

УДК 621.391

Б.М. Лобанов, В.А. Житко

ПРИМЕНЕНИЕ ОБЛАЧНЫХ ТЕХНОЛОГИЙ ПРИ РАСПОЗНАВАНИИ РЕЧИ

Рассматриваются особенности применения облачных интернет-технологий в задачах автоматического распознавания речи. Кратко описываются положительные особенности и основные модели построения современных облачных интернет-технологий. Рассматриваются существующие методы, технологии и архитектура систем автоматического распознавания речи, а также особенности облачной технологии распознавания речи, предоставляемой компанией Google. Описывается разработанная на основе этой технологии экспериментальная прикладная программа STENOGRAPH, а также приводятся предварительные результаты ее тестирования.

Введение

В связи с бурным ростом рынка интернет-поиска с использованием мобильных устройств разработкам в сфере упрощения ввода текстового запроса информации уделяется большое внимание. Наиболее перспективными в этой ситуации являются разработки, связанные с распознаванием речи, – современная альтернатива ручному вводу с клавиатуры. Однако в этом направлении существует масса различных проблем:

- большая вычислительная сложность и огромный объем баз данных, необходимых для корректного решения задачи распознавания речи;
- большое разнообразие используемых мобильных устройств и интернет-платформ.

Одним из возможных путей преодоления указанных трудностей является перенос на удаленный сервер системы распознавания речи. В связи с этим наиболее перспективным выглядит применение активно развивающихся в настоящее время облачных интернет-технологий, позволяющих легко наращивать по мере необходимости производительность и объем используемых баз данных.

1. Современные облачные интернет-технологии

Облачные технологии – это парадигма, предполагающая распределенную и удаленную обработку и хранение данных. Применение облачных технологий позволяет организовать повсеместный и удобный удаленный доступ к общему пулу вычислительных ресурсов, таких как устройства хранения данных, сети передачи данных, а также различные сервисы и приложения [1]. Одним из преимуществ данного подхода является снижение требований к вычислительным мощностям на стороне клиента при решении сложных вычислительных задач. Это позволяет интегрировать различные ресурсоемкие сервисы, например распознавание устной речи, в устройства с ограниченными ресурсами (мобильные телефоны, планшеты и др.). Облачные технологии имеют ряд положительных особенностей [2]:

1. Самообслуживание по требованию (on-demand self-service) – обеспечивает пользователю возможность самому определять вычислительные мощности конкретного облака.
2. Универсальный сетевой доступ (broad network access) – помогает организовывать однообразный доступ к облаку вне зависимости от используемого терминала.
3. Объединение ресурсов в пулы (resource pooling) – позволяет динамически изменять вычислительные мощности облака в зависимости от нагрузки.
4. Эластичность (rapid elasticity) позволяет предоставлять, расширять, сужать и убирать услуги в любой момент времени без дополнительных издержек взаимодействия с поставщиком.
5. Непрерывный учет потребления (measured service) – дает возможность исчислять объем услуг, предоставленный пользователю, в фактически потребленных ресурсах (объеме памяти, пропускной способности, количестве транзакций и пр.).

Перечисленные особенности делают облачные технологии особенно удобными для использования в развивающихся проектах, когда требования к вычислительным мощностям и базам данных постоянно изменяются.

Облачные сервисы могут предоставляться пользователю по трем основным моделям [2]: программное обеспечение как услуга (software as a service) – готовая платформа с уже установленным программным обеспечением. В этом случае пользователю нет необходимости покупать программное обеспечение самому, а достаточно лишь использовать то, которое предоставляет облачный сервис;

платформа как услуга (platform as a service) – облачная платформа, а также возможность разворачивать на ней любые приложения;

инфраструктура как услуга (infrastructure as a service) – некоторая облачная инфраструктура (виртуальные серверы), а также возможность конфигурировать ее под нужды пользователя.

Данные модели сервисов позволяют использовать облачные технологии в широком спектре проектов, начиная от небольших интернет-порталов до корпоративных сетей с множеством сервисов и клиентов с широкой географией.

Один из наиболее развитых сервисов технологии облачных вычислений предоставляется компанией Google [3]. Компания Google развивала технологию облачных вычислений более 10 лет. Многопользовательская инфраструктура Google позволяет сократить время развертывания приложений и их обновлений для всех пользователей. Быстрое внедрение программных решений с использованием облачных технологий Google обладает преимуществом по сравнению с другими системами. Веб-приложения на базе облачных технологий Google предоставляют пользователям доступ к облачным приложениям и хранимой информации с различных устройств, что повышает удобство и мобильность пользователя. Хранение данных в облаке, а не на компьютерах пользователей позволяет нескольким пользователям обращаться к информации и работать с ней одновременно, не беспокоясь о совместимости операционных систем, программного обеспечения и браузеров. Синхронная репликация позволяет синхронизировать данные и действия пользователя в режиме реального времени между несколькими центрами обработки данных. Если один из них по какой-либо причине становится недоступен, то система мгновенно обращается к резервному центру, при этом пользователь не испытывает никаких перебоев в обслуживании.

