ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ
INFORMATION SYSTEM AND TECHNOLOGIES

УДК 004.934.5 **DOI: 10.18413/2518-1092-2022-8-2-0-1**

Джумаев А.Б.

О ПРОГРАММНЫХ СРЕДСТВАХ СИНТЕЗА РУССКОЙ РЕЧИ

ООО «ЛЕРУА МЕРЛЕН ВОСТОК», Осташковское шоссе, д.1, г. Мытищи, 141031 Россия

e-mail: art2371@yadex.ru

Аннотация
В современных информационных технологиях важное место занимают интерактивные методы ввода информации с последующим использованием обработанных данных. Голосовой ввод, управление жестами, захват движения, дополненная реальность, – данные средства в настоящее время широко используются в профессиональной деятельности и повседневной жизни человека. Синтез речи является частью речевых технологий, к которым также относятся распознавание речи, семантика речи и ее перевод. В настоящее время значительное количество исследователей занимаются изучением вопроса синтеза речи при создании новых программных продуктов. Поэтому оценка программных средств синтеза и сравнение их между собой является актуальной задачей. В данной работе проведён краткий анализ некоторых стационарных программных средств синтеза русской речи и дана оценка известным приложениям.
Ключевые слова: синтез речи; речевые технологии; системы синтеза речи; программные средства синтеза русской речи
Для цитирования: Джумаев А.Б. О программных средствах синтеза русской речи // Научный результат. Информационные технологии. – Т.8, №2, 2023. – С. 3-10. DOI: 10.18413/2518-1092-2022-8-2-0-1

Jumaev A.B.

ABOUT RUSSIAN SPEECH SYNTHESIS SOFTWARE

LLC «LEROY MERLIN VOSTOK», 1 Ostashkovskoe highway, Mytishchi, 141031 Russia

e-mail: art2371@yadex.ru

Abstract
In modern information technologies, an important place is occupied by interactive methods for entering information with the subsequent use of processed data. Voice input, gesture control, motion capture, augmented reality - these tools are currently widely used in professional activities and everyday life. Speech synthesis is a part of speech technologies, which also include speech recognition, speech semantics and its translation. Currently, a significant number of researchers are studying the issue of speech synthesis when creating new software products. Therefore, the evaluation of synthesis software tools and their comparison with each other is an urgent task. In this paper, a brief analysis of some stationary software tools for the synthesis of Russian speech is carried out and an assessment is made of known applications.
Keywords: speech synthesis; speech technologies; speech synthesis systems; Russian speech synthesis software
For citation: Jumaev A.B. About Russian speech synthesis software // Research result. Information technologies. – T. 8, №2, 2023. – P. 3-10. DOI: 10.18413/2518-1092-2022-8-2-0-1

ВВЕДЕНИЕ

Синтез речи на основе текстовых данных является актуальной задачей современности, так как синтезированная речь используется в различных сферах деятельности человека, к примеру, в банковских системах голосового самообслуживания, транспортных компаний, при проведении телефонных опросов и т.д.

В настоящее время известно несколько компаний и их программных продуктов, которые обладают поддержкой русского синтезированного языка – Microsoft Speech SDK (Microsoft), L&H (Lernout & Hauspie Speech Products) и Digalo (Elan Informatique) и другие.

Несмотря на то, что в мире существует масса разработок, проблема синтеза речи до сих пор не решена, так как качество синтезированной речи только в отдельных случаях можно считать удовлетворительной. Основными проблемами являются низкие показатели разборчивости, естественности и эмоциональности, которые приводят к ошибкам и сложности восприятия синтезированной речи [1, 2].

В рамках данной работы рассмотрены одни из наиболее распространённых стационарных сервисов синтеза речи, с позиций анализа их функциональных возможностей.

КЛАССИФИКАЦИЯ ПРОГРАММНЫХ СРЕДСТВ СИНТЕЗА РЕЧИ

Разработка первых русскоязычных синтезаторов началась в 2000-х года [3-11]. В настоящее время представлена масса разнообразных решений, как коммерческих, так и бесплатных. К глобальным отличиям существующих технологий можно отнести:

1. Количество поддерживаемых голосов;
2. Формат работы (отдельное приложение, вэб-версия, часть системы, интегрируемая библиотека);
3. Тип распространения (коммерческий, бесплатный);
4. Платформа использования.

