Интеграция Asterisk с системой синтеза/распознавания речи ЦРТ

Всем известно, что Asterisk является очень гибкой программной телефонной платформой, на основе которой можно достаточно просто строить приложения, гибко адаптируемые под конкретные требования бизнеса. Например, простейшее IVR меню может быть построено за 20 минут, при этом наибольшее время потребуется для записи звуковых файлов для меню, подбор наилучшего варианта записи, подгонка фрагментов записи друг к другу (по интонации, громкости и пр.)

Представьте, что теперь не нужно записывать звуковые файлы, для использования в IVR системе. Диктор не может заболеть, потерять голос или уволиться, в конце концов. Подготовка и обновление фраз для IVR сводится к замене или изменению текста. Время, которое требуется для этой процедуры несоизмеримо меньше по сравнению с записью речевых сообщений диктором.

Этот факт позволит аутсорсинговым call-центрам вводить новые услуги (ветки меню для новых проектов) в кратчайшие сроки, что несомненно повышает эффективность call-центра в разы.

Распознавание речи так же способствует построению эффективных голосовых интерфейсов. Можно привести несколько ярких примеров меню, в которых использование тонального набора может запутать клиента и никогда не привести к требующейся ему информации:

1. Большое количество вариантов произнесения объекта (например географических ориентиров)

2. Построение меню с множественным выбором (нажмите 1 чтобы то, нажмите 2 чтобы другое и т.д.)

Именно здесь система распознавания русской речи поймет что сказал клиент, не заставляя его дослушивать до конца или требуя ввода одного единственного верного значения.

Ещё в начале работы с Asterisk я столкнулся с построением IVR, которое суммарно должно было отвечать за ввод более дюжины однотипных параметров. При этом я столкнулся с определенными сложностями:

1. Предлагать вводить посредством DTMF двузначные числа. При этом человек не может держать в памяти множество значений и скорее всего откажется от использования такого сложным меню

2. Предоставить возможность постраничного листинга динамически изменяемой информации

3. Предоставить в меню возможность выбора основных вариантов, для выбора специальных вариантов - организовать разговор с оператором

Я использовал последний вариант. Однако уже сейчас имеется возможность использовать эффективные средства синтеза/распознавания русской речи. Сразу стоит сказать, что это платные решения. OpenSource приложения имеются, но применительно к синтезу русской речи годятся к использованию скорее в тестовых лабораторных условиях. Существует возможность настройки распознавания русской речи в CMU Sphinx4, но распознавание работает не достаточно надежно в данный момент. В качестве примера можно привести следующие решения: синтез русской речи - Govorilka, Festival, eSpeak и др., распознавание русской речи на ограниченном словаре - Горыныч.

Технические особенности

Для синтеза речи используется технология синтеза речи VitalVoice от ЦРТ, которую по заверениям сотрудников компании ЦРТ, смело можно считать лучшей технологией синтеза русской речи на данный момент. Решение от ЦРТ включает в себя пять голосов, построенных на основе записи десятков часов речи профессиональных дикторов. Возможен синтез сложных текстов с корректной расстановкой ударений, правильной расшифровкой сокращений и интонационной окраской.

Ресурс распознавания речи ASR принимает грамматику в SRGS формате и возвращает результат распознавания в NLSML разметке. В настоящий момент поддержаны следующие теги, используемые в SRGS грамматиках и позволяющие строить более сложные меню:

1. confidence level - уровень уверенности распознавания (коэффициент достоверности)

2. n-best - возврат нескольких результатов распознавания с указанием confidence level для каждого

3. instance - семантическая интерпретация произносимых клиентом равнозначных слов и выражений

Также поддерживается режим barge-in, т.е. пользователь может произносить запросы не дожидаясь

окончания проигрывания синтезированной фразы в IVR.

SRGS имеет 2 эквивалентных формата - это означает что одна и та же грамматика может быть представлена одной из следующей форм:

XML – описание грамматики структурой XML файла;

ABNF – описание грамматики структурированным текстовым файлом.

Решение от «Центра Речевых Технологий» в данный момент поддерживает связь по протоколу MRCPv1 (RFC4463), поддерживаются uLaw и aLaw кодеки. Завершено тестирование ресурса синтеза речи TTS (plan-text и текст в формате SSML), тестирование распознавания речи продолжается.

Со стороны Asterisk поддержка протокола MRCP реализована посредством библиотеки UniMRCP. На основе этой библиотеки поддержка реализована и для FreeSWITCH, которая в свою очередь базируется на коде приложения unimrcp-client, поставляемого в составе исходных кодов библиотеки.

Краткая справка

Протокол MRCPv1

Протокол MRCPv1 (Media Resource Control Protocol) описан в RFC4463, представляет собой простой текстовый протокол, похожий на HTTP. Передаваемые MRCP запросы и ответы инкапсулируются в транспортный протокол RTSP. Передача голосовых данных производится посредством протокола RTP. Спецификация определяет ряд методов и параметров, которые участвуют в установлении и управлении сессиями синтеза/распознавания.