2. Облачная технология автоматического распознавания речи

С развитием интернет-технологий, а также коммуникационных возможностей мобильных устройств все более актуальной становится разработка разнообразных интернет-сервисов. Одним из многообещающих сервисов являются системы обработки и распознавания человеческой речи. Современные системы автоматического распознавания речи (АРР) представлены двумя основными технологиями, реализующими распознавание голосовых команд (Voice Command Recognition, VCR) и распознавание слитной речи (Speech-to-Text, STT).

Исторически VCR-технология возникла раньше STT-технологии, и уже в 1990-х гг. VCR была успешно использована в ряде коммерческих приложений (см., например, [4, 5]). В основе VCR-технологии положен принцип сопоставления распознаваемой речевой команды с набором эталонов методами динамического программирования. При объеме словаря до 1 000 слов VCR-технология позволяет достичь высокой достоверности распознавания и к настоящему времени достаточно широко используется в мобильных устройствах при голосовом наборе и навигации по сайтам.

STT-технология применяет теорию скрытых Марковских моделей, с помощью которых реализуется метод статистического сравнения распознаваемой фразы с эталонами. При этом используются акустические модели слов, а также грамматические модели фраз и предложений. Задача высококачественного распознавания слитной речи в рамках STT-технологии еще не решена нигде в мире, хотя достоверность распознавания уже достаточно высока для ее использования в ряде практических приложений. Дальнейшее возрастание точности распознавания зависит от качества используемых акустических и языковых моделей. Для обучения этих моделей необходим большой объем исходного речевого материала. Так, для создания акустических

моделей требуются сотни часов записей речи для нескольких тысяч дикторов. При этом для повышения устойчивости распознавания к помехам и искажениям необходимо использовать записи, созданные в различных каналах и условиях. Еще более жесткие требования предъявляются к разработке языковых моделей распознавания. Для обучения языковых моделей необходимо использовать корпуса текстов объемом от сотен миллионов словоформ до нескольких миллиардов. Подготовка и обработка такого объема обучающего материала – это сложная, кропотливая и весьма дорогостоящая работа [6].

Архитектура системы APP определяется местом обработки речевого сигнала: на клиентском компьютере или на сервере. Первая – наиболее привычная и распространенная архитектура – называется встроенной. При такой реализации вся обработка речевого сигнала и распознавание выполняются на компьютере клиента или мобильном устройстве. Данное решение обладает целым рядом недостатков. Прежде всего, это проблемы, связанные с большим разнообразием существующих архитектур компьютеров и мобильных устройств. Для мобильных устройств это еще и дополнительные ограничения, накладываемые на вычислительные ресурсы и память.

Второй вариант построения систем APP – более мощная и более гибкая альтернатива встроенной системе. Эти системы построены на основе клиент-серверной архитектуры с использованием технологий облачных вычислений. Клиентский компьютер (мобильный телефон, смартфон, планшет или нетбук) в такой архитектуре осуществляет только ввод и передачу речевого сигнала по цифровому каналу связи на удаленный сервер, а сервер выполняет основную работу – распознавание полученной последовательности данных. Кроме того, у этой архитектуры отсутствуют ограничения на вычислительные ресурсы клиентского компьютера, что дает возможность использовать современные и более сложные алгоритмы распознавания, а также централизованно поддерживать и обновлять серверную программу системы APP [7].

На сегодняшний день компания Google является лидером по предоставлению облачных технологий распознавания речи [8]. В течение последних пяти лет активно развивалась облачная технология распознавания речи Google Voice, и к настоящему времени существуют технологии распознавания речи для большинства европейских языков, включая русский, а также японский и китайский. Одним из немаловажных компонентов системы распознавания речи Google Voice является обучающая выборка звукозаписей человеческого голоса. Для системы Google Voice источником таких записей являются различные сервисы, предоставляемые Google и использующие речевые технологии. К ним относятся система распознавания речи и команд в системе Android, сервис диктовки писем Google Mail, телефонная справочная система Goog411 и др. [9]. Таким образом, обучающая выборка постоянно пополняется новыми образцами голосов как с особенностями произношения, так и эффектами, от которых зависят технические характеристики записи и передачи голоса на различных устройствах. К примеру, в 2011 г. обучающая выборка для английского языка составляла примерно 230 млрд записей слов, извлеченных из реальных запросов [10]. Для обработки таких объемов информации требуется около 70 лет процессорного времени, однако при использовании облачной технологии Google время сокращается до одного дня [9].