В рамках исследования были отобраны и разбиты на 2 группы приложения, которые могут быть использованы для работы с персональным компьютером под управлением системой Windows и обладают поддержкой русского языка. Основной критерий разделения на группы является формат работы приложения.

К стационарным приложениям относятся:

1. «Говорилка»;
2. «Балаболка»;
3. «Vociebot»;
4. «NaturalReaders»;
5. «Robot Talk»;

К веб-приложениям относятся:

1. «2yxa»;
2. «Apihost»;
3. «Texttospeech»;
4. «TexttoSpeechRobot»;
5. «VoxWocker»;
6. «Ivona»;
7. «Acapela»;
8. «Microsoft Azure»;
9. «Yandex SpeechKit»;
10. «VoiceMaker»;
11. «OddCast».

В рамках данной работы проведён анализ группы стационарных приложений по синтезу русской речи.

СТАЦИОНАРНЫЕ ПРОГРАММНЫЕ СРЕДСТВА СИНТЕЗА РЕЧИ

Приложение **«Говорилка» –** стационарное приложение для семейства ОС Windows, основной задачей которого является озвучивание произвольного текста установленным голосом, интерфейс программы представлен на рисунке 1. Приложение основано на технологии конкатенативного синтеза. Данная технология использует конкатенацию предварительно записанных примеров человеческой речи в единую звуковую последовательность.

К основным возможностям программы относятся:

1. Запись генерируемой речи в звуковые файлы.
2. Регулировка скорости чтения и высоты тона голоса.
3. Система «слежения за речью», которая позволяет видеть озвучиваемый текст.
4. Пополняемые словари произношения с возможностью корректировки произношения.
5. Поддержка сторонних голосов.

Приложение «Говорилка» предлагает в стандартном исполнении 4 голоса (2 мужских и 2 женских), но количество голосов может быть расширено при помощи сторонних баз синтезированных голосов.

Преимущества данной программы совпадают с её основными возможностями, к минусам можно отнести:

1. низкое качество озвучиваемого текста,
2. низкая естественность речи,
3. большое количество ошибок произношения

Основные недостатки вызваны маленькой акустической базой голосов из-за чего «склейки» незнакомых слов происходят грубым образом, вызывая неестественное звучание.

