

УДК 004.822:514

**В.А. Житко, Б.М. Лобанов**

## **ПРИКЛАДНЫЕ СИСТЕМЫ С РЕЧЕВЫМ ИНТЕРФЕЙСОМ**

*Рассматриваются три прикладные системы с речевым интерфейсом. Первые две из них предоставляют конечному пользователю возможность устно задавать вопрос и слышать ответ, создавая тем самым дополнение к традиционному вводу-выводу через клавиатуру и экран компьютера. Третья система (IntonTrainer) предоставляет пользователю возможность интерактивного голосового взаимодействия и предназначена для углубленного самообучения интонации устной речи.*

### **Введение**

Современный уровень развития вычислительной техники и ее повсеместное внедрение в человекомашинные системы управления делают актуальной организацию общения человека и компьютера в одной из наиболее удобных для человека форм – в форме речевого диалога на естественном языке. Принципиальными преимуществами использования речевого интерфейса в человекомашинных системах являются:

– удобство, естественность и простота общения, не требующие специальной подготовки, что существенно расширяет круг потенциальных пользователей вычислительных систем и повышает эффективность их применения;

– разгрузка зрительного канала при выводе информации и устранение ручных манипуляций при вводе, что увеличивает оперативность взаимодействия с компьютером и уменьшает число ошибок оператора;

– возможность использования в качестве терминалов телефонных аппаратов и существующей сети телефонной связи, что позволяет создавать системы массового обслуживания, в том числе с выходом в Интернет.

Создание речевого интерфейса (РИ) для широкого спектра прикладных человекомашинных систем наталкивается на значительные трудности, связанные с разработкой наукоемкого программного обеспечения (ПО) высокой сложности. Разработанные к настоящему времени речевые технологии достигли высокого уровня развития, что позволяет использовать различные готовые технические решения как компоненты при проектировании ПО разнообразных прикладных человекомашинных систем с речевым интерфейсом.

### **1. Компонентный подход при проектировании естественно-языкового интерфейса**

В связи с динамичным развитием рынка ПО большое внимание уделяется вопросам снижения сроков и стоимости его разработки. Это достигается целым комплексом мер, включающим в себя как технические решения, так и административные. Одним из элементов этого комплекса мер является использование готовых компонентов в разработке ПО. Такой подход называется компонентно-ориентированным программированием (англ. component-oriented programming, COP).

Программные компоненты – это повторно используемые части программных систем, предназначенные для достижения четких целей [1]. Компоненты могут применяться в различных системах, условиях и общих процессах, что накладывает на них повышенные требования как со стороны реализации, так и со стороны проектирования. Для компонентов важной является корректность выбора их функционала, а объем функциональных возможностей должен быть достаточным для выполнения независимой задачи и не превышать объем, затрудняющий разворачивание компонента в целевой системе.

Применение компонентов при проектировании и разработке программных продуктов позволяет повысить качество, ускорить разработку и сократить время до выхода на рынок данного программного продукта [2]. СОР предоставляет эффективный механизм управления структурой программных систем с возрастающей сложностью (управление сложностью), где важным элементом является функция управления не только структурой системы, но и каждым компонентом в отдельности, возможность адаптации компонентов к изменяющимся требованиям (управление изменениями). При этом созданный компонент можно использовать повторно в различных системах без необходимости его дополнительной адаптации [3, 4].

Ниже рассмотрены примеры разработки прикладных компьютерных программ с речевым интерфейсом (приложений) при использовании компонентного подхода к их проектированию. Приложения реализованы под операционные системы Windows (7, 8, 10) и Linux (Debian, Ubuntu).

## 2. Приложение с элементами речевого диалога

Целью разработки данного программного приложения с элементами речевого диалога является предоставление рядовым пользователям (в том числе детям дошкольного и школьного возраста) возможности ознакомления в игровой форме с современным речевым интерфейсом, включающим автоматическое распознавание и синтез речи. В рамках приложения разработан программный комплекс, объединивший в себе режимы работы, которые демонстрируют различные аспекты использования речевого интерфейса: режим «Попугай» – повтор синтезированным голосом любых сказанных пользователем фраз; «Сказка» – простейший диалог с пользователем (например, детская сказка про белого бычка); «Переводчик» – устно-речевой перевод для английского, белорусского и русского языков; «Калькулятор» – возможность производить арифметические действия, задаваемые голосовыми командами.