Рассматриваемая реализация распознавания речи включает в себя три модели: акустическую, лексическую и языковую [8]. Акустическая модель ответственна за распознавание фонем, она учитывает все возможные варианты произношения, а также другие особенности, такие как тип используемого микрофона (качество записи), фоновые шумы, возраст и пол говорящего и многое другое. Наиболее важным в данном случае является объем обучающей выборки: чем он больше, тем лучше будет результат распознавания. В лексической модели фонемы объединяются в слова на основе словарей, в которых указаны различные варианты произношения слов. Языковая модель объединяет слова, используя статистический подход. На основе анализа поисковых запросов, текстов Интернета выделяются вероятности взаимного расположения слов в предложении.

3. Экспериментальная прикладная программа STENOGRAPH

Для оценки современного состояния разработки облачных интернет-технологий распознавания речи, а также возможности построения на их основе готовых речевых приложений был проведен эксперимент по использованию русскоязычной версии системы распознавания Google Voice для решения задачи стенографирования устной речи. Подобный программный продукт мог бы найти достаточно широкое применение во многих сферах деятельности (бизнес, медицина, веб-приложения) в качестве средства, заменяющего (дополняющего) клавиатурный ввод.

В качестве входных требований к проекту были взяты следующие положения:

- программное средство должно реализовываться в виде приложения под ОС Windows;
- работа с программным средством должна вестись не только в ручном стартстопном режиме (как это предусмотрено системой Google Voice), но и в автоматическом режиме путем дополнительного включения в его состав детектора речи, определяющего начало и конец голосового сигнала;
- программное средство должно обладать простым и понятным графическим интерфейсом;
- к программному средству должны предъявляться минимальные требования по установке и настройке.

В связи с этим в качестве средств реализации были выбраны язык программирования C++, который обеспечивает максимальную производительность при захвате и обработке речевого сигнала, и библиотека построения пользовательских интерфейсов Qt Quick, которая позволяет в кратчайшие сроки разрабатывать кроссплатформенные пользовательские интерфейсы.

В экспериментальной системе автоматического стенографирования выделены следующие программные компоненты:

- детектор речи;
- компонент сжатия и обработки речевого сигнала;
- компонент для работы с распознаванием речи Google;
- графический пользовательский интерфейс.

Детектор речи является необходимым дополнительным компонентом, обусловленным используемой архитектурой удаленного распознавания речи. В этом случае канал передачи данных (в нашем случае интернет-соединение) является «бутылочным горлышком», ограничивающим максимально возможный объем передачи данных в заданный промежуток времени. При передаче по каналу слишком длинного отрезка речевого сигнала происходит недопустимо длительная задержка ответной реакции системы распознавания. Кроме того, при этом возникает значительный риск получения ошибочных результатов распознавания. Ввод коротких отрезков речи возможен при использовании ручного стартстопного режима, предоставляемого системой Google Voice. Однако для решения поставленной здесь задачи стенографирования устной речи этот режим оказывается крайне неудобным. Обычно диктовка осуществляется короткими фразами, заканчивающимися паузами. После произнесения каждой из них пользователь, как правило, желает убедиться в правильности распознавания и при необходимости повторить ее более четко.

Задача детектора речи заключается в автоматической локализации полезного сигнала, т. е. в определении начала и конца произнесенной фразы. Это обеспечивает автоматический пофразовый ввод речи. В экспериментальной программной системе использован относительно простой, но достаточно эффективный алгоритм определения полезного потока речи, основанный на расчете усредненной текущей энергии звуковой волны входного сигнала. Перед началом работы в течение одной секунды производится замер фонового шума для определения порога срабатывания детектора речи. Порог вычисляется как произведение средней энергии фонового шума на коэффициент, заданный в настройках программы (по умолчанию равный 2,0).

Далее ведется запись в кольцевой буфер задержки сигнала (по умолчанию задержка равна 1,0 с) и одновременно вычисляется текущая средняя энергия звуковой волны входного сигнала (по умолчанию время усреднения равно 0,4 с). Если порог будет превышен, задержанный входной сигнал передается на вход компонента сжатия и обработки сигнала, а передача сигнала будет вестись до тех пор, пока значение средней энергии не станет меньше значения порога.

Принцип работы детектора речи поясняется на примере произнесенной фразы «Раз два три» (рис. 1).

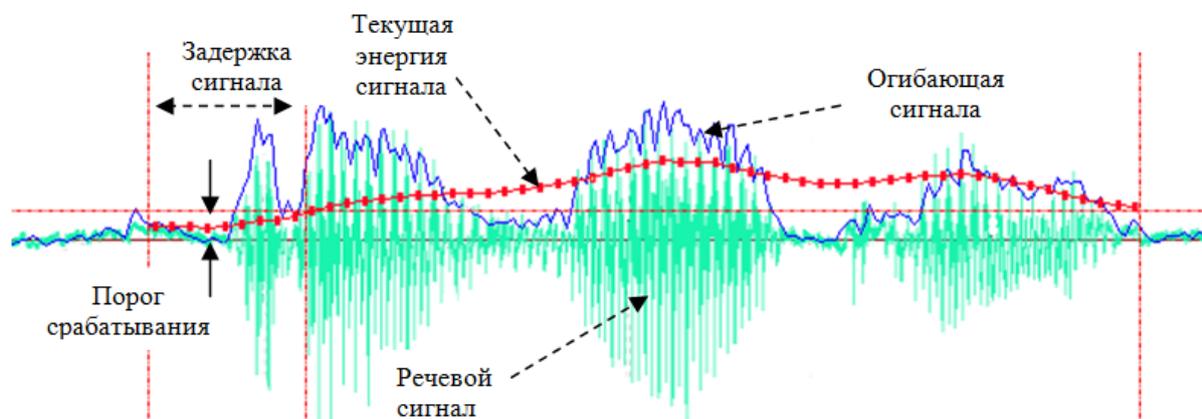


Рис. 1. Пояснения к принципу работы детектора речи

Полученный полезный сигнал в модуле сжатия и обработки сигнала подготавливается для отправки на удаленный сервер распознавания речи Google Voice. Для этого сигнал кодируется в открытом формате FLAC (Free Lossless Audio Codec) [12] с заданными характеристиками (частота дискретизации 16 кГц, моно). При этом происходит сжатие сигнала кодеком, что уменьшает объем передаваемых данных и как следствие сокращает время ожидания результата распознавания. Использование open-source-кодека имеет и другие преимущества: простоту использования сторонними разработчиками, единообразие получаемого сервером формата данных. Каждый компонент программы работает асинхронно и имеет свои пулы данных для нивелирования эффектов, связанных с задержками в работе каждого компонента. Такие задержки появляются как на этапе записи и кодирования речевого сигнала в файл, так и на этапе отправки его на удаленный сервер.

Полученная запись речевого сигнала прикрепляется к get-запросу на распознавание, отправляемому на сервер Google. Get-запрос имеет следующую форму: <http://www.google.com/speech-api/v1/recognize?client=chromium&lang=ru-Ru&maxresults=3>. В поле lang get-запроса указывается используемый язык (в нашем случае русский), а в поле maxresults – количество возвращаемых вариантов распознавания. Ответ на такой запрос от сервера приходит в формате JSON (JavaScript Object Notation) [13]. Ниже приведен пример ответа на запрос распознавания:

```
{
  «status»:0,
  «id»:«5e34348f2887c7a3cc27dc3695ab4575-1»,
  «hypotheses»: [
    {«utterance»:«один два три»,«confidence»:0.7581704},
    {«utterance»:«1 2 3»},
    {«utterance»:«один 2 три»}
  ]
}
```

В этом примере если поле status равно 0, запись успешно распознана, если 5 – запись не распознана. Поле id – это уникальный идентификатор запроса; поле hypotheses – результат распознавания, в нем есть два подполя: utterance – распознанная фраза и confidence – достоверность распознавания. Полученные данные можно анализировать и выдавать пользователю в наиболее подходящем виде. Например, числовые комбинации – в виде чисел или слов.

На рис. 2 описанный алгоритм более подробно представлен в графическом виде.

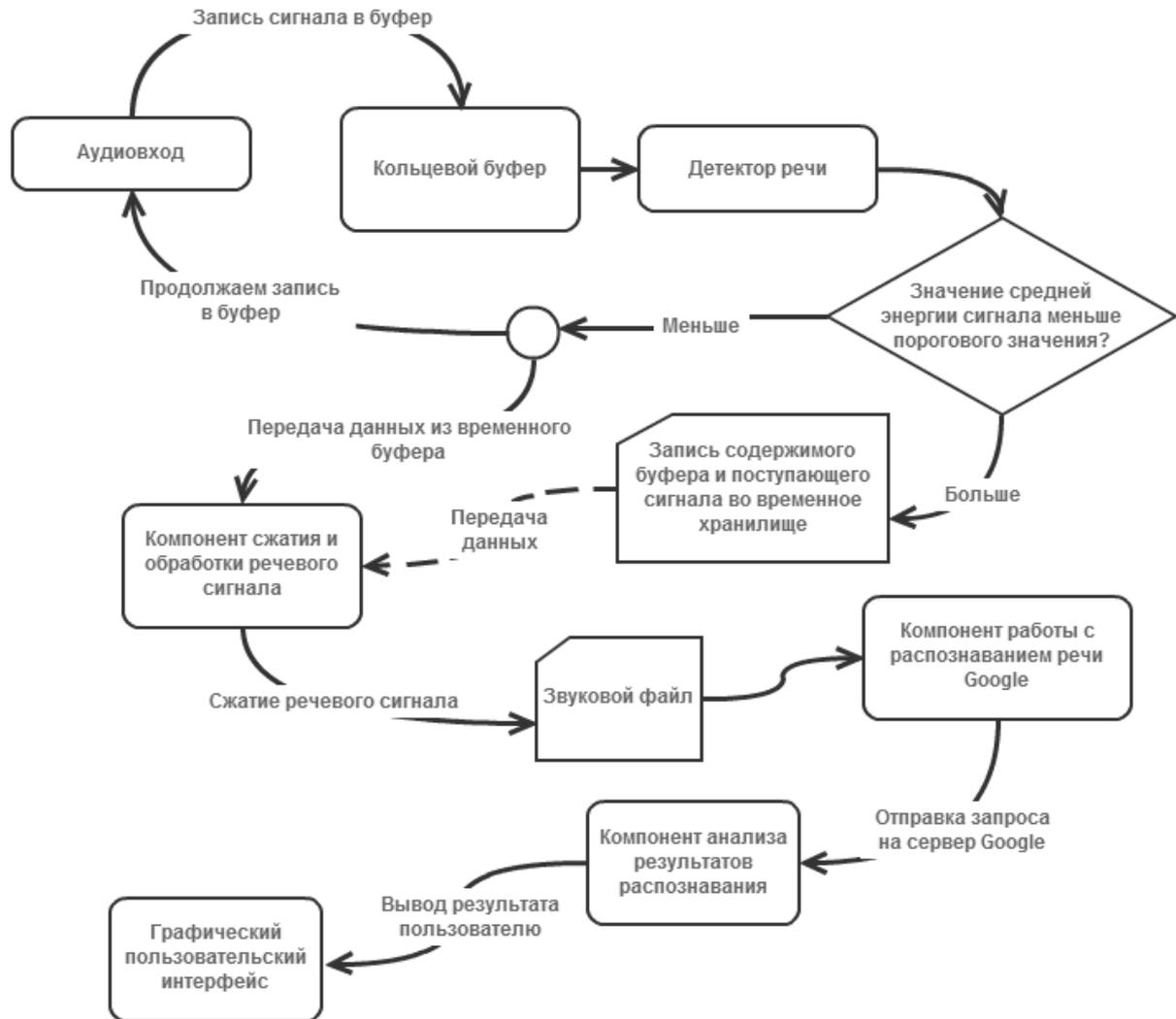


Рис. 2. Схема алгоритма работы прикладной программы STENOGRAPH

4. Пользовательский интерфейс и предварительные результаты тестирования

Рассмотрим главное окно программы STENOGRAPH (рис. 3).

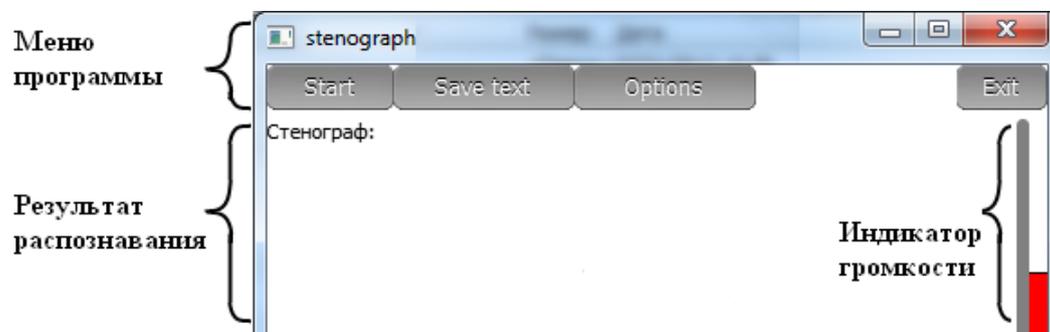


Рис. 3. Структура рабочей области программы STENOGRAPH

Меню программы включает шесть функциональных кнопок:
Start|Stop – запуск и остановка записи;

- Play – проиграть последнюю запись (только в ручном режиме);
 - Recognize – распознать последнюю запись (только в ручном режиме);
 - Save text – сохранить набранный текст в файл;
 - Options – настройки программы;
 - Exit – выход.
- В режиме Options открывается окно настроек программы (рис. 4).