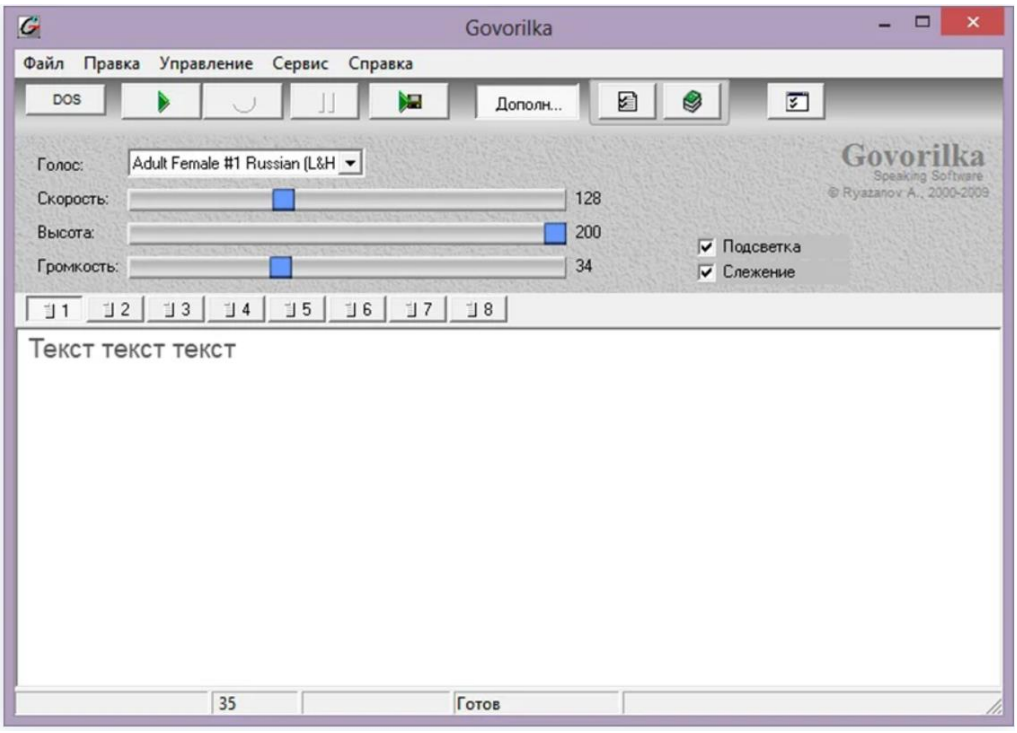

Рис. 1. Интерфейс программы «Говорилка»
Fig. 1. The interface of the «Govorilka» program

Приложение «**Балаболка»** – стационарное приложение для семейства ОС Windows, основной задачей которого является озвучивание произвольного текста, интерфейс программы представлен на рисунке 2. Данное приложение в своей основе так же, как и «Говорилка» использует метод синтеза речи на основе конкатенативного метода. Отличием данного речевого синтезатора от «Говорилки» является то, что для работы данной программы используется речевой

синтезатор операционной системы компьютера, то есть Microsoft Speech API системы Windows. К основным возможностям относятся:

1. Контроль воспроизведения речи
2. Возможность воспроизведения текста из буфера обмена
3. Озвучка набираемого текста
4. Экспорт речи в звуковые файлы

Недостатки данной программы совпадают с ранее выявленными для «Говорилки», дополнительно можно выделить:

1. Для замены голоса необходимы дополнительные плагины.
2. Отсутствует отдельный голос.

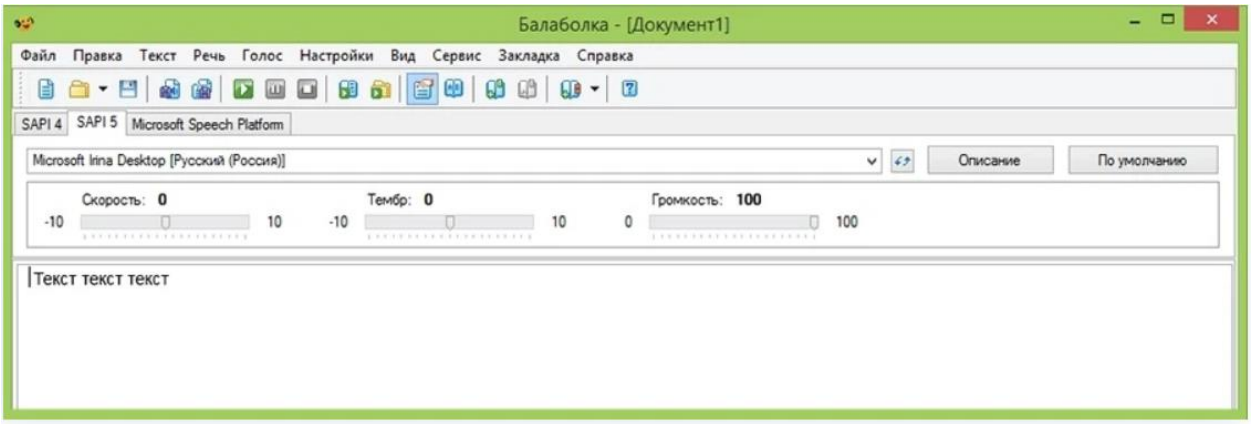

Рис. 2. Интерфейс программы «Балаболка»
Fig. 2. The interface of the «Balabolka» program

Приложение **«Voicebot» –** средство для программирования голосовых команд для управления системными службами ПК, интерфейс программы представлен на рисунке 3. В основе приложения, как и в предыдущих используется конкатенативный синтез. Данное средство не предназначено для целевого использования по озвучке текста и используется для узконаправленного использования по управлению персональным компьютером. Как и у сервиса «Балаболка» используется речевой синтезатор операционной системы компьютера. Данное программное обеспечение распространяется по платной подписке с бесплатным 30-дневным пробным периодом. «Voicebot» в процессе своей работы по заранее подготовленным командам производит распознавание речи пользователя и озвучку своих действий в рамках выполнения команд.

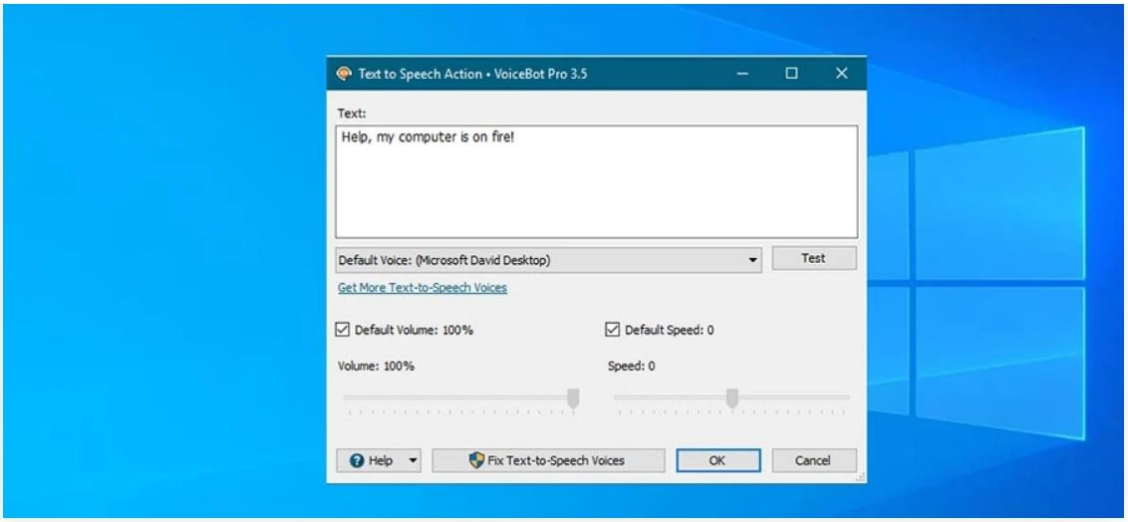

Рис. 3. Интерфейс программы «Voicebot»
Fig. 3. The interface of the « Voicebot » program

Приложение «**NaturalReaders**» **–** универсальное программное обеспечение, которое имеет, как стационарную версию приложения для персонального компьютера, так и веб-версию для генерации речи, интерфейс программы представлен на рисунке 4. В связи с ограничениями средство лишилось поддержки русского языка в обычном доступе, теперь для использования русского языка необходим пакет «Plus» в рамках платной подписки. Доступны 2 голоса (1 мужской и 1 женский), которые являются базовыми голосами для программного средства «Говорилка» и основаны на той же технологии. Так как «NaturalReaders» обладает двумя версиями приложения – стационарную и веб версии для них есть ряд отличающих особенностей:

1. Стационарная версия приложения может быть использована только на 1 компьютере, на котором была произведена установка средства;
2. При покупке лицензии для приложения будут разблокированы от 2 до 6 «премиальных» голосов, доступ к остальной библиотеке голосов можно получить за дополнительную плату в 40$ за каждый голос;
3. Для полноценной работы приложения не требуется подписка, что предполагает единоразовую оплату.

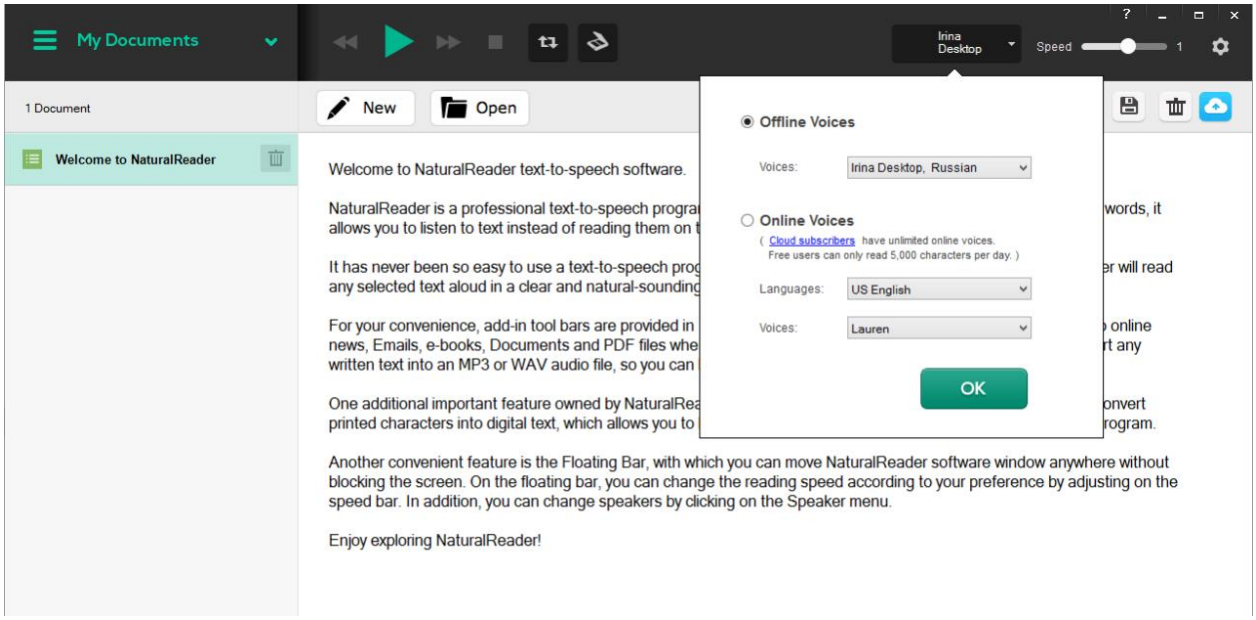

Рис. 4. Интерфейс программы «NaturalReaders»
Fig. 4. The interface of the «NaturalReaders» program

Приложение **«Robot Talk»** – бесплатное приложение (рис. 5) Windows Store, которое обладает 5 голосами (3 мужских и 2 женских), из которых 1 русский. Для синтеза речи используется та же технология, что и у предыдущих программных средств за исключением того, что Microsoft Speech API использует дополнительный модифицированный голос английской речи. Функциональные возможности приложения позволяют изменять тембр голоса и скорость речи. Так же присутствует возможность сохранения аудиофайлов. Недостатком данной программы является отсутствие русского интерфейса программы.