В рассматриваемой системе использованы:

– собственные разработки авторов, к которым относятся компоненты записи и детектирования речи, работы с синтезаторами белорусской и русской речи [5, 6], графический пользовательский интерфейс и логические компоненты работы системы;

– сторонние разработки, находящиеся в свободном доступе в Интернете, а именно компоненты работы с Google Voice [7] и Google Translate [8].

Общая структурная схема системы, реализующей названные виды режимов работы, изображена на рис. 1.

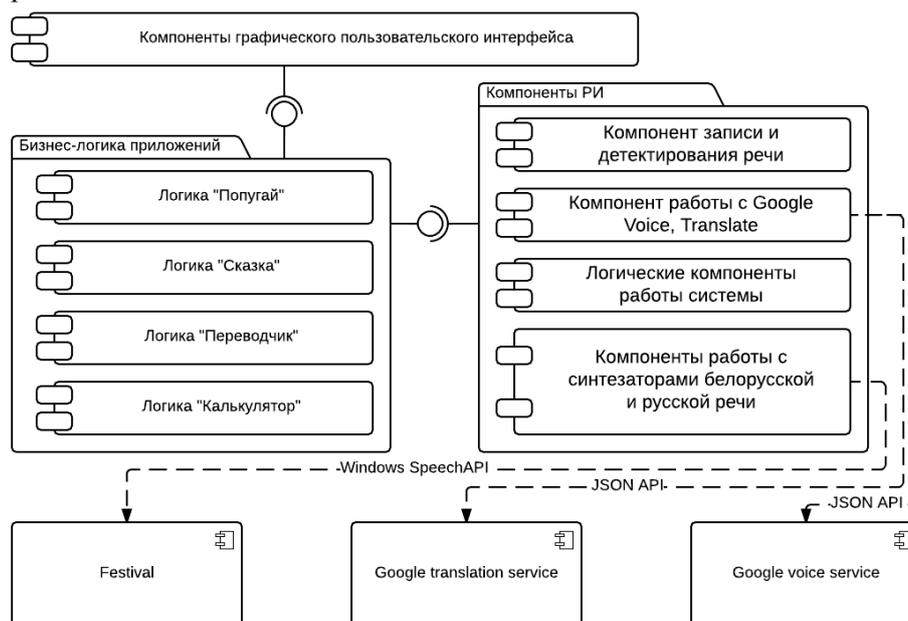


Рис. 1. Структурная схема игрового приложения

Алгоритм работы приложения, который описывает обработку произносимого сигнала и логику взаимодействия с компонентами распознавания речи, представлен на рис. 2.

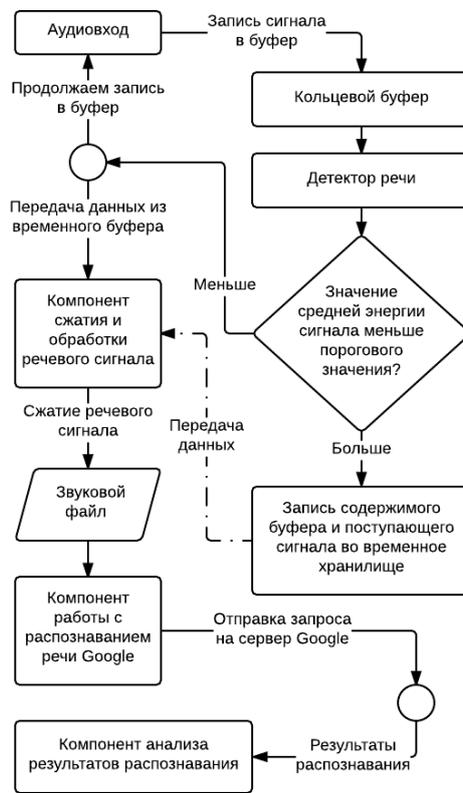


Рис. 2. Алгоритм работы игрового приложения

Главное окно пользовательского интерфейса приложения (рис. 3) позволяет выбрать один из режимов работы.

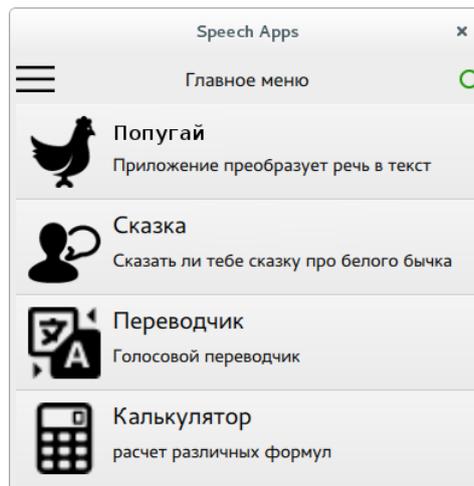


Рис. 3. Главное окно игрового приложения

В режиме «Попугай» (рис. 4) система повторяет синтезированным голосом любые сказанные пользователем фразы. При этом происходит их автоматическое распознавание, преобразование в текст и синтез речи по распознанному тексту. В этом режиме пользователь имеет возможность

произносить произвольные фразы и прослушивать результат распознавания синтезированным голосом, усваивая таким образом специфику работы с речевым интерфейсом системы.



Рис. 4. Пример работы игрового приложения в режиме «Попугай»

В режиме «Сказка» (рис. 5) демонстрируется элементарный диалог в стиле докучной сказки про белого бычка, рассказываемой с большим успехом родителями детям дошкольного возраста.

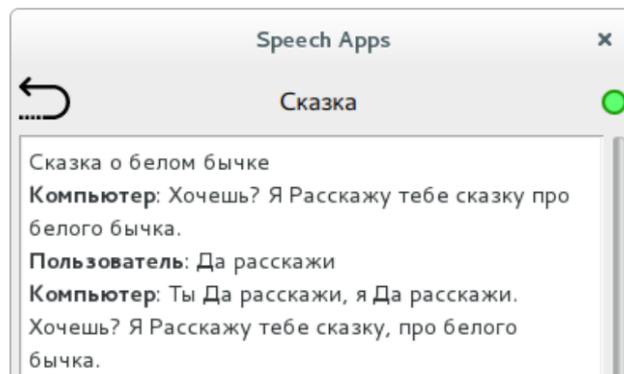


Рис. 5. Пример работы игрового приложения в режиме «Сказка»

Пример текста диалога, озвучиваемого синтезатором, где показаны распознанные компьютером реплики пользователя, приводится ниже.

Компьютер: Хочешь? Я расскажу тебе сказку про белого бычка.

Человек: Да.

Компьютер: Ты Да, я Да. Хочешь? Я расскажу тебе сказку про белого бычка.

Человек: Расскажи.

Компьютер: Ты Расскажи, я Расскажи. Хочешь? Я расскажу тебе сказку про белого бычка.

Человек: Не хочу сказку.

Компьютер: Ты Не хочу сказку, я Не хочу сказку. Хочешь? Я расскажу тебе сказку про белого бычка.

Человек: Дурак.

Компьютер: Ты Дурак, я Дурак. Хочешь? Я расскажу тебе сказку про белого бычка.

Человек: Замолчи наконец.

Компьютер: Ты Замолчи наконец, я Замолчи наконец. Хочешь? Я расскажу тебе сказку про белого бычка.

В режиме «Переводчик» (рис. 6) осуществляется распознавание речи на английском или русском языке, преобразование речи в текст, перевод распознанного текста на английский, белорусский или русский язык. Полученный текст отправляется в соответствии с указанным пользователем языком на вход одного из синтезаторов речи.

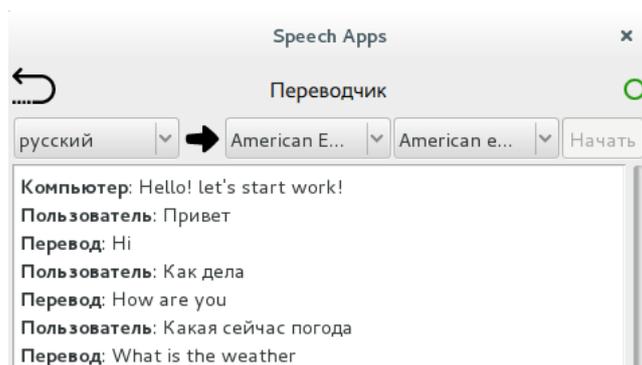


Рис. 6. Пример работы игрового приложения в режиме «Переводчик»

В режиме «Калькулятор» (рис. 7) пользователь может осуществлять арифметические действия, задаваемые голосовыми командами. Вывод результата осуществляется голосом в виде синтезированной речи.



Рис. 7. Пример работы игрового приложения в режиме «Калькулятор»

Более подробные сведения об особенностях программной реализации данного игрового приложения изложены в работе [9].

### 3. Справочная система по правилам дорожного движения

Для недостаточно опытного водителя было бы очень полезным получить оперативную справку по правилам дорожного движения (ПДД) и выбору конкретного действия в возникшей ситуации без необходимости остановки для просмотра справочного пособия. Решение такой проблемы становится возможным с помощью справочной системы (ПДД-приложения), в которой для водителя реализована возможность голосового запроса, его анализа и получения голосового ответа, рекомендующего его дальнейшие действия в соответствии с ПДД.

В качестве предметной области для системы были выбраны реальные ПДД, действующие на территории Республики Беларусь. Отметим, что набор ПДД характеризуется структурированностью исходных данных, что существенно облегчает решение поставленной задачи. В процессе взаимодействия с пользователем данная система в качестве речевого ответа производит выбор текста наиболее релевантного ПДД и, если необходимо, задает уточняющие вопросы, снимающие возможную неоднозначность запроса.

База ПДД представлена в виде набора отдельных правил с указанием дополнительных ключевых слов для конкретного правила. Существует также список всех ключевых слов предметной области. Цикл работы справочной системы начинается с ввода пользователем голосового запроса, который преобразуется в текст на естественном языке. По введенному в систему тексту строится его формальное описание в памяти системы. Все предшествующие результаты применяются при анализе последующих запросов, что позволяет системе сохранять ход диалога с пользователем и задавать уточняющие вопросы.

В описываемой системе использованы:

- собственные разработки авторов, к которым относятся компоненты записи и детектирования речи, компоненты работы с синтезаторами белорусской и русской речи [5, 6], разбора и анализа текста, графического пользовательского интерфейса; логические компоненты работы системы; база знаний ПДД; компонент принятых решений;

- сторонняя разработка – хорошо зарекомендовавшая себя система распознавания русской речи CMU Sphinx [10], находящаяся в свободном доступе в Интернете.

На рис. 8 представлена общая структурная схема ПДД-приложения, реализующего работу справочной системы с РИ.

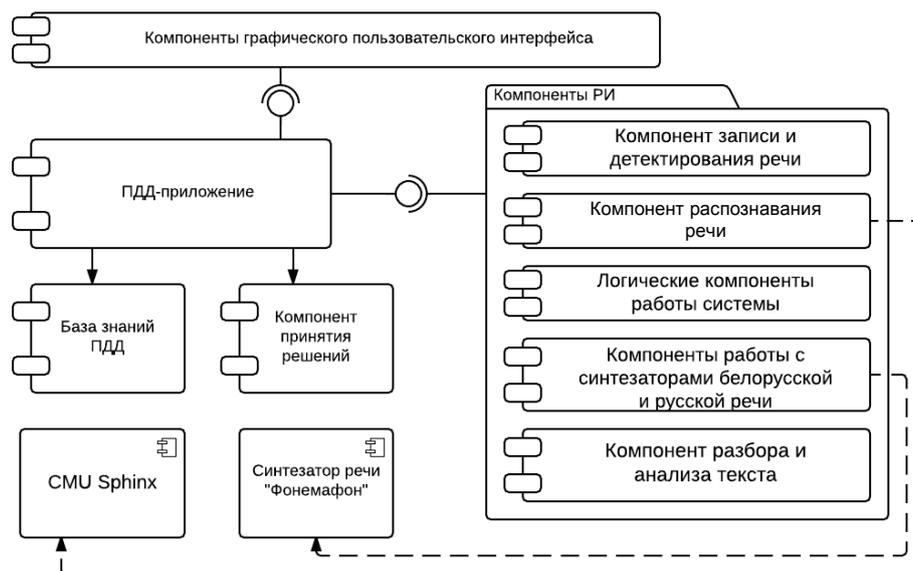


Рис. 8. Общая структурная схема ПДД-приложения

Более подробные сведения об особенностях программной реализации данного игрового приложения изложены в работе [11].