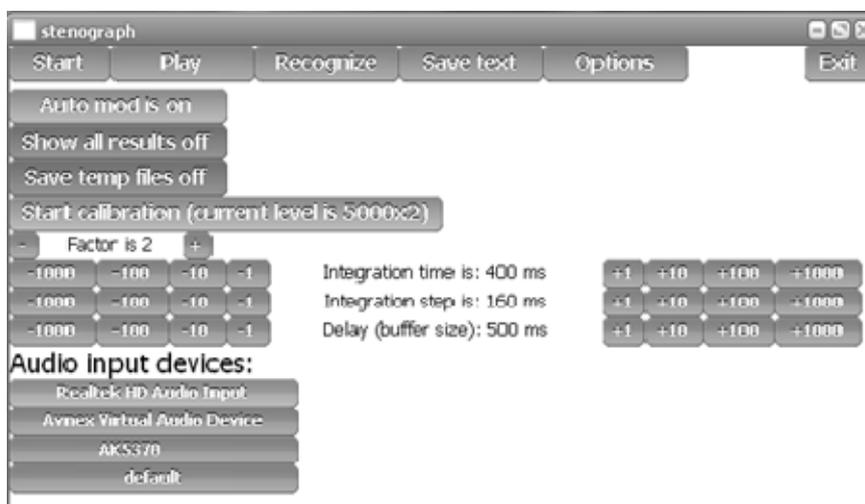


Рис. 4. Настройки программы STENOGRAPH

Окно настроек содержит:

- переключатель Auto mod on/off – выбрать режим работы: автоматический либо ручной;
- переключатель Show all result on/off – вывести все или только наилучшие варианты распознавания;
- переключатель Save temp files on/off – сохранить или удалить произносимые звуковые файлы;
- кнопку Start calibration – запустить режим калибровки уровня входного шума;
- кнопку Audio input devices – выбрать тип устройства ввода речевого сигнала.

Кроме перечисленных основных кнопок окно содержит и ряд дополнительных, используемых при экспериментальной отладке программы в случае смены окружающих акустических условий (например, микрофона, устройства ввода речи) и в других случаях. К таким дополнительным кнопкам относятся:

Factor is (+/-) – для подстройки порога срабатывания детектора речи путем увеличения или уменьшения коэффициента, заданного по умолчанию 2,0;

Integration time is (+/-) – для подстройки постоянной времени интегратора детектора речи путем ее увеличения или уменьшения, заданной по умолчанию 400 мс;

Integration step is (+/-) – для подстройки временного интервала считывания сигнала путем его увеличения или уменьшения, заданного по умолчанию 160 мс;

Delay (buffer size) (+/-) – для подстройки временной задержки речевого сигнала путем ее увеличения или уменьшения, заданной по умолчанию 500 мс.

Последовательность действий при работе с программой:

1. Запустить программу stenograph.exe.
2. Выбрать из меню пункт Options.
3. Нажать кнопку Start calibration. В данном режиме будет произведена калибровка уровня фонового шума. Для этого в течение одной секунды требуется сохранять тишину.
4. Для закрытия окна настроек повторно нажать в меню пункт Options.
5. Для начала записи выбрать из меню пункт Start.
6. Теперь можно начать надиктовку фраз. Для более корректной работы фразы должны быть не длиннее 5 с (рекомендуется, но не обязательно).

Если выбран режим Show all result off, в окне главной программы (рис. 5, а) появляется с новой строки наилучший вариант распознавания каждой фразы. Если же выбран режим Show all result on, в каждой строке выводятся также и другие, менее вероятные результаты (рис. 5, б).

7. По окончании работы необходимо выбрать из меню пункт Stop.

8. Для сохранения введенного текста нужно выбрать из меню пункт Save file. Текст сохранится в файл txt.



Рис. 5. Примеры отображения результатов распознавания:

- а) режим Show all result off – дважды повторенные четыре фразы из текста песни «Подмосковные вечера»;
 б) режим Show all result on – десятикратное повторение слова «шестьсот»

Для предварительной оценки потребительских качеств программы было проведено предварительное испытание надежности распознавания для двух наиболее практически важных категорий слов: названий цифр от 0 до 9, речевых ответов «Да» и «Нет».