Библиотека UniMRCP

Библиотека поддерживает версии протокола MRCPv1 и MRCPv2. Поддерживается работа с ресурсами распознавания, синтеза, верификации и записи речи. Для реализации протокола MRCPv2 используется библиотека SofiaSIP, библиотека кросс-платформенная и выпущена под лицензией Apache License 2.0, что позволяет использовать её в коммерческих решениях. Библиотека позволяет реализовать как MRCP клиент, так и сервер, в поставке имеются примеры реализации, которые служат хорошим примером для разработки своего приложения.

Кроме того автор активно работает над новыми возможностями, буквально на днях появился релиз 0.9.0 (0.8.0 был выпущен 3 месяца назад), основными нововведениями которого явились:

- Поддержка 64-битных ОС
- Поддержка ресурса записи
- Поддержка для RFC4733/RFC2833 и Inband DTMF генерации и распознавания
- Реализация RTCP стека

Модуль app_mrcp

На данный момент модуль представляет из себя базовую реализацию поддержки MRCP протокола, однако уже готовую к эксплуатации. Модуль работоспособен с Астериском всех версий начиная с 1.4.

MRCPSynth — синтез речи

Приложение получает на вход текст, предназначенный для синтеза и ряд параметров, таких как:

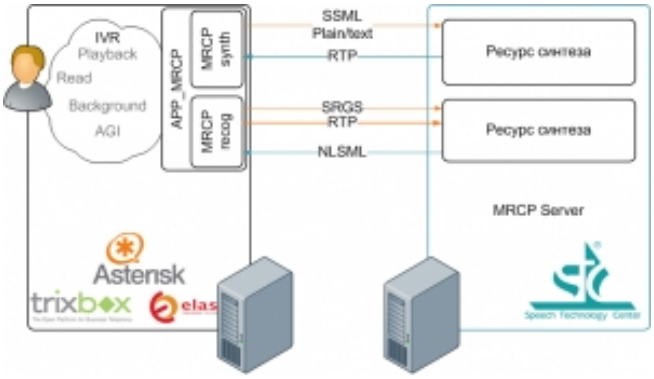
- Файл на жестком диске для записи синтезированной речи
- Параметры синтезированного голоса (возраст, пол, голос)
- Профиль сервера распознавания

Работает как приложение Playback, может производить запись полученного синтезированного голоса в файл, доступный для дальнейшего использования.

MRCPRecog — распознавание речи

Приложение получает на вход грамматику, после установления сессии с MRCP сервером передает все голосовые пакеты к ресурсу распознавания. Как только ресурс распознавания примет решение о том что распознавание закончено (по таймауту или успешному распознаванию в соответствии с

грамматикой) приложение прекращает выполнение и устанавливает переменную диалплана RECOG_RESULT, которая содержит результат в формате NLSML. Приложение аналогично приложению Background — воспроизведение звукового файла прерывается успешным вводом пользователя.



Пример использования

Ниже приведен теоретический пример использования в простейшем IVR с распознаванием ответа пользователя, синтезом приветствий и предварительно-синтезированными отдельными фрагментами меню. приложение nlsml-wrapper в данном примере должно выделять результат распознавания и управлять переходами внутри меню. Dialplan приведен для демонстрации использования приложений.

```
exten => ivr,1,Answer
exten => ivr,n,Wait(0.5)
exten => ivr,n,MRCPSynth(Добро пожаловать в компанию Рога и Копыта!,v(maria))
exten => ivr,n,MRCPRecog(ivr/grammar-sales.srgs,t(5000)f(synth/ivr-main-promt)b)
exten => ivr,n,AGI(agi://10.0.0.1/nlsml-wrapper,ivr,${RECOG_RESULT})
exten => ivr,n+5(tech),MRCPSynth(Техотдел,v(vladimir))
exten => ivr,n,GoToIf($["${ALARM}" = "1"]?alarm:norm)
exten => ivr,n(alarm),Playback(synth/tech-alarm-notification)
exten => ivr,n(norm),MRCPRecog(ivr/grammar-sales.srgs,t(5000)f(synth/ivr-tech-promt)b)
exten => ivr,n,AGI(agi://10.0.0.1/nlsml-wrapper,tech,${RECOG_RESULT})

exten => ivr-
dictate,n,MRCPSynth(${DICTATE_TEXT},v(${DICTATE_VOICE})w(synct/${DICTATE_NAME}.ulaw))
```

Планы

В планах на февраль 2010 — достаточно серьезная переработка UniMRCP клиента для Asterisk для возможности полной поддержки синтеза и распознавания русской речи от ЦРТ:

- Добавление команд CLI
- Добавление AMI событий и команд
- Поддержка SNMP
- Встроенное кеширование синтезированных фраз
- Парсинг NLSML в вид, пригодный к использованию в диалплане

Благодарности
Выражаю благодарность компании Центр Речевых Технологий и лично Кондратьеву Виталию за предоставленную возможность протестировать синтез и распознавание русской речи совместно с Asterisk.