#### 4. Многоязычная система обучения интонации речи IntonTrainer

Интонация играет важную роль в восприятии речи человеком. Зачастую в зависимости от интонационного оформления логически идентичные высказывания могут иметь различное семантическое значение. Интонация представляет собой важнейшее средство передачи как собственно языковой, так и социокультурной информации. В практике преподавания существует мнение, что иноязычный акцент особенно ярко проявляется в интонации, поэтому при обучении и преподавании иностранных языков ей следует уделять особое внимание. Следует иметь в виду, что функции интонации в речи очень многообразны и нарушения в этой области могут привести к серьезным смысловым изменениям, а также создать неверное впечатление о говорящем человеке. Очень часто учащимся сложно контролировать адекватность своей интонации (особенно при отсутствии музыкального слуха). Существующие лингафонные курсы и оборудование обеспечивают только слуховую обратную связь контроля правильности интонирования речи, что явно недостаточно.

В мировой практике имеется устойчивая тенденция к освоению языка с помощью компьютера (Computer Assisted Language Learning, CALL). CALL является широко распространенной междисциплинарной областью, которая включает также подобласть обучения произношению на иностранных языках с использованием речевых технологий.

Описываемая в настоящей работе система IntonTrainer (рис. 9) предназначена для углубленного обучения интонации устной речи. Обучение основано на сравнении эталонных интонационных портретов заданного набора фраз с интонацией фраз, произносимых учащимися.

Акустическая база данных включает набор стандартных фраз, который достаточно полно покрывает многообразие интонационных эталонов и их основных вариаций, используемых в изучаемом языке. Система основана на интерактивном взаимодействии с обучаемой личностью, обеспечивая слуховую и зрительную обратную связь, а также численную оценку правильности интонирования фразы. В настоящее время система ориентирована на обучение интонации устной речи на русском или английском языках. Настройка на другие языки может быть достаточно просто осуществлена путем создания соответствующих акустических баз данных.

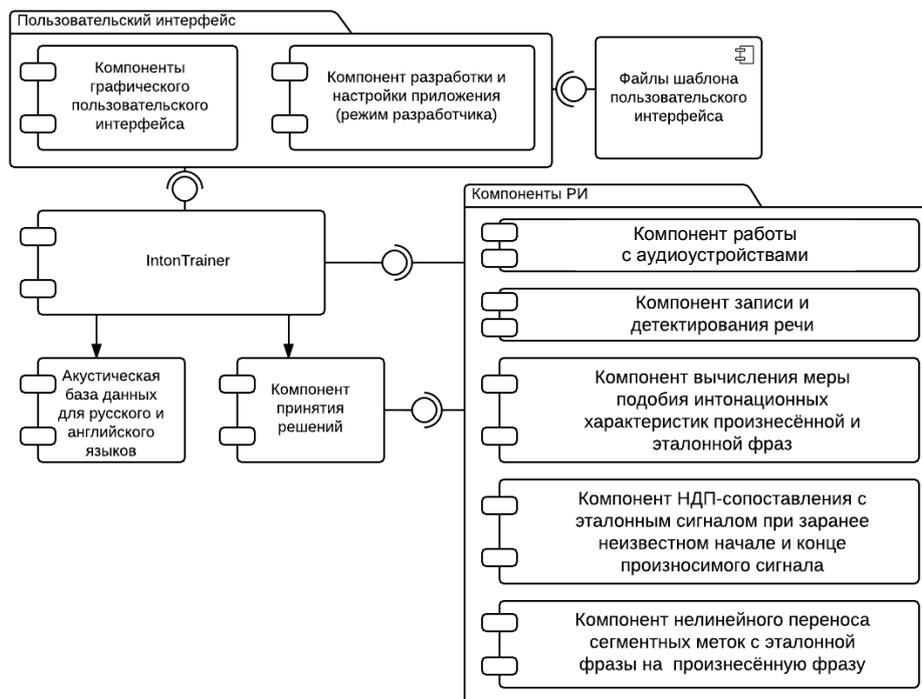


Рис. 9. Структурная схема приложения IntonTrainer

В системе IntonTrainer использованы:

