

УДК 681.3

Синтезаторы речи с MS SAPI интерфейсом для скринридеров

Г.В. Юсим

Современные специализированные компьютерные системы, ориентированные на пользователей с полной потерей зрения, которые не могут непосредственно использовать визуальную информацию для взаимодействия с компьютером, чаще всего предусматривают в своём составе программный модуль чтения экрана (screen reader). Этот модуль воспринимает информацию, отображаемую на экране дисплея, и делает её доступной путём вывода посредством таких устройств как синтезатор речи или регенерируемый брайлевский дисплей. При этом для управления компьютером пользователь использует обычную или специальную клавиатуру или же устройство голосового ввода.

К числу таких систем относятся Hal 95 for Windows 95 фирмы Dolphin Systems, JAWS for Windows фирмы Freedom Scientific., outSPOKEN for Windows фирмы ALVA Access Group Inc., VIRGO NT и VIRGO 4 фирмы BAUM Products GmbH, WINDOTS фирмы F.H. Papenmeier GmbH & Co и др.

Среди этих систем наиболее гибким и универсальным инструментом для работы инвалидов зрения с компьютером является система JAWS, которая в настоящее время поддерживается фирмой Freedom Scientific. С помощью JAWS незрячий пользователь имеет практически ту же функциональность и возможность такой же быстрой работы на компьютере, как и зрячий пользователь. За счёт развитых средств для написания скриптов JAWS обеспечивает эффективную работу с произвольными программами под Windows и, в частности, с программами Microsoft Office, Internet Explorer, FineReader и др. Важной особенностью JAWS является то, что в нём основной упор сделан на использование намного более дешёвого (чем тактильного) озвученного интерфейса. При этом система JAWS обладает таким достоинством как стандартизация средств подключения к ней синтезаторов речи. В частности, в JAWS реализован стандартный MS Speech API интерфейс,

предложенный фирмой Microsoft для использования в разработках в области анализа и синтеза речи [1].

Отметим, что для успешного применения JAWS необходимы качественные синтезаторы речи, которые обеспечивают выразительное воспроизведение произвольного текста и управление темпом речи в зависимости от характера текста и уровня подготовки пользователя. Отсутствие же до последнего времени украинско - и русскоязычных синтезаторов речи, которые удовлетворяли бы требованиям стандартов, предъявляемым к TTS (text-to-speech) машинам, затрудняет распространение JAWS и других аналогичных систем, как в Украине, так и в других странах бывшего Советского Союза.

Наиболее качественным русскоязычным синтезатором является, по нашему мнению, синтезатор "Голосовая мышь" Клуба голосовых технологий (г. Москва), реализованный на базе разработанного на факультете лингвистики МГУ метода компиляции и синтеза по коэффициентам линейного прогноза (КЛП-синтеза). К сожалению, однако, этот синтезатор не удовлетворяет требованиям упомянутых стандартов и, в частности, в нём отсутствуют интерфейсы, необходимые для использования синтезатора вместе с наиболее эффективными программными средствами, разработанными для незрячих пользователей. С другой стороны, зарубежные русскоязычные синтезаторы, которые появились в последнее время - "Nicolai"(французской фирмы Elan), "Boris" и "Svetlana" (Microsoft), значительно уступают по своим качественным характеристикам воспроизведения речи первому из указанных синтезаторов.

Разработанные в Общественном Объединении "Окно в мир" (г. Киев) украинско – и русскоязычные синтезаторы "Vikno" [2,3] так же, как и "Голосовая мышь", используют в качестве базы метод компиляции и КЛП-синтеза. При этом синтезаторы реализованы как TTS –машины с архитектурой и обязательными интерфейсами, соответствующими требованиям MS Speech API. Каждый из этих синтезаторов состоит из двух основных объектов - Engine COM-объекта, который обеспечивает режим процессирования текста в озвученную речь и Engine Enumerator COM-объекта, который перечисляет режимы, поддерживаемые TTS-машиной. Два других обязательных объекта, необходимые для функционирования TTS-машины (Text-to-speech Enumerator COM-объект, который перечисляет TTS-машины, установленные в

операционной системе, и Multimedia Audio-Destination COM-объект, который позволяет приложению создать аудио-целевой объект) входят непосредственно в состав MS Speech API [1].

