменными представлениями и их произношением.

2 Система распознавания и синтеза речи — см. [3].

3.1.27 **текст:** Информация, воспринимаемая человеком и представленная с использованием символов какой-либо системы письменности.

3.1.28 **универсальный набор кодированных символов:** Набор символов с сопоставленными им числовыми кодами и способами кодирования.

Примечание — Универсальный набор кодированных символов — см. [4], а также стандарт Юникод. В зависимости от года издания международный стандарт полностью или частично совпадает с Юникодом (так, [4] соответствует Юникоду 13,0).

3.1.29 **фонемная база:** Обобщенное наименование для совокупности упорядоченных звуковых фрагментов в любом формате хранения цифрового звука, используемой в процессе конкатенативного синтеза речи.

3.1.30 **фонетический алфавит:** Алфавит, предназначенный для фонетической транскрипции языка устной речи.

3.1.31 **цифровой текст:** Текст, символы которого представлены числовыми значениями из какого-либо кодового набора символов.

3.1.32 **Юникод:** Международный стандарт индустрии информационных технологий для кодирования, представления и обработки текста, составленного на большинстве мировых систем письменности.

Примечания

1 Стандарт поддерживается консорциумом Юникода [5], и по состоянию на март 2020 года версия Юникод 13.0 содержит 143 859 символов, охватывающих 154 современных и исторических наборов букв, а также несколько наборов символов для смайликов.

2 Набор символов Юникода синхронизирован с [4] в части соответствия символов и их кодов.

3.1.33 **язык разметки синтеза речи:** Язык разметки, основанный на XML и разработанный консорциумом W3C для унифицированного представления информации, управляющей параметрами синтеза речи, такими как произношение, громкость, высота тона, скорость и другие.

Примечание — Язык разметки синтеза речи — см. [6].

3.2 В настоящем стандарте применены следующие сокращения:

- API — (Application Programming Interface) интерфейс прикладного программирования, программный интерфейс приложения;
- IPA — (International Phonetic Alphabet) международный фонетический алфавит;
- PCM — (Pulse Code Modulation) импульсно-кодовая модуляция;
- PLS — (Pronunciation Lexicon Specifications) спецификация лексикона произношения;
- SSML — (Speech Synthesizers Markup Language) язык разметки синтеза речи;
- X-SAMPA — (Extended Speech Assessment Methods Phonetic Alphabet) расширенный фонетический алфавит методов оценки речи.

4 Технические требования

4.1 Система речевого вывода должна удовлетворять требованиям настоящего стандарта независимо от того, какие компоненты и модули задействованы в синтезе речи по тексту, являются ли эти компоненты и модули частью операционной среды или разработаны сторонними разработчиками.

Если компоненты операционной среды, используемые при синтезе речи по тексту, не позволяют системе речевого вывода соответствовать требованиям настоящего стандарта, то система речевого вывода должна использовать аналогичные по функционалу сторонние компоненты, позволяющие реализовать требования настоящего стандарта.

Система речевого вывода должна реализовывать все требования настоящего стандарта независимо от того, существует ли возможность реализовать их на стороне пользовательского приложения, взаимодействующего с системой речевого вывода. Если система речевого вывода и прикладная программа, взаимодействующая с ней, составляют единый программный продукт и не предполагают раздельного использования, что явно оговорено в лицензионном соглашении и условиях эксплуатации, то требования к системе вывода речи относятся ко всему продукту целиком, а выбор внутреннего компонента, который будет реализовывать требования настоящего стандарта, остается на усмотрение разработчика.

Примечание — Примером единого программного продукта может быть программа экранного доступа, поставляемая совместно со встроенными модулями синтеза речи по тексту, которые поддерживают работу исключительно с данной программой экранного доступа, не могут быть отделены от продукта без нарушения лицензионного соглашения и не содержат открытых интерфейсов для взаимодействия с другими прикладными программами. В то же время, такие модули могут использовать компоненты операционной среды и любые компоненты, входящие в состав поставки программного продукта.

4.2 Система речевого вывода должна соответствовать:

а) всем требованиям используемой операционной среды, предъявляемым к прикладным программам;

б) требованиям операционной среды, предъявляемым к прикладным программам, использующим системный API для поддержки синтеза речи по тексту, если система вывода речи использует указанный API;

в) требованиям прикладной программы, совместно с которой используется данная система.