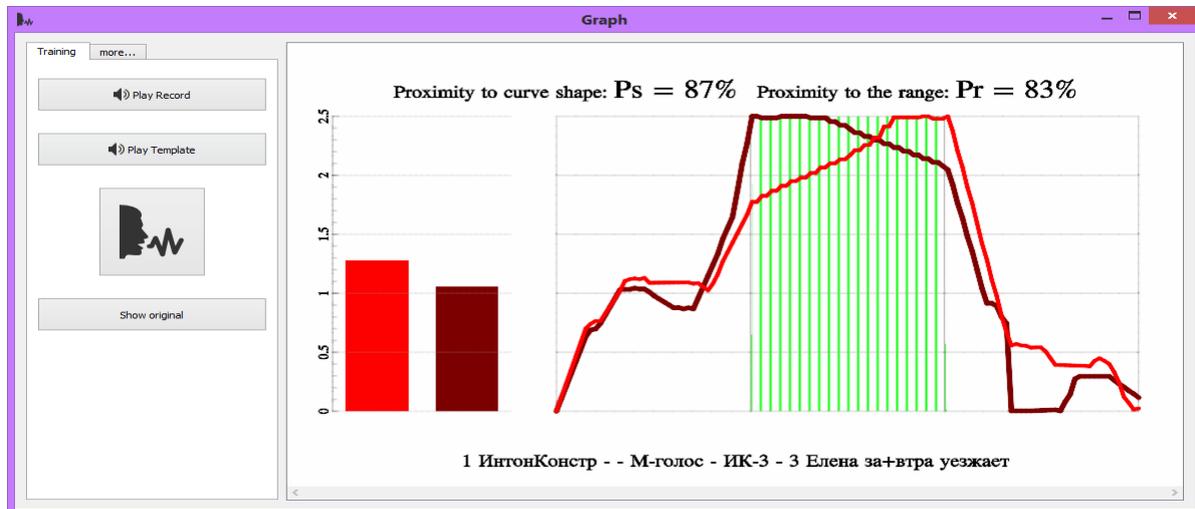
– собственные разработки авторов, к которым относятся компоненты работы с аудиоустройствами, записи и детектирования речи, непрерывного ДП-сопоставления (алгоритмы динамического программирования) с эталонным сигналом при заранее неизвестных начале и конце произносимого сигнала [12], нелинейного переноса сегментных меток с эталонной фразы на произнесенную фразу [6], вычисления меры подобиия интонационных характеристик произнесенной и эталонной фраз [13], графического пользовательского интерфейса; логические компоненты работы системы; акустическая база данных для русского и английского языков и компонент принятия решений;

– сторонняя разработка – хорошо зарекомендовавший себя программный комплекс анализа и обработки речевых сигналов, который находится в свободном доступе в Интернете [14].

