

СИСТЕМНЫЙ АНАЛИЗ И УПРАВЛЕНИЕ SYSTEM ANALYSIS AND PROCESSING OF KNOWLEDGE

УДК 004.75

DOI: 10.18413/2518-1092-2018-3-3-0-2

Кузнецов Д.А.¹
Кузнецов А.В.¹
Тезин А.В.¹
Басов О.О.²

**СРАВНИТЕЛЬНЫЙ АНАЛИЗ СИНТЕЗАТОРОВ РЕЧИ
ДЛЯ ПОДСИСТЕМЫ ОПОВЕЩЕНИЯ ИНТЕЛЛЕКТУАЛЬНОГО
ЗАЛА СОВЕЩАНИЙ**

¹) Федеральное государственное казённое военное образовательное учреждение высшего образования «Академия Федеральной службы охраны Российской Федерации», ул. Приборостроительная, д. 35, г. Орёл, 302034, Россия

²) Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики», Кронверкский пр., д. 49, г. Санкт-Петербург, 197101, Россия

e-mail: wvxp@mail.ru, oficerow@mail.ru, kvaa77@mail.ru, oobasov@mail.ru

Аннотация

Развитие концепции интеллектуального зала совещаний влечет за собой реализацию взаимодействия человек-машина, посредством которого осуществляется обмен информацией между пользователями и устройством управления интеллектуальным залом совещания. Одним из способов такого обмена является воспроизведение синтезированной речи на основе обрабатываемых данных. В настоящее время существует два типа синтезаторов речи: с ограниченным и неограниченным словарем. Устройства первого класса хранят речь в виде слов и предложений, которые выводятся в определенной последовательности при синтезе сообщения. Эти синтезаторы используют модель компилятивного синтеза и параметрическое представление. Устройства второго типа требуют большого объема знаний акустических и лингвистических ограничений. Такие синтезаторы речи используют метод полного синтеза по правилам. Для сравнения существующих решений по синтезу речи вынесен ряд оценочных критериев. При анализе учтена возможная необходимость модернизации программы под конкретную реализацию с учетом требований, предъявляемых заказчиком.

Ключевые слова: интеллектуальное пространство; зал; исполнительные модули; освещение; автоматизация; климат-контроль; жизнеобеспечение.

UDC 004.75

Kuznetsov D.A.¹
Kuznetsov A.V.¹
Tezin A.V.¹
Basov O.O.²

**THE COMPARATIVE ANALYSIS OF THE SPEECH SYNTHESIZERS
FOR THE NOTIFICATION SUBSYSTEM OF SMART HALL**

¹) Federal state military educational institution of higher professional education "Academy of the Federal security service of the Russian Federation", 35 Priborostroitelnaya St, Orel, 302034, Russia

²) Saint Petersburg National Research University of Information Technologies, Mechanics and Optics, 49 Kronverkskiy prospekt, St. Petersburg, 197101, Russia

e-mail: wvxp@mail.ru, oficerow@mail.ru, kvaa77@mail.ru, oobasov@mail.ru

Abstract

Development of the concept of the intellectual hall of meetings involves realization of interaction of people technician through which exchange of information between users and the control unit by the intellectual hall of a meeting is carried out. One of ways of exchange of information between the actuation device of the intellectual hall of a meeting and a wide range of users is reproduction of the synthesized speech on the basis of the processed data. Now there are two types of synthesizers of the speech: with the limited and unlimited dictionary. Devices of the first class store the speech in the form of words and offers which are displayed in a certain sequence at synthesis messages. These synthesizers use model of kompilyativny synthesis and parametrical representation. Devices of the second type demand the large volume of knowledge of acoustic and linguistic restrictions. Such synthesizers of the speech use a method of full synthesis by rules. For comparison of the existing decisions on synthesis of the speech a number of estimated criteria is taken out. In the analysis possible need of modernization of the program under concrete realization taking into account requirements imposed by the customer is considered.

Keywords: intellectual space; room; executive modules; lighting; automation; climate control; life support.

ВВЕДЕНИЕ

Важной составляющей концепции интеллектуального зала совещаний [1] является реализация взаимодействия человек-машина, посредством которого осуществляется обмен информацией между пользователями и устройством управления интеллектуальным залом совещания [2]. Управляющее устройство (ПЭВМ) осуществляет прием информации от пользователей посредством различных датчиков, модулей, путем распознавания речи и изображений [3]. Одним из способов передачи информации от ПЭВМ [4] к широкому кругу пользователей является воспроизведение синтезированной речи на основе обрабатываемых данных. В связи с этим стоит задача выбора синтезатора речи, пригодного к применению в подсистеме оповещения интеллектуального зала совещаний.