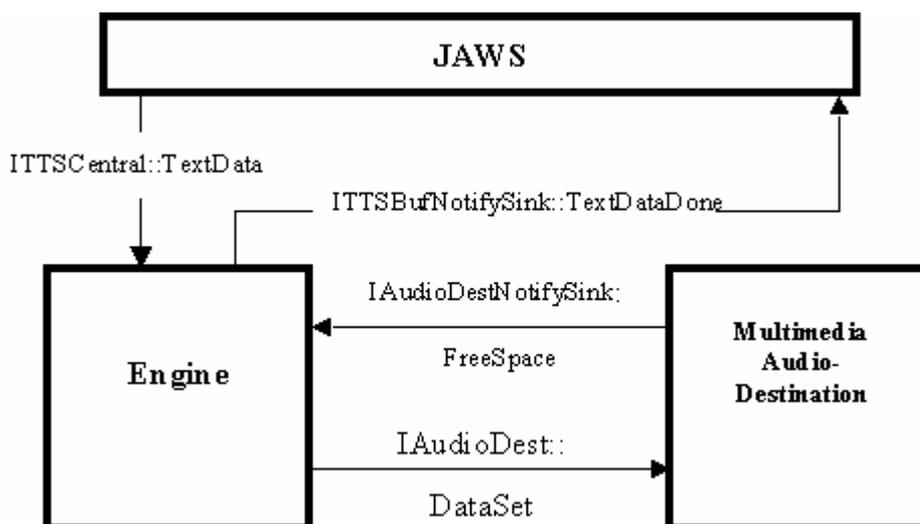


Рис.1

Упрощённая схема (рис.1) показывает взаимодействие приложения (в нашем случае JAWS) с синтезатором при воспроизведении текстовой информации. С помощью функции ITTSCentral::TextData JAWS передаёт объекту Engine текст, который запоминается в одном или нескольких буферах. Engine процессирует текст из буфера, передавая порциями преобразованный в цифровую форму речевой сигнал с помощью функции IAudioDest::DataSet. При этом посредством функции IaudioDestNotifySink::FreeSpace проверяется наличие свободного места в буфере Multimedia Audio-Destination - объекта, который последовательно опустошается в процессе воспроизведения текста. С помощью функции ITTSBufNotifySink::TextDataDone JAWS информируется о том, что закончено процессирование очередного входного буфера.

Как уже отмечалось, эта схема является упрощённой. На самом деле для синхронизации работы JAWS с синтезатором при непрерывном воспроизведении текста (по строкам, предложениям или абзацам) используются более сложные механизмы, связанные, в частности, с обработкой маркеров, которые JAWS привязывает к словам воспроизводимого текста. Кроме того, поскольку в соответствии с методом КЛП-синтеза процессирование текста и его мелодичное оформление осуществляется по так называемым синтагмам (отрезкам текста, ограниченных разделительными знаками), то алгоритм синхронизации обработки учитывает и это обстоятельство.

Разработанные синтезаторы имеют все обязательные интерфейсы и соответствующие им функции-члены с поддержкой как ANSI так и Unicode реализаций каждого из этих интерфейсов.

Перечислим основные интерфейсы с кратким описанием выполняемых ими функций.

ITTSAttributes.

Управляет атрибутами TTS-машины, такими, в частности, как тембр, скорость речи, громкость голоса.

ITTSBufNotifySink.

Используется TTS-машиной, чтобы сообщать приложению об изменениях в буфере, который содержит воспроизводимый текст.

ITTSCentral.

Управляет воспроизведением текста TTS-машиной, включая отправку текста машине, остановку и продолжение воспроизведения, получения информации о тексте и регистрацию сообщений. Этот интерфейс предусматривает также обработку специальных управляющих тегов (tags), которые могут вставляться в текст с целью управления его воспроизведением. Это такие теги как Pau, Pit, Spd, Vol, Mrk, которые позволяют в процессе воспроизведения вводить паузы заданной длительности, изменять тембр, скорость, громкость, а также информировать приложение о появлении маркеров. Эти управляющие теги могут быть вставлены в текст не только заранее, но и "на лету" (то есть непосредственно в процессе воспроизведения текста), для чего в интерфейсе предусмотрена функция-член **ITTSCentral::Inject.**

ITTSDialogs.

Отображает Windows - диалоговые окошки, которые позволяют пользователю получать информацию о TTS-машине, дают доступ к управляющим элементам и редактировать словарь ударений (рис.2).

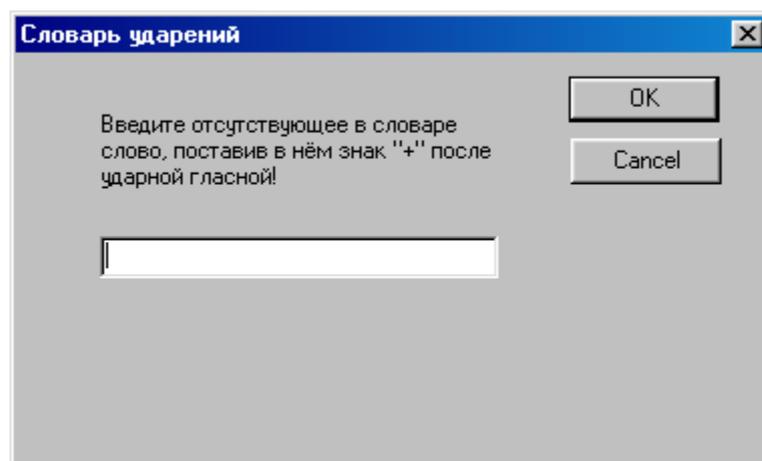


Рис.2

ITTSNotifySink.