4.3 Система речевого вывода, используемая в составе специального рабочего места, должна соответствовать, в том числе и в части предоставляемого API, требованиям программы экранного доступа, которой оснащено данное рабочее место.

Примечания

1 Для систем речевого вывода, входящих в единый программный продукт с программой экранного доступа (4.1), требование данного пункта выполняется по умолчанию только для программы экранного доступа из указанного единого программного продукта.

2 Для систем речевого вывода, входящих в единый программный продукт с программой экранного доступа (4.1), требования настоящего пункта в отношении программ экранного доступа, установленных на специальное рабочее место и не входящих в единый программный продукт с системой речевого вывода, не является обязательным.

4.4 Система речевого вывода должна обеспечивать устойчивую совместную работу с пользовательским приложением, обрабатывать все варианты входных данных, согласно 4.7, в том числе и содержащие ошибки, обрабатывать любые сочетания символов во входных данных без аварийного завершения работы.

4.5 Вспомогательное программное обеспечение, входящее в состав системы речевого вывода и предназначенное для взаимодействия с пользователем при помощи пользовательского графического интерфейса, должно удовлетворять требованиям доступности, предъявляемым ГОСТ Р 52872.

Для системы речевого вывода, используемой в составе специального рабочего места, указанное вспомогательное программное обеспечение должно быть доступно для программы экранного доступа, которой оснащено данное рабочее место.

4.6 Система речевого вывода в качестве входных данных должна принимать:

а) цифровой текст для преобразования в речь, представленный в кодировках, поддерживаемых используемой операционной средой;

б) цифровой текст совместно с метаданными, представленными с использованием SSML;

в) транскрибированный текст в виде отдельных слов, представленный с использованием знаков IPA в кодах Юникода. Дополнительно может поддерживаться представление знаков IPA по системе X-SAMPA;

г) информацию о произношении, представленную с использованием PLS.

Данные требования не являются обязательными для системы речевого вывода в рамках единого программного продукта (4.1); в таком случае требования к входным данным определяются прикладной программой, использующей систему вывода речи и входящую в тот же единый программный пакет.

4.7 Система речевого вывода, в составе специального рабочего места или взаимодействующая с программой экранного доступа, должна поддерживать:

а) корректное произношение любых символов Юникода, в том числе состоящих из более чем одной кодовой точки (например, основной и комбинируемые знаки). Символы Юникода, представленные последовательностью основного и комбинируемых знаков, должны произноситься как единый символ в потоке речи (то есть при чтении слов, предложений), и как последовательность отдельных символов (знаков), входящих в его состав, при посимвольном чтении. Для символов Юникода, не входящих в алфавит языка синтезированной речи, должно озвучиваться их локализованное название либо определенное Юникодом, либо подготовленное на основе англоязычного названия, определенного Юникодом;

б) произношение любого сочетания букв, в том числе и невозможного или ошибочного по правилам произношения языка синтезированной речи;

в) произношение числовых данных.

4.8 Синтезированная речь должна соответствовать современным орфоэпическим нормам используемого языка (для русского языка см. [4]).

В частности, для русского языка должны выполняться:

а) правильное произношение слов с буквой «ё» в случаях, когда в тексте буква «ё» представлена буквой «е»;

б) распознавание и правильное произношение сокращений и аббревиатур. Под правильным подразумевается принятое в современном русском языке произношение самих сокращений и аббревиатур, а не их раскрытие.

4.9 Система речевого вывода должна поддерживать пользовательский словарь произношений, а также предоставлять пользователю возможность внесения новых, редактирования и удаления существующих записей из данного словаря.

Данное требование не распространяется на электронно-цифровые комплексы массового применения, не использующие аутентификацию или не поддерживающие долговременное хранение учетных данных пользователей (например, банкоматы или иные компьютеризированные терминалы самообслуживания).

Возможность изменять пользовательский словарь может быть реализована как в виде вспомогательного приложения с графическим пользовательским интерфейсом, удовлетворяющим 4.6, так и в форме текстовых файлов словаря произношений, доступных для чтения и редактирования пользователем при помощи текстового редактора.