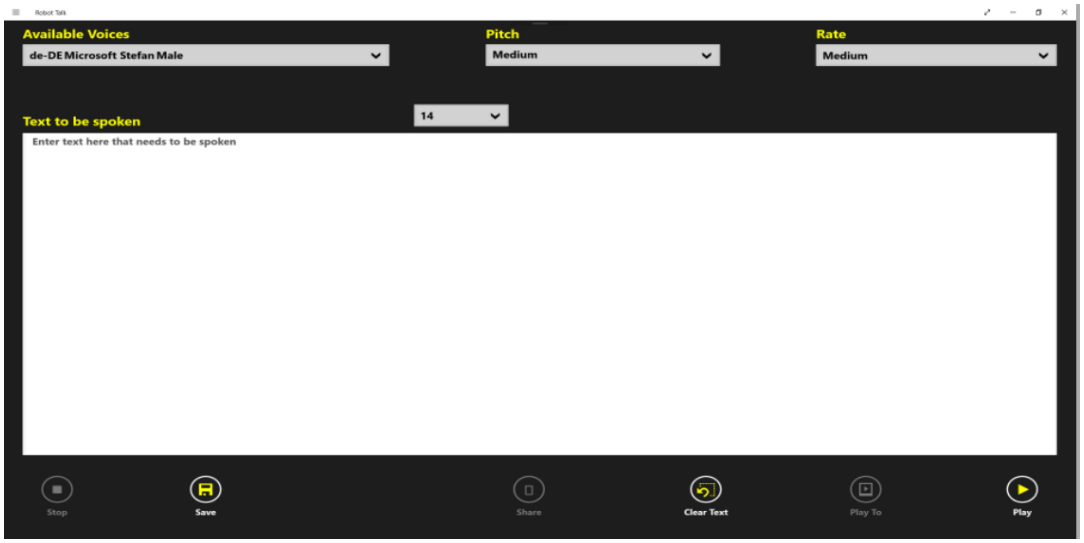

Рис. 5. Интерфейс программы «Robot Talk»
Fig. 5. The interface of the «Robot Talk» program

РЕЗУЛЬТАТЫ АНАЛИЗА СТАЦИОНАРНЫХ СРЕДСТВ

При наличии большого количества стационарных приложений их функционал не сильно отличается, а глобальные функции идентичны для всех программ. В зависимости от программы присутствует разброс по количеству голосов, но качество озвучивания у всех представленных аналогов находится на низком уровне и не может сравниться с естественной речью. Низкое качество синтезированной речи вызвано маленькой речевой базой и большим количеством ошибок при озвучивании. Малым исключением из списка является «Говорилка», так как она обладает функцией пополняемого словаря, что позволяет уменьшить количество ошибок синтеза речи в отличии от всех других представленных систем.

В таблице представлены сводные данные о стационарных средствах синтеза речи.

Таблица

Некоторые характеристики стационарных синтезаторов речи для русского языка

Table

Some characteristics of stationary speech synthesizers for the Russian language

Система синтеза	Доступные голоса	Преимущества	Недостатки	Комментарий
Говорилка	Николай Анна Digital Male voice Digital Female Voice	1. Большой выбор голосов; 2. Пополнение словаря произношений 3. Возможность установки сторонних голосов.	Низкое качество звучания	Оптимальное решение по функциональным возможностям
Балаболка	Встроенный в систему (Microsoft speech API)	1. Контроль воспроизведения речи; 2. Возможность воспроизведения текста из буфера обмена	1. Низкое качество звучания 2. Отсутствуют голоса, кроме предустановленных в систему	

Система синтеза	Доступные голоса	Преимущества	Недостатки	Комментарий
Vociebot	Встроенный в систему (Microsoft speech API)	Автоматизация задач управления ПК	Отсутствуют голоса, кроме предустановленных в систему	Узконаправленное ПО для автоматизации задач у правления ПК
NaturalReaders	Максим Ирина	Обладает, как стационарной версией, так и веб-версией	1. Высокая стоимость подписки 2. Ограничение на использование русскоязычных голосов	
Robot Talk	Встроенный в систему (Microsoft speech API)	-	1. Низкое качество звучания 2. 1 русскоязычный голос 3. Интерфейс программы на английском языке	Отсутствуют преимущества, т.к. модифицированный голос доступен только для английского языка

ЗАКЛЮЧЕНИЕ

В работе рассмотрены стационарные программные средства синтеза русской речи. При обзоре стационарных программных средств синтеза выявлена закономерность использования идентичных голосов, которые распространяются в рамках платного контента. Отличительной чертой является дополнительный функционал в виде эмоциональной окраски и регулировки тембра синтезированного голоса. Несмотря на то, что сервисов достаточно много, бесплатных решений доступно малое количество, при этом качество синтезированной речи – роботизированное в сравнении с естественной речью. Современной тенденцией является требование от ИТ-разработчиков создания облачных синтезаторов речи, а не стационарных, о чем свидетельствует значительное количество доступных средств синтеза речи в виде веб-приложений. Так как технологии синтеза речи находятся в процессе постоянного развития, то разработка новых синтезаторов русской речи доступна не только крупным ИТ-компаниям, но и частным разработчикам. Доступность данной технологии и спрос со стороны потребителя делает разработку новых программных средств синтеза речи актуальной задачей.

Список литературы

1. D.B.Keele, JR., Evaluation of Room Speech Transmission Index and Modulation Transfer Function by the Use of Time Delay Spectrometry, Techron, Div. Crown International, Inc., Elkhart, Indiana, 46517, USA.

2. A method for subjective performance assessment of the quality of speech voice output devices. ITU-T Recommendation P. 85. ITU-T, 1994.

3. Корольков В.А., Главатских И.А., Таланов А.О. Синтез естественной русской речи при помощи метода Unit Selection // Тр. XXXVI межд. Филолог. Конф. «Формальные методы анализа русской речи». Россия, 2008.

4. Джумаев А.Б. Синтезирование русской речи при помощи метода Unit Selection / VIII Международная научно-техническая конференция «Информационные технологии в науке, образовании и производстве», Белгород, 2020 г. – С. 43-46.

5. French N., Steinberg J. Factors Governing the Intelligibility of Speech Sounds // J.Acoust. 6 oc. Am. – 1947. – Vol. 19, No 1.

7. Fletcher H., Galt F. Perception of Speech and its Relation to Telephony // J. Acoust Soc. Am. – 1950. – Vol. 22, No 2.

7. Kryter K.D. Methods for the calculation and use of the articulation index // J. Acoust Soc. Am. – 1962. – Vol. 34. – P. 1689–1697.

8. ANSI S3.5-1997, American National Standard Methods for Calculation of the Speech Intelligibility Index – American National Standards Institute, New York. – 1997.

9. Беранек Л. Расчет речевых систем связи // Proceedings of the IRE. – 1947. – September. – P. 880-890.

10. Steeneken H.J.M., Houtgast T. RASTI: A Tool for Evaluating Auditoria // Bruel &Kjaer Technical Review No. 3 – 1985. – P.13-39.

11. Steeneken H.J.M., Houtgast T. RASTI: The Modulation Transfer Function in Room Acoustics // Bruel & Kjaer Technical Review No.3 – 1985. – P.1-12.

References

1. D.B.Keele, JR., Evaluation of Room Speech Transmission Index and Modulation Transfer Function by the Use of Time Delay Spectrometry, Techron, Div. Crown International, Inc., Elkhart, Indiana, 46517, USA.

2. A method for subjective performance assessment of the quality of speech voice output devices. ITU-T Recommendation P. 85. ITU-T, 1994.

3. Korolkov V.A., Glavatskikh I.A., Talanov A.O. Synthesis of natural Russian speech using the Unit Selection method // Tr. XXXVI Int. Philologist. Conf. "Formal Methods for the Analysis of Russian Speech". Russia, 2008.

4. Dzhumaev A.B. Synthesizing Russian speech using the Unit Selection method / VIII International Scientific and Technical Conference "Information Technologies in Science, Education and Production", Belgorod, 2020 – P. 43-46.

5. French N., Steinberg J. Factors Governing the Intelligibility of Speech Sounds // J.Acoust. 6 oc. Am. – 1947. – Vol. 19, No 1.

7. Fletcher H., Galt F. Perception of Speech and its Relation to Telephony // J. Acoust Soc. Am. – 1950. – Vol. 22, No 2.

7. Kryter K.D. Methods for the calculation and use of the articulation index // J. Acoust Soc. Am. – 1962. – Vol. 34. – P. 1689–1697.

8. ANSI S3.5-1997, American National Standard Methods for Calculation of the Speech Intelligibility Index – American National Standards Institute, New York. – 1997.

9. Beranek L. Calculation of speech communication systems // Proceedings of the IRE. – 1947. – September. – P. 880-890.

10. Steeneken H.J.M., Houtgast T. RASTI: A Tool for Evaluating Auditoria // Bruel &Kjaer Technical Review No. 3 – 1985. – P.13-39.

11. Steeneken H.J.M., Houtgast T. RASTI: The Modulation Transfer Function in Room Acoustics // Bruel & Kjaer Technical Review No. 3 – 1985. – P.1-12.

Джумаев Артём Бахтиёрович, аналитик ИТ-процессов отдела поддержки пользователей ООО «Леруа Мерлен Восток»

Jumaev Artem Bakhtierovich, IT process analyst of the User Support Department of Leroy Merlin Vostok LLC