Используется TTS-машиной для сообщения приложению о старте или остановке воспроизведения речи.

ITTSEnum.

Перечисляет и выбирает TTS-режимы.

Поскольку все интерфейсы реализованы в TTS-машине в виде COM объектов, то каждый такой объект поддерживает также IUnknown-интерфейс. Как известно [1], такой интерфейс содежит в себе три обязательных метода - Queryinterface, AddRef i Release.

Следует отметить, что использование объектно-ориентированного подхода и OLE - технологии позволяет разработанной TTS-машине иметь множественные реализации, допуская одновременную работу с несколькими приложениями.

Важной особенностью русскоязычного синтезатора является наличие в нём словаря ударений слов достаточно полного словарного запаса русского языка, который содержится в известном словаре Зализняка (более 2 млн. слов). Электронная версия этого словаря требует 32 мб. памяти. Однако, разработанный метод сжатия информации позволил ограничиться памятью всего в 1мб., обеспечив возможность работы со словарём в реальном масштабе времени

Программная реализация предусматривает возможность пополнения словаря путём вызова стандартной интерфейсной функции ITTSDialogs::LexiconDlg. К сожалению, в JAWS не предусмотрена возможность вызова этой функции. Поэтому сейчас для пополнения словаря предлагается использовать программу Attstest.exe (рис.3), которая входит в состав инструментария MS Speech API и предназначена для тестирования TTS-машин.

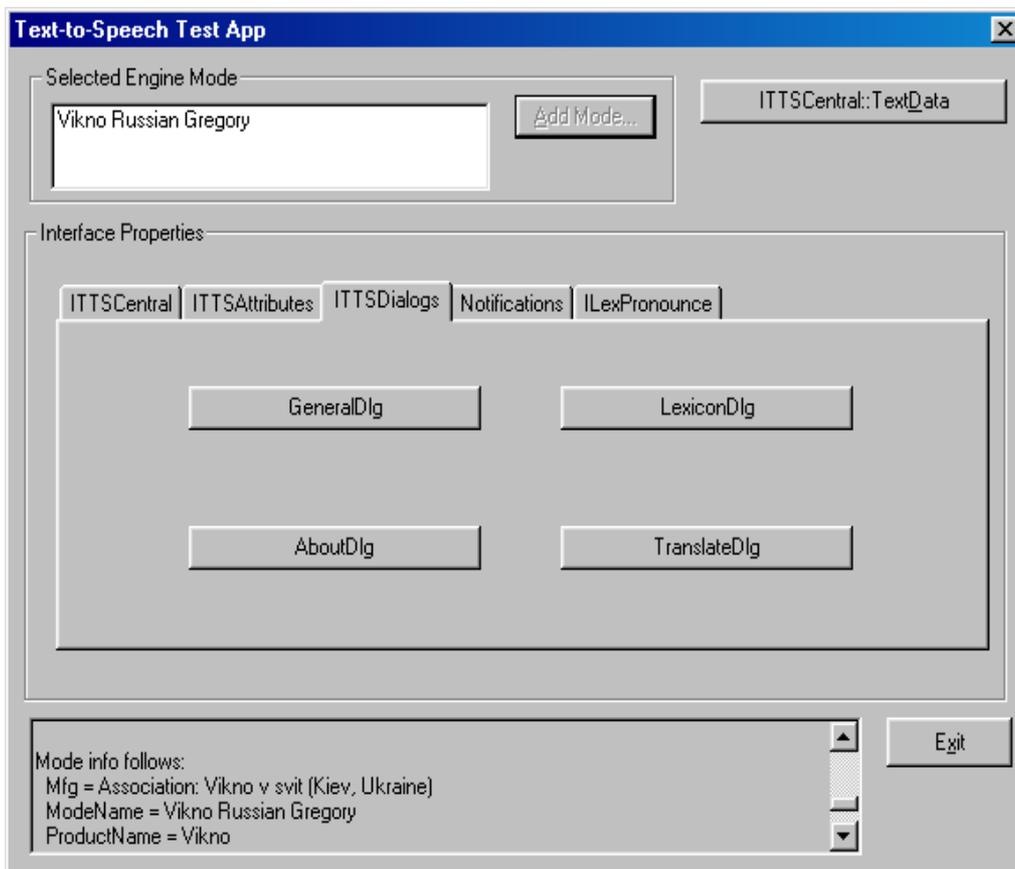


Рис. 3

Для чтения английского текста в синтезаторе предусмотрены следующие возможности:

1. Нахождение в тексте определённых слов и фраз и замена их русским переводом с использованием специального словаря, который может пополняться.
2. Транскрибирование с глубиной просмотра до 4-х букв.
3. Спеллинг, решение о котором принимается на основе соотношения гласных и заглавных букв, а также их месторасположения в слове.