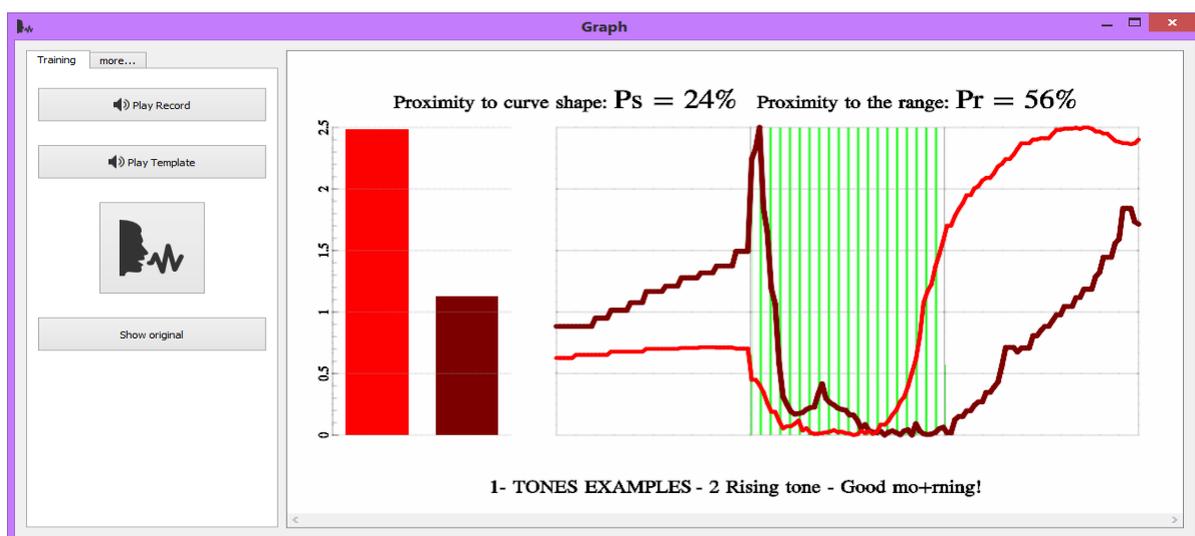
Прототип приложения написан под операционные системы Windows (7, 8, 10) и Linux (Debian, Ubuntu). В работе приложения используются библиотека Qt для реализации отладочного пользовательского интерфейса и ядра приложения и библиотека MathGL для вывода графиков. Для реализации основного пользовательского интерфейса применяется приложение, написанное на HTML/JS (используется ReactJS-фреймворк). Алгоритмы расчета и обработки данных написаны на C с использованием библиотек GNU Scientific Library, Speech Signal Processing Toolkit (SPTK) и OpenAL.

В процессе обучения пользователь оценивает степень правильности интонирования фразы визуально в виде графиков изменения частоты основного тона и в виде столбчатой диаграммы диапазонов изменения данной частоты эталонной и произнесенной фраз, а также в виде численных оценок степени совпадения ее кривых ( $P_s$ ) и диапазонов их измене-

ния (Pr). При правильном интонировании фразы пользователь видит высокие оценки Ps/Pr и схожие графики универсального мелодического портрета (УМП) (рис. 10, а). В случае ошибочного интонирования фразы пользователь может увидеть, в каком месте фразы он допустил ошибку по расхождению графиков УМП (рис. 10, б).



а)



б)

Рис. 10. Рабочее окно приложения IntonTrainer: а) на русском языке; б) на английском языке

Более подробные сведения об особенностях программной реализации данного приложения изложены в [15]. Для читателей доступны также две демоверсии системы IntonTrainer, ориентированные на изучение интонаций русского и английского языков (<https://intontrainer.by/>).

## Заключение

Рассмотренные первые два примера прикладных систем с речевым интерфейсом предоставляют конечному пользователю возможность устно задавать вопрос и слышать ответ от системы, создавая дополнение к традиционному вводу-выводу через клавиатуру и экран компьютера, что делает интерфейс более естественным и дружелюбным для пользователя.

Система интонационного тренинга IntonTrainer сможет найти применение не только при обучении интонации иностранной речи, но и в практике обучения нормативной интонации дикторов радио и телевидения, операторов колл-центров, а также людей других профессий, где требуется идеальное владение устной речью.

Представленный набор спроектированных и реализованных компонентов речевого интерфейса со строго заданным функционалом позволяет упростить разработку других подобных систем с речевым интерфейсом, а в дальнейшем упрощает их сопровождение и отладку.

Разбиение системы на отдельные независимые компоненты дает возможность интеграции речевого интерфейса в различные сторонние разработки и проекты, а также интеграции различных подходов и методов в рамках одного проекта, эффективно используя их лучшие стороны.