ОСНОВНАЯ ЧАСТЬ**Описание синтезаторов речи**

Синтезаторы речи являются специальными программами, которые предоставляют возможность перевести текстовую информацию в озвученные человеческим голосом предложения. В предлагаемом интеллектуальном зале совещания предлагается использовать данные синтезаторы в подсистеме оповещения для отображения информации о текущем состоянии в зале (температура, влажность), о состоянии исполнительных модулей (положение), для оповещения участников совещания в процессе конференции (кто выступает, кто задает вопрос и т.п.), а также для интерактивного общения с пользователями.

Существует два типа синтезаторов речи: с ограниченным и неограниченным словарем [5].

Устройства первого класса хранят речь в виде слов и предложений, которые выводятся в определенной последовательности при синтезе сообщения. Эти синтезаторы используют модель компилятивного синтеза и параметрическое представление.

Модель компилятивного синтеза

Данная модель предполагает синтез речи путем «склеивания» записанных образцов отдельных звуков, записанных заранее [6]. В этом случае составляется база данных звуковых фрагментов, как правило, не меньше слова. Такой способ синтеза речи обеспечивает высокое качество синтезируемой речи, так как позволяет воспроизводить форму естественного речевого сигнала. Важное достоинство метода заключается в том, что не требуется знаний об устройстве речевого тракта и структуре языка, однако существенным недостатком является необходимость в большом объеме памяти, а также трудность в генерации высококачественной речи при попытке соединения сообщений в более длинное из-за разных значений параметров речевой волны. Таким

образом, модель компилятивного синтеза достаточно проста в реализации и эффективна в системах, где не требуется синтез заранее неизвестных предложений.

Параметрическое представление

Метод параметрического представления сигнала разработан с целью устранения недостатков модели компилятивного синтеза. Этот способ представляет речевую волну в виде определенных параметров. Такой подход уменьшает объем требуемой памяти для словаря и дает большую гибкость по сравнению с моделью компилятивного синтеза [7]. Параметры отражают характерную информацию во временной или частотной области. Один из вариантов параметрического представления речевого сообщения – формантный, который генерирует речь искусственным путем, создавая нужный набор резонансов. Такая система оперирует параметрами основного тона и формантами [8]. Изменения этих параметров дают возможность изменять интонацию и временные характеристики сообщения. Такое представление является эффективным способом хранения речевой информации по сравнению с методом модели компилятивного синтеза. Другим важным преимуществом метода является гибкость, так как смысловая информация содержится в формантах, а мелодическая – периоде основного тона и временном распределении речи. Таким образом формантный подход требует меньшего объема памяти, но ему необходимо произвести больше вычислений, чтобы воспроизвести речевой сигнал, а также требуется знание моделей речеобразования [9].

Полный синтез по правилам

Этот способ также использует компилятивный и параметрический методы синтеза, слогов. Метод синтеза по правилам базируется на знании акустических и лингвистических ограничений и не использует непосредственно элементы речи. Для запоминания этой информации не требуется большого количества памяти, но, чтобы извлечь из нее параметры, необходимы знания эксперта. Анализ текста включает в себя определение фонетических, слоговых, морфемных и синтаксических форм, а также вычисление семантической информации [10]. Данный метод дает полную свободу моделирования параметров и позволяет воспроизводить любой текст, а также экономит память, однако синтезированная речь звучит значительно хуже реальной. Кроме того, такая система сложна в разработке. Поэтому был разработан метод синтеза по правилам с использованием предварительно запомненных отрезков речи. Обычно в качестве таких отрезков используются сегменты, содержащие половину согласного и половину примыкающего к нему гласного. Такой сегмент называется полуслогом. Качество такого синтеза не соответствует качеству реальной речи даже при использовании больших объемов памяти для хранения словоформ, однако такой метод синтеза речи является более качественным по сравнению с простым методом синтеза по правилам.

Синтезатор речи Asapela

Синтезатор речи, который получил название «Алёна» разработанный компанией Asapela. Работает на SAPI5, SAPI4 с частотой 22 КГц. По качеству синтезируемой речи «Алёна» опережает аналогичные русские голоса, имея намного более приятный тембр голоса и интонацию.

Программа поддерживает стандартные программные интерфейсы SAPI4 и SAPI5 и совместима со множеством речевых приложений для чтения с экрана. В наличии имеется полный набор инструментов для контроля параметров голоса, таких как скорость чтения, громкость, высота, паузы, настройки эквалайзера и режима чтения (чтение фразы, чтение отдельных слов и посимвольное чтение).

