

О технологиях взаимодействия «человек-компьютер» в системах компьютерной телефонии

Давид Петросян

Институт проблем информатики и автоматизации НАН РА

Аннотация

В работе приводится краткий обзор основных технологий взаимодействия пользователя с компьютером в системах компьютерной телефонии (СТП). Рассмотрены возможные области и границы применимости данных технологий в информационных, информационно-справочных и сетевых системах СТП.

Не останавливаясь на общепринятом механизме взаимодействия - ввода команд \ данных (клавиатура\ мышь) и отображения информации (монитор\ печать), рассмотрим другие возможные альтернативные или комбинированные механизмы взаимодействия пользователя с компьютером в системах компьютерной телефонии (СТП). (под термином «компьютер» имеется ввиду компьютер (сервер) СТП).

Очевидно, что наиболее естественным способом «общения» с компьютером для человека является голосовое общение, учитывая, к тому же, то обстоятельство, что в системах компьютерной телефонии в качестве терминала чаще всего используется телефон.

Рассмотрим возможные технологии голосового взаимодействия с системой СТП

- Технологии распознавания речи (Automatic Speech Recognition – ASR)
- Технологии синтеза речи (Text To Speech – TTS)
- «Конструирование» речи
- Комбинированные технологии

ASR используется для голосового ввода в компьютер по телефону команд и данных. Технологии синтеза речи находят применение в сетевых приложениях СТП, в частности, для передачи на телефон преобразованных в голосовую форму сообщений, полученных по e-mail (одна из технологий, применяемых в системах Unified Messaging [1]).

Проблемно-ориентированные технологии «конструирования» речи («сборка» из голосовой базы данных заготовок фрагментов и составление из них соответствующего предложения-справки в ответ на запрос пользователя) широко используются в информационных, информационно-справочных системах СТП.

Наконец, комбинированные технологии являются гибридными системами, отдельные элементы упомянутых выше технологий. Учитывая, что основным «терминалом» в системах СТП служит телефон, рассмотренные технологии являются базовыми технологиями систем компьютерной телефонии.

Приведем возможные варианты применения этих технологий.

Технологии распознавания речи (ASR)

Выделим два основных направления ASR

- Системы распознавания слитной речи для индивидуального пользователя
- Системы ограниченного распознавания речи.

Детальное рассмотрение и анализ известных систем ASR выходит за рамки тематики настоящей работы. Приведем обобщенные характеристики известных систем ASR и области СТИ, где они могут найти (или находят) применение.

Отметим также, что данное направление ИТ динамично развивается и совершенствуется, соответственно, приводимые оценки могут претендовать на достоверность только в настоящее время.

Рассмотрим обобщенные параметры нескольких распространенных (2002г) систем ASR (IBM ViaVoice Gold, Philips Free Speech 2000, Dragon NaturallySpeaking Preferred, Горыныч Проф. (более поздняя версия – Combat)

Все эти системы рассчитаны, в первую очередь, на обеспечение речевого ввода текста в документ. «Горыныч» позволяет вводить русскую речь, остальные английскую. Точность распознавания речи – около 90% для английских версий, Русская версия имеет невысокую точность распознавания (около 50-60 %). При установке программного пакета «мастер установки» просит наговорить ряд тестов, назначение которых адаптировать программу к особенностям вашего голоса, произношения, окружающего фона и т.п. Эта процедура может длиться около 60 минут. Отметим сразу, что данная технология непосредственного отношения к СТИ в настоящее время не имеет и приводится здесь только для полноты обзора.

Системы ограниченного распознавания речи

Распознавание речи в данных системах может осуществляться в двух режимах - с настройкой на голос конкретного пользователя и без настройки.

Первый режим в системах СТИ чаще всего используется в защищенных системах для «опознавания» звонящего по голосу. Данная технология удобна для организации персональных голосовых систем, в которых команды на основные действия могут отдаваться голосом.

Более широкое распространение получила технология распознавания речи без настройки на голос конкретного пользователя. Только подобная технология может обеспечить массовый голосовой доступ к различным информационным, информационно-справочным системам компьютерной телефонии.

Работа систем распознавания речи без настройки на голос пользователя строится на основе словарей распознавания. Словарь составляется на основе анализа речи нескольких сотен людей. В данное время существуют три типа подобных словарей: цифровой, алфавитно-цифровой и словарь ограниченного распознавания слитной речи (первые два словаря предполагают произнесение слов по одному), содержащий цифры от нуля до девяти и слова «да» и «нет».

Цифровые словари доступны для нескольких десятков языков, включая русский. Алфавитно-цифровые словари и словари для ограниченного распознавания слитной речи разработаны для существенно меньшего числа языков. Для русского языка таких словарей пока нет. К сожалению, подобные технологии для армянского языка, по имеющейся у нас информации, не разработаны.

Сегодня доступны два русских цифровых словаря. Первый содержит слова «ноль».... «девять» и слово «остановиться», второй – слова «да», «нет», «остановиться», «переносится», «аннулируется», «помогите». Такое разделение сделано для того, чтобы увереннее различать похожие по звучанию слова («да» и «два»). При этом, физически оба словаря размещаются в одном файле, но в каждый момент может работать только один. Предусмотрена возможность переключения словарей программно «на лету».

Упомянутые выше системы ограниченного распознавания речи ориентированы на работу с аппаратными средствами компьютерной телефонии – Scbus-совместимых DSP платформ (DSP-Digital Signal Processing), открытых для разработки речевых технологий, таких как ASR, TTS и др. К примеру, аппаратная система распознавания Antares 200/50 реализует распознавание команд, произнесенных на русском языке без настройки на голос говорящего одновременно по 32

телефонным каналам. Программное обеспечение разработано фирмой Voice Control Systems (VCS).