### Список литературы

1. Szyperski, C. Component Software: Beyond Object-Oriented Programming / C. Szyperski. – N. Y. : Addison-Wesley, 1998. – 411 p.
2. Cheesman, J. UML Components: A Simple Process for Specifying Component-Based Software / J. Cheesman, J. Daniels. – Boston : Addison-Wesley, 2000. – 208 p.
3. Wang, A. Ju. An Component-Oriented Programming / A. Ju. Wang, K. Qian. – 1st ed. – N. Y. : Addison-Wesley, 2005. – 336 p.
4. Brown, A.W. The current state of CBSE / A.W. Brown, K.C. Wallnau // IEEE Software, sept. 1998. – Los Alamitos, CA, USA, 1998. – Vol. 15, iss. 5. – P. 37–46.
5. Гецевич, Ю.С. Система синтеза белорусской речи по тексту / Ю.С. Гецевич, Б.М. Лобанов // Речевые технологии. – 2010. – № 1. – С. 91–100.
6. Лобанов, Б.М. Компьютерный синтез и клонирование речи / Б.М. Лобанов, Л.И. Цирульник. – Минск : Бел. наука, 2008. – 344 с.
7. Google Cloud Speech API [Electronic resource]. – 2016. – Mode of access : <https://cloud.google.com/speech/docs>. – Date of access : 20.05.2017.
8. Google Cloud Translation API [Electronic resource]. – 2016. – Mode of access : <https://cloud.google.com/translate/docs>. – Date of access : 20.05.2017.
9. Житко, В.А. Микросервисный подход к проектированию речевых интерфейсов / В.А. Житко, Б.М. Лобанов // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS–2016) : мат. VI Междунар. науч.-техн. конф. (Минск, 18–20 февр. 2016 г.); под ред. В.В. Голенкова. – Минск : БГУИР, 2016. – С. 383–388.
10. CMU Sphinx Speech Recognition Toolkit [Electronic resource]. – 2016. – Mode of access : <https://sourceforge.net/projects/cmuspinx>. – Date of access : 20.05.2017.
11. Житко, В.А. Справочная система с речевым интерфейсом / В.А. Житко, Ю.С. Гецевич, Б.М. Лобанов // Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS–2013) : мат. III Междунар. науч.-техн. конф. (Минск, 21–23 февр. 2013 г.) / редкол. : В.В. Голенков (отв. ред.) [и др.]. – Минск : БГУИР, 2013. – С. 505–510.
12. Lobanov, B.M. Automatic Recognition of Sounds Combination in Running Speech Signal / B.M. Lobanov, G.S. Slucker, A.P. Tizik // Автоматическое распознавание и синтез речи : сб. науч. тр. – Минск : Ин-т техн. кибернетики НАН Беларуси, 2000. – С. 78–88.
13. Lobanov, B.M. Comparison of Melodic Portraits of English and Russian Dialogic Phrases / B.M. Lobanov // Компьютерная лингвистика и интеллектуальные технологии : по мат. ежегодной Междунар. конф. «Диалог» (Москва, 1–4 июня 2016 г.). – М. : РГГУ, 2016. – Вып. 15(22). – С. 382–392.
14. Speech Signal Processing Toolkit (SPTK) [Electronic resource]. – 2016. – Mode of access : <https://sourceforge.net/projects/sp-tk>. – Date of access : 20.05.2017.

---

15. Lobanov, B.M. A prototype of the computer system for speech intonation training / B.M. Lobanov, V.A. Zhitko // Open Semantic Technologies for Intelligent Systems (OSTIS-2017) : мат. VI Междунар. науч.-техн. конф. (Минск, 18–20 февр. 2017 г.) ; под ред. В.В. Голенкова. – Минск : БГУИР, 2017. – С. 383–388.

Поступила 21.06.2017

*Объединенный институт проблем  
информатики НАН Беларуси,  
Минск, Сурганова, 6  
e-mail: lobanov@newman.bas-net.by,  
zhitko.vladimir@gmail.com*

**V.A. Zhitko, B.M. Lobanov**

### **SOME EXAMPLES OF APPLIED SYSTEMS WITH SPEECH INTERFACE**

Three examples of applied systems with a speech interface are considered in the article. The first two of these provide the end user with the opportunity to ask verbally the question and to hear the response from the system, creating an addition to the traditional I / O via the keyboard and computer screen. The third example, the «IntonTrainer» system, provides the user with the possibility of voice interaction and is designed for in-depth self-learning of the intonation of oral speech.