Синтезатор речи Demagog

Программа предназначена для чтения вслух текстовых файлов с использованием пакетов речевых функций Microsoft Speech API 4/5 (SAPI4 и SAPI5). Вместо чтения вслух текст может быть ускоренно записан в аудиофайл формата WAV или MP3. Таким образом можно создавать аудиокниги. Имеется возможность автоматического выделения тембром и громкостью прямой речи персонажей. Поддерживаются словари корректировки произношения популярного формата

DIS. Синтезатор поддерживает пять языков: английский, русский, украинский, эсперанто и испанский. При необходимости программа позволяет добавить новые языки.

Синтезатор речи Infovox 4

Infovox4 является инновационным программным обеспечением, позволяющим слабовидящим пользователям прослушивать текстовую информацию используя высококачественный синтезатор речи. Программное обеспечение Infovox4 представляет собой USB накопитель, позволяя пользователю подключиться к любому компьютеру без предварительной установки. Кроме того, программа может быть предустановлена на трех ПК. Предустановленные версии программы будут работать независимо от наличия флеш-накопителя. ПО включает в себя программы чтения с экрана, программу создания голосовых MP3-файлов, а также широкий диапазон выбора языков и голоса. Для каждого языка доступны две версии голосов преобразования текста в речь: голоса высокого качества High Quality (HQ) для обычного, приятного чтения и Colibri голоса с улучшенной разборчивостью речи на высокой скорости чтения. Программа обеспечивает поддержку 62 голосов на 26 языках, поддерживает SAPI4 и SAPI5 интерфейсы, обеспечивает интеграцию со многими программами экранного доступа. Синтезатор имеет меню на основе графического интерфейса, что облегчает навигацию при помощи программы экранного чтения, а также имеет возможность регулировки темпа речи, высоты тона (только для Colibri голосов), длительности пауз и настройки эквалайзера для каждого голоса. Возможно выбрать 3 доступных режима чтения: непрерывный, по словам и по буквам.

Синтезатор речи Балаболка

Программа предназначена для чтения вслух текстовых файлов. Для воспроизведения звуков человеческого голоса могут использоваться любые речевые синтезаторы, установленные на компьютере. Программа умеет читать вслух содержимое буфера обмена, показывать текст, содержащийся в AZW, AZW3, CHM, DjVu, DOC, DOCX, EML, EPUB, FB2, HTML, LIT, MOBI, ODS, ODT, PDB, PDF, PRC, RTF, TCR, WPD, XLS, XLSX файлах, произносить набираемый на клавиатуре текст. Текст может быть сохранен в виде аудиофайла (поддерживаются форматы WAV и MP3). Балаболка умеет работать с различными версиями пакета речевых функций Microsoft Speech API (SAPI). Он позволяет изменять скорость, тембр и громкость речи.

Выбор синтезатора речи для подсистемы оповещения интеллектуального зала совещаний

Исходя из вышесказанного, проведем сравнение перечисленных синтезаторов речи (табл. 1) по следующим критериям: доступность (цена), количество используемых голосов, тип интерфейса, наличие дополнительных настроек.

Таблица 1

Сравнение синтезаторов речи

Table 1

Comparison of synthesizers of the speech

Вид синтезатора речи	Критерии оценки			
	Доступность	Количество используемых голосов	Тип интерфейса	Наличие дополнительных настроек
Acapela	Платная	2	SAPI4/5	Да
Demagog	Бесплатная	2*	SAPI4/5	Да
Infovox 4	Платная	62	SAPI4/5	Да
Балаболка	Бесплатная	2*	SAPI4/5	Да

*-возможна установка дополнительных голосов

Из представленных данных видно, что платные программы обладают расширенным функционалом относительно бесплатных, но для решения поставленной задачи это преимущество

не является значимым, поэтому целесообразнее использовать один из бесплатных синтезаторов речи с открытым кодом Demagog или Балаболка.

ЗАКЛЮЧЕНИЕ

Анализ существующих синтезаторов речи показал, что не смотря на их большое многообразие, все они базируются на использовании одного и того же речевого ядра и отличаются лишь функциональными возможностями и качеством воспроизводимой речи. Однако применительно к подсистеме оповещения интеллектуального зала данные факторы не являются важными. Важной является необходимость модернизации программы под конкретную реализацию с учетом требований, предъявляемых заказчиком. Исходя из этого, можно сделать вывод о возможности использования в подсистеме оповещения интеллектуального зала совещаний бесплатного программного обеспечения с открытым кодом, такого как Балаболка и Demagog.

Работа выполнена при финансовой поддержке фонда РФФИ (проект № 18-07-00380).