Испытания проводились в реальных акустических условиях лаборатории с использованием внутреннего микрофона ноутбука. В испытаниях участвовали три диктора: двое мужчин и одна женщина. Средняя надежность распознавания названий цифр и речевых ответов составила около 90 %. При этом основной массив ошибок распознавания чисел приходился не на цифры 3 и 7, а на речевой ответ «Да». Однако если в соответствии с решаемой прикладной задачей на определенных этапах диалога речь идет только о распознавании названий цифр или речевых ответов, то надежность их распознавания можно существенно повысить, если использовать режим Show all result on. Этот факт хорошо виден на рис. 6, где приведены варианты принятия решения для цифр и речевых ответов при их десятикратном повторении.

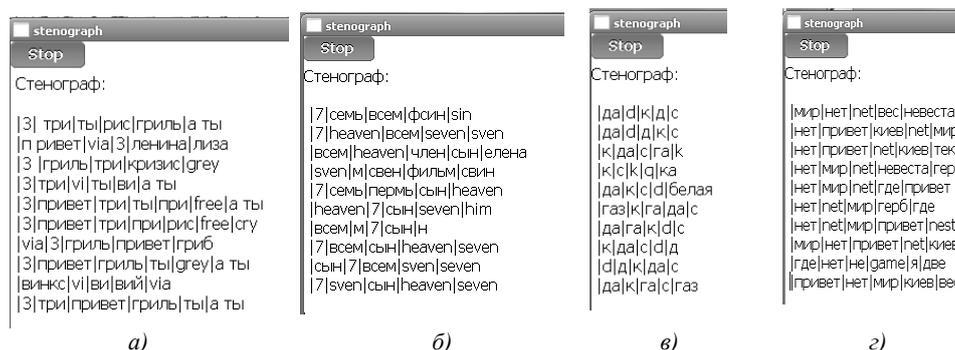


Рис. 6. Варианты принятия решения для цифр и речевых ответов

Важной особенностью рассматриваемой системы является то, что надежность распознавания практически одинакова для отдельного и слитного произнесения слов. Система характеризуется также весьма обширным допустимым словарем распознавания. Однако ощутимым недостатком системы является то, что зачастую незнакомое ей слово она не отвергает, а заменяет другим словом, близким по звучанию. Так, при распознавании двух строф стихотворения Пушкина (рис. 7, а) первая строфа распознана безошибочно, а вторая лишь отдаленно напоми-

нает истинный ее текст, а при распознавании слитно произнесенных названий мировых столиц (рис. 7, б) произошли следующие замены: Андорра – помидоры, Осло – Ош, Рейкьявик – видеоролик, Скопье – скобки, Дели – диалект, Астана – пастернак.

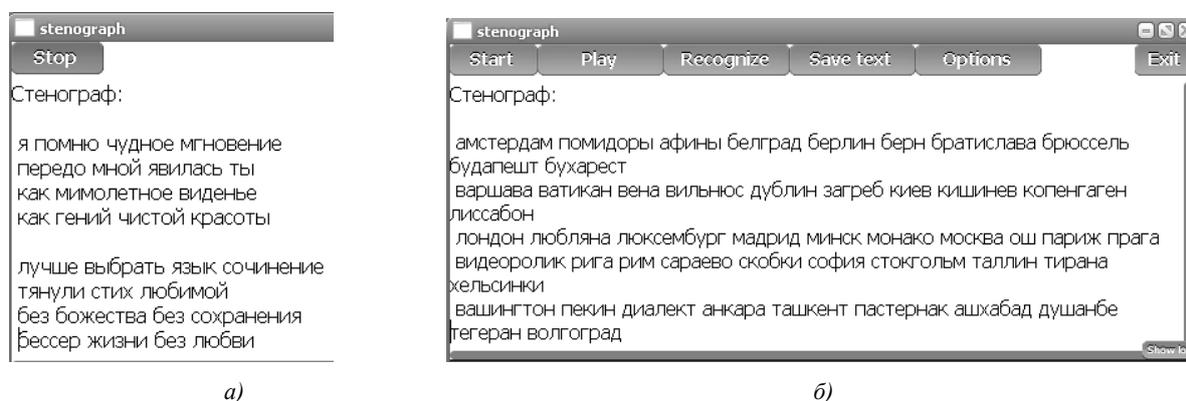


Рис. 7. Результат распознавания: а) двух строк стихотворения Пушкина;
 б) слитно произнесенных названий мировых столиц