Доступна также программная система распознавания средствами центрального процессора, а не аппаратно. При этом используются существенные вычислительные ресурсы компьютера при значительно меньшем количестве одновременно обслуживаемых каналов (до 4-х). Это программа VRSoft той же фирмы, работающая только под управлением MS-DOS. Качество распознавания и словари абсолютно такие же, как при аппаратном распознавании. Система VRSoft работает совместно с голосовой платой фирмы Dialogic D/41 ESC [2].

Технологии синтеза речи

Наиболее привлекательным в использовании технологий синтеза речи в системах СТИ является предоставление возможности телефонному пользователю прослушать поступившую на его e-mail адрес электронную почту.

Данная технология имеет как самостоятельное значение, так и в качестве одного из базовых элементов Unified Messaging. В настоящее время доступно множество программных продуктов, осуществляющих эти функции для английского и большинства европейских языков.

Существенно скромнее представлен русскоязычный сектор TTS. Отметим еще раз отсутствие продуктов (по крайней мере, доступной информации) о армянской версии TTS.

Несмотря на оптимистические заявления российских разработчиков «e-mail на телефон», судя по отзывам в Интернете пользователей подобной системе в Москве, функционирование системы в настоящее время оптимизма не внушает.

Несовершенство русских версий TTS сдерживает развитие данных технологий в русскоязычном секторе Интернет. Предлагаемые зарубежные решения (в частности, Philips) не могут полностью удовлетворить имеющийся спрос на системы TTS, т.к. не полностью учитывают специфику русского языка, к тому же, наличие пусть даже совершенной технологий формирования речи требует наличия еще двух важных компонент – поддержку и пополнение голосовой базы фоном (аллофонов) и постоянное пополнение словаря системы.

Конечно, со временем данная проблема найдет свое решение, но в настоящее время, видимо, уместно говорить только об опытных системах сетевых СТИ, использующих русские версии TTS. Отметим один, общий для всех языковых версий TTS, недостаток. При воспроизведении текста системой TTS незнакомые слова и словосочетания (аббревиатуры, редкие собственные имена и фамилии и т.п.), отсутствующие в словаре чаще всего воспроизводятся по буквам, что не воспринимается, или плохо воспринимается на слух.

В данной части работы автор не имеет возможности привести полный список ссылок на использованные первоисточники. Только по тематике TTS Интернет содержит более 17 тысяч (!) наименований (поисковая система Google, тема – «синтез речи»).

Выборочный список источников информации приводится в [3-10].

«Конструирование» речи

Под данным определением подразумеваются элементы информационных, информационно-справочных систем СТИ, которые передают по телефонному запросу тематическую информацию, состоящую из законченных предложений или блоков. Фрагменты заранее «наговариваются» и хранятся в базе фрагментов.

Предложение формируется из постоянной, не меняющейся части и переменных составляющих, которые «собираются» из соответствующих фрагментов (например, «.....сегодня вторник (среда...суббота), восьмое (десятое.....тридцать первое) января (марта....). Местное время XX часов YY минут». Выделены переменные фрагменты. В большинстве случаев «переменные» - это числительные, дни недели, месяцы, ограниченные списки и т.п.

Понятно, что числительные в голосовой базе целесообразно хранить не в виде «от нуля до N» (если диапазон используемых цифр достаточно велик), а в виде голосовых базовых компонент, из которых «собирается» заданное число. Например, число 895 собирается из следующих компонент: «восемьсот» + «девяносто» + «пять». Это позволяет существенно сократить количество хранимых фрагментов. Данная, в общем, не сложная технология находит широкое применение в разнообразных СТИ системах.

Комбинированные технологии

Данные технологии взаимодействия телефонного пользователя с системой СТИ строятся по принципу разделения процедур ввода пользователем в систему данных и команд (которые производятся кнопочным набором на телефоне в тоновом режиме) от процедур голосового вывода информации и запросов системы.

Учитывая ограниченные возможности «кнопочного» ввода, подобные системы строятся, чаще всего, на основе диалога системы с пользователем, при котором голосовые запросы системы предполагают либо ввод цифровых данных (пароли доступа, идентификационные коды, номера телефонов и т.п.), либо предлагается в ответ на запрос нажать определенную кнопку (или одну из двух функционально альтернативных кнопок, или последовательную комбинацию кнопок) предложенную в запросе.

В многоплановых информационно-справочных системах распространены голосовые тематические меню, из которых пользователь может выбрать интересуемый раздел и, соответственно, прослушать содержание раздела.

Существенным в рассматриваемых системах является наличие грамотно построенного аппарата подсказок (подсказка ожидаемых действий пользователя, назначенные функциональные кнопки и др.) Подобные комбинированные технологии, несмотря на то, что несколько осложняют работу пользователя с системой, являются наиболее распространенными, ввиду их простоты и надежности в работе.

Литература

1. П.Иванов, Универсальные системы обработки сообщений
журн. «Сети» (Network World), №№ 4,8,9 2000г.
2. www.Comptek.ru
3. www.Tehnovoise.cse.lt/it/main.htm
4. www.Élan.fr
5. www.Forcecomputer.com
6. www.Ilsl.com
7. www.Locusdialog.com
8. www.Speech.be.philips.com
9. www.Speechworks.com
10. www.Vocalis.com

Կոմպյուտերային հեռախոսակապի համակարգերում «մարդ-կոմպյուտեր» փոխանցողությունների տեխնոլոգիաների մասին

Դ. Պետրոսյան

Ամփոփում

Աշխատանքում ներկայացվում է կոմպյուտերային հեռախոսակապի համակարգերում օգտագործողի և կոմպյուտերի փոխանցողությունների տեխնոլոգիաների կրճատ տեսությունը: