

УХО ДЛЯ РОБОТА

Принципиально новый подход к компьютерному распознаванию слов предложили российские ученые. С его помощью человек сможет командовать даже самым примитивным мобильником.

Доп. информация: кандидат физико-математических наук старший научный сотрудник Института радиотехники и электроники РАН Вячеслав Евгеньевич Анциперов, (095) 203-19-92, antciperov@cplir.ru

Принципиально новый подход к компьютерному распознаванию слов предложили российские ученые. С его помощью человек сможет командовать даже самым примитивным мобильником.

Разумное существо без труда узнает знакомое слово независимо от того, каким голосом и с какой интонацией оно произнесено. “Шесть” или “восемь” остаются для человека шестеркой или восьмеркой, как их ни произнеси - громко или шепотом, взволновано или спокойно, голосом старика или ребенка, мужчины или женщины. Мозг человека мгновенно отфильтрует смысловую часть от массы фоновых звуков.

Для машины каждый вариант голоса уникален. Поэтому программу для распознавания речи как правило приходится учить. В результате обучения в памяти кремниевых мозгов появляется огромная библиотека, хранящая тысячи возможных вариантов произнесения одних и тех же слов, например, названий цифр. Услышав слово, компьютер переберет библиотеку и почти наверняка в ней найдется нечто, похожее на услышанное.

Подход, который предложили ученые из Института радиотехники и электроники РАН, скорее человеческий, чем машинный: компьютер под их руководством фильтрует индивидуальные особенности, то есть выделяет только самую основную и отбрасывает все несущественное. В результате машина приобретает способность даже распознавать отдельные звуки и в уме составлять из них знакомые слова.

В результате памяти всего в 1 килобайт процессору хватит, чтобы уверенно узнавать все цифры и некоторые простые команды, правда, произнесенные пока только по-русски. Несколько десятков человек – мужчин и женщин, с дикцией безупречной и, наоборот, далекой от идеала –

пытались сбить с толку сообразительную программу, произнося цифры то шепотом, то дрожащим от волнения голосом. Но компьютер удачно отметал эмоциональные частоты как несущественные.

“Разработанный и созданный нами опытный программный интерфейс системы голосового ввода данных и команд управления предназначен для массовых мобильных электронных устройств, – рассказывает руководитель проекта кандидат физико-математических наук Вячеслав Анциперов. - Может быть, самое важное, принципиально новое в нашей работе – это то, что мы ориентируясь на понятие иерархической структуры речи выделили ее существенные элементы. Как в музыкальном произведении можно распознать есть более и менее высокие уровни организации – ритм, основная тема, аранжировка, так и мы научились в речевом потоке, то есть в широком спектре частот, выделять диапазоны, которые несут основную смысловую нагрузку. Оказалось, что это очень небольшая часть звуков человеческой речи – всего до 1 КГц. Остальное – из области психофизики. Так мы максимально упростили задачу для компьютера. И еще одно – мы научили его распознавать отдельные звуки, что подчас весьма нелегко. В результате наша система выигрывает как по быстродействию, так и по затратам процессорного времени и памяти по сравнению со всеми известными аналогичными системами. А это путь к эффективным речевым процессорам, пока еще никем не пройденный”.