Заключение

По результатам работы прикладной программы STENOGRAPH и проведенных экспериментов по распознаванию речи можно сделать следующие выводы:

- программа STENOGRAPH обеспечивает достаточно надежное автоматическое интерактивное взаимодействие пользователя с облачной системой распознавания речи Google Voice;
- использование общепринятого и открытого формата сжатия речевого сигнала (flac) не создает большой нагрузки на канал передачи данных, что позволяет создавать системы, активно реагирующие на голосовые команды пользователя;
- достаточно высокая для практики надежность распознавания может быть достигнута в случае, если в прикладной диалоговой системе используется режим Show all result on;
- отсутствие механизмов обратной связи не позволяет создавать системы с настройкой под определенного пользователя, что могло бы значительно улучшить надежность распознавания для дикторов с особенностями произношения;
- отсутствие возможности влиять на используемые акустические модели слов, а также грамматические модели фраз и предложений не позволяет производить настройку приложения на определенную предметную область, что могло бы существенно улучшить надежность распознавания.

Облачная система распознавания русской речи Google Voice обладает положительными свойствами, такими как способность к распознаванию дискретной и слитной речи, потенциальная устойчивость к смене диктора. Система использует большой по объему словарь русских слов и словоформ, однако их количество не может обеспечить безошибочное стенографирование текстов произвольной тематики.

Список литературы

1. Public Cloud Service Definition. Public Version 1.5 // VMware, Inc. [Electronic resource]. – 2010. – Mode of access : <http://www.vmware.com/files/pdf/VMware-Public-Cloud-Service-Definition.pdf>. – Date of access : 01.08.2012.
2. Mell, P. The NIST Definition of Cloud Computing. Recommendations of the National Institute of Standards and Technology / P. Mell, T. Grance // U.S. Department of Commerce [Electronic resource]. – NIST Special Publication, 2011. – Mode of access : <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>. – Date of access : 01.08.2012.

3. Schalk, C. Google Cloud Technologies Overview / C. Schalk // Cloud Computing Expo, 2010 [Electronic resource]. – Mode of access : <http://www.slideshare.net/cschalk/google-cloud-technologies-overview>. – Date of access : 01.08.2012.
4. Lobanov, B.M. Continuous Speech Recognizer for Aircraft Application / B.M. Lobanov, T.V. Levkovskaya // Proc. of the 2nd Intern. Workshop «Speech and Computer» – SPECOM'97. – Cluj-Napoca, 1997. – P. 97–102.
5. An Intelligent Answering System Using Speech Recognition / B.M. Lobanov [et al.] // Proc. of the 5th European Conf. on Speech Communication and Technology – EUROSPEECH'97. – Rhodes-Greece, 1997. – Vol. 4. – P. 1803–1806.
6. Распознавание речи // Центр речевых технологий [Электронный ресурс]. – 2001–2012. – Режим доступа : <http://www.speechpro.ru/technologies/recognition>. – Дата доступа : 01.08.2012.
7. Маковкин, К.А. Удаленная система автоматического распознавания речи / К.А. Маковкин // Речевые технологии. – № 4. – 2009. – С. 70–96.
8. Manjoo, F. Now You're Talking / Farhad Manjoo // The Slate Group. – Washington Post Company [Electronic resource]. – 2012. – Mode of access : http://www.slate.com/articles/technology/technology/2011/04/now_youre_talking.single.html. – Date of access : 01.08.2012.
9. Singhal, A. Knocking down barriers to knowledge / Amit Singhal // Google Official Blog [Electronic resource]. – 2011. – Mode of access : <http://googleblog.blogspot.com/2011/06/knocking-down-barriers-to-knowledge.html>. – Date of access : 01.08.2012.
10. Enge, E. Search Algorithms with Google Director of Research Peter Norvig / E. Enge // Stone Temple Consulting [Electronic resource]. – 2011. – Mode of access : <http://www.stonetemple.com/search-algorithms-with-google-director-of-research-peter-norvig>. – Date of access : 01.08.2012.

Поступила 16.08.12

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

B.M. Lobanov, V.A. Zhitko

APPLICATION OF CLOUDY TECHNOLOGIES IN SPEECH RECOGNITION

Features of cloud Internet technologies in automatic speech recognition are considered. Positive peculiarities and basic models of modern cloud Internet technologies are briefly described. The existing methods, technology and architecture of automatic speech recognition, and also features of cloud technology for speech recognition provided by Google are considered. STENOGRAPH, experimental applied software based on this technology, its user interface and preliminary results of the testing are described.