Для подключения синтезатора, например, к программе JAWS4.51 необходимо добавить в файл Jfw.ini, который находится в каталоге ..\JAWS4.51, и имеет вид :

```
[Synthesizers]
Synth1Name=eloq
Synth1LongName=Eloquence for JFW
Synth1Driver=eloq
Synth2Name=Microsoft
Synth2LongName=Microsoft Text-to-Speech Engine
Synth2Driver=mssapi
Synth3Name=NoSpeech
Synth3LongName=No Speech
Synth3Driver=nospeech
[Options]
AskRunJAWSFirst=0
```

в раздел [Synthesizers] следующие три строки:

```
Synth4Name=Vikno
Synth4LongName= Russian Text-to-Speech Engine
Synth4Driver=mssapi
```

Разработанный русскоязычный синтезатор [2] был использован в качестве основы для создания украинскоязычного синтезатора [3]. При этом практически без изменений сохранилась внутренняя структура синтезатора, архитектура его интерфейсов и алгоритм взаимодействия синтезатора с приложениями. Изменения связаны, главным образом, с лингвистическим и информационным обеспечением синтеза, который учитывает специфические особенности украинского языка. К числу таких изменений относятся следующие:

- введения в состав аллофонной базы 17 новых единиц, что связано с особенностями выговора украинских букв "г" и "ц" в различных контекстах;
- существенная переработка правил буквенно-кодовых преобразований, объединённых в блоках транскрипции и аллофонного кодирования;

- введение блока транскрибирования украинских текстов в соответствующие русскоязычные эквиваленты;
- формирование и использование украинского словаря ударений;
- использование англо-украинского словаря для перевода характерных англоязычных вставок в украинских текстах.

Для подключения украинскоязычного синтезатора в систему JAWS 4.51 необходимо добавить в указанный выше файл Jfw.ini в раздел [Synthesizers] следующие три строки:

```
Synth5Name=Ukraine
Synth5LongName=Ukrainian Text-to-Speech Engine
Synth5Driver=mssapi
```

В заключение отметим, что разработанные синтезаторы функционируют в операционных системах Windows 95\98\Me\NT\2000\XP.

Список литературы:

1. Microsoft Speech SDK 4.0 , ©1995-1998 Microsoft Corporation.
2. Юсім Г.В. Комп'ютерна програма "Синтезатор російської мови з стандартним MS SAPI-інтерфейсом". Свідоцтво про реєстрацію авторського права на твір № 6144, 02.09.2002. Міністерство освіти і науки України. Державний департамент інтелектуальної власності.
3. Юсім Г.В., Терзян Т.К. Комп'ютерна програма "Синтезатор української мови з MS SAPI-інтерфейсом". Свідоцтво про реєстрацію авторського права на твір № 7112, 12.02.2003. Міністерство освіти і науки України. Державний департамент інтелектуальної власності.

Синтезаторы речи с MS SAPI интерфейсом для скринридеров

Аннотация

Рассматриваются украинско – и русскоязычная программы-синтезаторы речи, предназначенные для работы со специализированными средствами, ориентированными на незрячих

пользователей. Программы позволяют озвучивать произвольные тексты, написанные на украинском и русском языках с возможными англоязычными вставками. Программы содержат все обязательные интерфейсы, соответствующие требованиям стандарта MS SpeechAPI, с поддержкой как ANSI так и Unicode реализаций каждого из этих интерфейсов. Работают в операционных системах Windows 95/98/Me/2000/NT/XP.

Синтезаторы мовлення з MS SAPI інтерфейсом для скрінрідерів

Анотація

Розглядаються українсько – та російськомовна програми-синтезатори мовлення, призначені для роботи зі спеціалізованими засобами, що орієнтовані на незрячих користувачів. Програми дозволяють озвучувати довільні тексти, написані українською та російською мовами з можливими англомовними вставками. Програми мають усі обов'язкові інтерфейси, що відповідають вимогам стандарту MS SpeechAPI, з підтримкою як ANSI так і Unicode реалізацій кожного з цих інтерфейсів. Працюють в операційних системах Windows 95/98/Me/2000/NT/XP.