Список литературы

1. Кузнецов Д.А. Предпосылки создания интеллектуального зала совещаний / Кузнецов Д.А., Офицеров А.И., Кузнецов А.В., Чистяков С.В., Басов О.О. // Научный результат. Информационные технологии. 2018. Том 3. Вып. 2. С. 44-50.
2. Нэкэшима Х. Руководство по интеллектуальному окружению и Умной окружающей среде / Агэдждэн Х. К., Аугусто Х. К. // Спрингер. 2010. 1294 с.
3. Ронжин Ал.Л. Система аудиовизуального мониторинга участников совещания в интеллектуальном зале / Ронжин Ал.Л., Ронжин Ан.Л. // Доклады ТУСУРа, № 1 (22), ч. 1, 2011. С. 153-157.
4. Филлингер А. Промежуточное программное обеспечение и Метрология для Распространяющегося будущего / Хэмчи Ай., Дегре С., Дидач Л., Роуз Т., Фискус Дж., Стэнфорд В // IEEE Распространяющиеся Вычислительные Мобильные и Повсеместные Системы. Издание 8, номер. 3, 2009. С. 74-83.
5. Рыбин С.В. Синтез речи / Рыбин С.В. // Учебное пособие по дисциплине «Синтез речи». Университет ИТМО. 2014. 92 с.
6. Пугач А.С. Сравнительный анализ методов синтеза речи / Пугач А.С. // Молодой ученый. 2016. 156 с.
7. Сорокин В.Н. Синтез речи / Сорокин В.Н. // М.: Наука. 1992. 392 с.
8. Пугач. А.С. Базовые принципы построения системы синтеза речи / Пугач А.С. // Молодой ученый. 2017. С. 16-19.
9. Фролов А.В. Синтез речи. Современные решения [Электронный ресурс]. – Режим доступа: URL: http://www.frolov-lib.ru/books/hi/ch07.html#_Тoc153187665 (Дата обращения 5.09.2018).
10. Рахимова М.И. Синтез и анализ лингвистической системы // Молодой ученый. 2016. С. 1159-1161.

References

1. Kuznetsov D. A. Prerequisites of creation of the intellectual hall Meetings / Kuznetsov D.A., Ofitserov A.I., Kuznetsov A.V., Chistyakov S.V., Basov O.O. //Scientific result. Information technologies. 2018. Volume 3. Issue 2. Page 44-50.
2. Nakashima H. Handbook of Ambient Intelligence and Smart Environments / Aghajan H. K., Augusto J. C. // Springer. 2010. 1294 p.
3. Ronzhin A. The system of audiovisual monitoring of participants of the meeting in the intellectual hall / Ronzhin Al., Ronzhin An. // Reports ТУСУРа, № 1 (22), p. 1, 2011, p. 153-157.
4. Fillinger A. Middleware and Metrology for the Pervasive Future /., Hamchi I., Degré S., Diduch L., Rose T., Fiscus J., Stanford V. // IEEE Pervasive Computing Mobile and Ubiquitous Systems. Vol. 8, num. 3, 2009. pp. 74-83.
5. Rybin S.V. Synthesis of speeches / Big fishes of S.V.//the Manual on discipline "Synthesis of the speech". ITMO university. 2014. 92 pages.

6. Pugach A.C. Comparative analysis of methods of synthesis speech / Pugach A.C. //Young scientist. 2016. 156 p.
7. Sorokin V.N. Synthesis Speech / Sorokin V.N.//M.: Science. 1992. 392 p.
8. Pugach A.C. Basic principles of creation of system of synthesis speech / Pugach A.C. //Young scientist. 2017. p. 16-19.
9. Frolov A.V. Synthesis of the speech. Modern decisions [An electronic resource]. – Access mode: URL: http://www.frolov-lib.ru/books/hi/ch07.html#_Toc153187665 (Date of the address 9/5/2018).
10. Rakhimova M.I. Synthesis and analysis of linguistic system//Young scientist. 2016. p. 1159-1161.

Кузнецов Денис Андреевич, студент Академии Федеральной службы охраны Российской Федерации
Кузнецов Андрей Викторович, кандидат технических наук, сотрудник Академии Федеральной службы охраны Российской Федерации

Тезин Александр Васильевич, кандидат технических наук, доцент, сотрудник Академии Федеральной службы охраны Российской Федерации

Басов Олег Олегович, доктор технических наук, профессор кафедры речевых информационных систем

Kuznetsov Denis Andreevich, student Academy of the Federal security service of the Russian Federation

Kuznetsov Andrey Viktorovich, candidate of technical sciences, Academy of the Federal security service of the Russian Federation

Tezin Aleksandr Vasil'evich, candidate of technical sciences, Associate Professor, Academy of the Federal security service of the Russian Federation

Basov Oleg Olegovich, doctor of technical sciences, Professor of Department of Speech Information Systems