4.10 Система речевого вывода должна обеспечивать вывод звука посредством соответствующих интерфейсов, стандартизованных или специфицированных для базовых операционной среды и аппаратной платформы, и формировать звуковой поток данных на аппаратное средство вывода звука. Взаимодействие с указанными средствами может выполняться как при помощи API, предоставляемого операционной средой, так и программными средствами самой системы речевого вывода, если необходимый API операционной среды отсутствует или не позволяет воспроизвести звук в том качестве, которое определено настоящим стандартом.

4.11 Система речевого вывода должна преобразовывать текст быстрее, чем он выдается в виде синтезированной речи через аппаратные средства вывода звука.

Данное требование не распространяется на системы речевого вывода, не участвующие в момент своей работы в интерактивном взаимодействии с пользователем, и осуществляющие вывод звуковых данных не через аппаратное средство воспроизведения звука.

Примечания

1 Данное требование, в частности, означает, что паузы между единицами речи должны определяться значениями соответствующих параметров, а не задержками, возникающими в ходе преобразования текста в речь.

2 Примером работы системы речевого вывода, не участвующей в интерактивном взаимодействии с пользователем, может быть работа такой системы совместно с приложением, которое преобразует текстовые данные в синтезированную речь и сохраняет результат в виде одного или нескольких звуковых файлов.

4.12 Система речевого вывода должна синтезировать звуковой поток, параметры которого эквивалентны или выше параметров звукового потока, кодированному линейной PCM с частотой дискретизации 22 050 Гц, размерностью отсчетов 16 бит и числом каналов 1 (моно).

Если аппаратные средства устройства поддерживают воспроизведение звука только с частотой дискретизации или размерностью отсчетов, ниже требуемых, то система речевого вывода должна синтезировать звук с частотой дискретизации и размерностью отсчетов максимально близких к требуемым.

Если в системе речевого вывода используется фонемная или речевая база, то указанные параметры должны быть результатом представления звуковых данных в таких базах, а не результатом передискретизации, изменения разрядности отсчетов или числа каналов звукового потока, полученного с использованием исходной фонемной или речевой базы.

Система речевого вывода может синтезировать звуковой поток с более высокой частотой дискретизации, более высокой размерностью отсчета и большим числом каналов, если воспроизведение звука с такими параметрами поддерживается операционной средой и аппаратной платформой.

Если система речевого вывода поддерживает различные варианты параметров звукового потока, пользователю должна быть предоставлена возможность выбрать вариант по своему усмотрению.

4.13 Система речевого вывода должна предоставлять возможность пользователю по своему усмотрению выбрать аппаратное устройство для вывода звука, если в аппаратно-технической платформе присутствует более одного такого устройства, а операционная среда поддерживает работу с ними. Система речевого вывода должна предоставлять такой выбор независимо от того, предоставляет ли его пользовательское приложение.

4.14 Система речевого вывода должна обеспечивать естественность звучания синтезированной речи.

Для оценки естественности звучания допускается использовать способ, основанный на 10.5 ГОСТ Р 50840—95, когда группа аудиторов прослушивает фрагменты синтетической речи и сравнивает их звучание с образцом голоса диктора, принимавшего участие в создании фонемной базы синтезатора, оценивая схожесть звучания по пятибалльной шкале.

Для методов синтеза речи, не использующих записи естественной речи дикторов-доноров, естественность звучания допускается оценивать, предлагая аудиторам прослушать различные фрагменты синтезированной речи и оценить их естественность звучания по пятибалльной шкале.

Примечание — Естественность звучания речи субъективно зависимая характеристика, и для систем речевого вывода, используемых в составе специального рабочего места или взаимодействующих с программой экранного доступа, достижение высокой естественности звучания не должно быть приоритетным по сравнению с комфортом, низкой утомляемостью и высокой эффективностью при выполнении пользователем своих трудовых обязанностей, а также других действий с использованием системы речевого вывода.

4.15 Система речевого вывода, используемая в составе специального рабочего места, должна поддерживать синтез речи на русском и английском языках. Дополнительно данная система речевого вывода может поддерживать синтез речи на национальном языке субъектов Российской Федерации, если специальное рабочее место установлено на территории, где такой язык употребляется при обучении или в профессиональной деятельности.

Система речевого вывода должна поддерживать синтез речи на языке, для которого локализовано пользовательское приложение.

4.16 Система речевого вывода может поддерживать синтез речи на любых языках.

4.17 Система речевого вывода должна обеспечивать необходимую разборчивость синтезированной речи при любых поддерживаемых значениях скорости речи. Разборчивость речи определяется по ГОСТ Р 50840.

4.18 Система речевого вывода должна поддерживать изменение скорости речи при неизменной высоте тона голоса. В качестве нормальной скорости устанавливается нормальный темп речи по ГОСТ Р 50840.

Диапазон поддерживаемых значений должен охватывать от 50 % до 200 % от значения нормальной скорости и содержать не менее пяти неубывающих промежуточных значений, равноудаленных друг от друга.

Примечание — Поскольку абсолютные значения скорости речи являются малоинформативными для пользователя, то во вспомогательном приложении, предоставляющем возможность пользователю по своему усмотрению установить значение скорости речи, рекомендуется использовать либо набор предопределенных значений с наименованиями «очень медленно», «медленно», «нормально», «быстро», «очень быстро», «сверхбыстро», либо безразмерную или процентную шкалу.

4.19 Система речевого вывода, используемая совместно с программой экранного доступа или в составе специального рабочего места, должна содержать хотя бы один голос, поддерживающий максимальное значение скорости речи не менее четырехкратного значения нормальной скорости.

4.20 Система речевого вывода должна поддерживать изменения основного тона голоса при неизменной скорости речи. Диапазон поддерживаемых значений должен охватывать от 50 % до 150 % от значения нормальной высоты голоса.

4.21 Система речевого вывода должна поддерживать изменения громкости речи во всем диапазоне значений, определяемых операционной средой и программным обеспечением.

Примечание — Изменение громкости должно затрагивать только звуковые данные, создаваемые системой речевого вывода, и не влиять на общий уровень громкости в операционной среде или затрагивать звуковые данные других программ.

4.22 Система речевого вывода должна предоставлять возможность изменения, переключения, включения и отключения пользователем по его выбору поддерживаемых параметров речи (обязательных и необязательных в соответствии с настоящим стандартом), таких как высота, громкость, скорость, длительность пауз между словами и предложениями, используемый голос, интонирование, эмоциональная окраска, распознавание и переключение языка, выбор параметров звука и аудиоустройства для воспроизведения синтезированной речи.

Внесенные пользователем изменения должны сохраняться между периодами работы системы речевого вывода, исключая аппаратно-программные комплексы массового применения, не использующие аутентификацию или не поддерживающие долговременное хранение учетных данных пользователей (например, банковские или иные компьютеризированные терминалы самообслуживания).

Значения параметров, установленные пользователем, имеют более высокий приоритет, чем значения, используемые системой речевого вывода по умолчанию.

Разработчик системы речевого вывода может предоставить пользователю возможность изменять другие параметры речи. Пользователю должна быть предоставлена возможность в любой момент вернуть изменяемые параметры к их исходным значениям.

Исходные значения изменяемых параметров должны соответствовать настоящему стандарту и выбираться разработчиком на основе экспертной оценки работы системы речевого вывода.

4.23 Система речевого вывода должна предоставлять пользователю возможность изменить названия символов для их произношения при посимвольном чтении, а также возможность отключить произнесение символов Юникода, не входящих в алфавит языка синтезируемой речи.

Внесенные пользователем изменения должны сохраняться между периодами работы системы речевого вывода.

Названия символов, установленные пользователем, имеют более высокий приоритет, чем названия символов, используемые системой речевого вывода по умолчанию.

Отключение произнесения символов Юникода должно быть возможно как для отдельного символа, так и для диапазонов/наборов символов.

Требования, указанные в данном пункте, не распространяются на аппаратно-программные комплексы массового применения, не использующие аутентификацию или не поддерживающие долговременное хранение учетных данных пользователей (например, банковские или иные компьютеризированные терминалы самообслуживания).

4.24 Система речевого вывода, используемая в составе специального рабочего места или взаимодействующая с программой экранного доступа, должна для русского языка поддерживать режим особого произношения, при котором звучащая синтезированная речь максимально идентична входному тексту независимо от правил произношения. В частности: озвучиваются все буквы, в том числе и непроизносимые в обычном режиме; все гласные звуки произносятся одинаково четко независимо от того, являются ли они ударными или нет; все согласные звуки произносятся четко без замены их парными глухими или звонкими.

Примечания

1 Данный режим позволяет пользователю анализировать на слух грамматические и орфографические характеристики читаемого текста, выявлять многие ошибки эффективнее, чем с использованием современных систем проверки орфографии.

2 Выбор способа реализации режима особого произношения предоставляется разработчику системы речевого вывода. К примеру, это может быть и набор специальных настроек для того же голоса, что используется в обычном режиме, либо голос, используемый специально для особого произношения, с иными, чем у основного голоса, характеристиками.

4.25 Система речевого вывода может предоставлять режим особого произношения для любого из поддерживаемых языков синтезируемой речи.

4.26 Система речевого вывода, работающая в режиме особого произношения, должна соответствовать требованиям настоящего стандарта [при этом для перечисления г) 4.29 удовлетворительным считается степень выраженности от нуля до двух баллов включительно].

4.27 Система речевого вывода, работающая в режиме особого произношения, должна отключать эмоциональную окраску синтезированной речи, если иное не установлено пользователем.

4.28 Система речевого вывода должна предоставлять пользователю возможность включать и отключать режим особого произношения в любой момент работы системы речевого вывода.

4.29 Система речевого вывода должна соответствовать требованиям качества речи. Следующие признаки синтезированной речи являются причинами ее низкого качества:

а) картавость — искажение сонорных «л», «р»;

б) плаксивость — дрожание основного тона;

в) гнусавость — звучание речи с излишней назализацией («произношение в нос»),

г) механический голос — монотонность речи, потеря естественности звучания речи,

д) дребезжание, хрип — неровный, резкий и прерывистый звук;

е) помехи — шумы и призвуки в речевых паузах;

ж) ошибки в произношении,

и) замена звуков, их выпадение или добавление;

к) отсутствие пауз или их неправильная, произвольная длительность;

л) неверный ритм чтения;

м) неровная или неверная интонация;

н) неверная или чрезмерная эмоциональная окраска;

п) нарушения плавности речи (дефекты в речевом сигнале): прерывистость, скачки, «бульканье», стук и т. п.

Для выявления указанных признаков и оценки их выраженности используется способ, который базируется на ГОСТ Р 50840 «измерение качества речи методом оценки по селективным признакам» и состоит в прослушивании аудиторами фрагментов синтезированной речи, созданных в реальном времени оцениваемой системой речевого вывода по произвольному тексту. Каждый из аудиторов по результатам прослушивания определяет наличие признака в синтезированной речи и оценивает степень его выраженности по пятибалльной шкале. Прослушивание выполняется для минимального, нормального и максимального значений всех параметров речи, изменения которых должны поддерживаться системой речевого вывода по требованиям настоящего стандарта.

Система речевого вывода, используемая в составе специального рабочего места или взаимодействующая с программой экранного доступа, получившая для требуемых значений параметров речи по каждому из указанных признаков среднюю оценку степени выраженности не выше одного балла, считается соответствующей настоящему стандарту в части требований по качеству речи.

Система речевого вывода, используемая в составе специального рабочего места, должна проходить указанную процедуру оценки качества речи непосредственно на данном специальном рабочем месте с применением программы экранного доступа.

4.30 Система речевого вывода, используемая в составе специального рабочего места или взаимодействующая с программой экранного доступа, должна обеспечивать единообразие произношения идентичных фрагментов текста, интонационную и эмоциональную нейтральность синтезированной речи при озвучивании информации, предоставляемой программой экранного доступа при отсутствии метаданных, явно определяющих интонационные и/или эмоциональные характеристики речи.

Примечание — Как правило, текст, поступающий от программы экранного доступа для преобразования в речь, содержит названия, состояние, роль и содержимое элементов управления пользовательского графического интерфейса, а также уведомления о происходящих на экране изменениях. Программы экранного доступа используют изменения параметров речи, чтобы сообщить пользователю о происходящих изменениях, не нагружая его слух текстовыми описаниями событий. Неожиданные изменения в произношении этой информации, а также изменение интонации и наличие эмоциональной окраски могут создавать у пользователя с нарушенным или отсутствующим зрением ложное впечатление о происходящих на экране событиях или оказывать на пользователя раздражающее воздействие.

4.31 Система речевого вывода, используемая в составе специального рабочего места или взаимодействующая с программой экранного доступа, не должна синтезировать речь или создавать звуковые сигналы, негативно влияющие на пользователя, вызывающие повышенную утомляемость и снижающие его работоспособность.

Примечания

1 Требования данного пункта распространяются на звуковые сигналы (обособленные или сопровождающие синтезированную речь), на которые пользователь системы вывода речи не может повлиять (отключить и выбрать другой вариант сигнала) при помощи пользовательского приложения или вспомогательных программ в составе системы речевого вывода.

2 В качестве методики оценки негативного влияния и снижения работоспособности может быть использован опрос (анкетирование) пользователей системы речевого вывода с последующей статистической обработкой результатов, либо методика, принятая уполномоченным органом, проверяющим на соответствие настоящему стандарту. Применяемые для обработки результатов опроса (анкетирования) пользователей статистические и аналитические методы должны быть приемлемыми для указанных целей и обоснованными.

3 Трудовая деятельность сама по себе приводит к временному снижению трудоспособности и утомлению работника. Акустическое воздействие системы речевого вывода на пользователя может приводить к дополнительному снижению работоспособности и производительности труда. В качестве параметра для оценки влияния системы речевого вывода на работоспособность пользователя может быть использовано сравнительное изменение работоспособности пользователя за тот же временной отрезок трудовой деятельности или при выполнении того же объема работ, когда в качестве устройства вывода информации используется тактильный (брайлевский) дисплей со строкой не менее 40 ячеек. В таком случае подбор пользователей должен осуществляться с привлечением экспертов для оценки умения пользователя воспринимать информацию при помощи тактильного дисплея. Для целей тестирования режим работы системы речевого вывода и тактильного дисплея устанавливается в соответствии с пожеланиями пользователя. Продолжительность трудовой деятельности, в течение которой выполняется тестирование, не может быть менее 4 ч и включает 15-минутные перерывы для короткого отдыха каждый час.

4.32 Система речевого вывода должна включать в свой состав руководство пользователя в виде электронного документа, соответствующего ГОСТ Р 52872.

Для системы речевого вывода, используемой в составе специального рабочего места, руководство пользователя должно быть доступно для чтения при помощи программы экранного доступа, которой оснащено данное рабочее место.

4.33 Для системы речевого вывода должны быть представлены условия эксплуатации, содержащие минимальные и рекомендуемые значения параметров аппаратно-программного оснащения, необходимого для функционирования системы речевого вывода. Условия эксплуатации должны быть представлены в доступной для пользователя форме.

4.34 Производительность системы речевого вывода должна обеспечивать необходимую для пользователя скорость предоставления выходных речевых данных на системах, удовлетворяющих минимальным аппаратно-программным требованиям, указанным в условиях эксплуатации системы речевого вывода.

4.35 Аппаратно-программное оснащение специализированного рабочего места должно удовлетворять рекомендуемым аппаратно-программным требованиям, указанным в условиях эксплуатации системы речевого вывода, или превосходить их.

Библиография

[1] IPA symbols with Unicode decimal and hex codes: https://www.internationalphoneticalphabet.org/ipa-charts/ipa-symbols-with-unicode-decimal-and-hex-codes/

[2] SAMPA computer readable phonetic alphabet: gost.e-ecolog.ru

[3] Pronunciation Lexicon Specification (PLS) Version 1.0: gost.e-ecolog.ru

[4] ИСО/МЭК 10646:2020 Информационные технологии. Универсальный набор кодированных символов (UCS) (ISO/IEC 10646:2020) [Information technology — Universal coded character set (UCS)]

[5] Unicode — The World Standard for Text and Emoji: gost.e-ecolog.ru

[6] Speech Synthesis Markup Language (SSML) Version 1.1: https://www.w3.org/TR/speech-synthesis/

УДК 681.327.8:006.354 ОКС 11.180.30

Ключевые слова: система речевого вывода, электронно-цифровые интерактивные комплексы, синтезатор речи

Редактор *Е.В. Зубарева*
Технический редактор *В.Н. Прусакова*
Корректор *М.В. Бучная*
Компьютерная верстка *Л.А. Круговой*

Сдано в набор 26.08.2021. Подписано в печать 17.09.2021. Формат 60×84¼. Гарнитура Ариал.
Усл. печ. л 1,86. Уч.-изд. л 1.70.

Подготовлено на основе электронной версии, предоставленной разработчиком стандарта

Создано в единичном исполнении в ФГБУ «РСТ»
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