Комп'ютерні засоби розпізнавання усної мови

Вадим Данник
Міжнародний науково-навчальний центр інформаційних технологій та систем
40, просп. Акад. Глушкова, Київ 252022, Україна
Електронна пошта: vadym@technopark.kiev.ua

Абстракт

Описуються впроваджені програмні та апаратні системи для розпізнавання усної мови, коротко описано методи, на яких вони базуються та сфери застосування. Наведено порівняльні характеристики, вказано основні тенденції розвитку.

Вступ

Використання технології розпізнавання мови почалось з невеликих програм, що працювали з десятком слів та могли керуватись голосовими командами. Розвиток цієї технології сьогодні дозволяє без довгої процедури настроювання на голос диктора, без використовування коштовних мікрофонів чи спеціалізованого обладнання усно керувати побутовими пристроями, вдосконалювати системи обмеження доступу, розширяти можливості людей з фізичними вадами тощо. Системи зі спеціалізованими словниками "розв'язують руки" медикам, журналістам, юристам, секретарям та рядовим користувачам сучасних персональних комп'ютерів. Активно розробляються та набувають поширення автоматичні довідникові системи з доступом з телефоного апарату.

Так, наприклад, фірма Sharper Image виробляє голосові номеронабирачі для телефонів, AutoLink OnBoard випускає автомобільні прилади з голосовим керуванням. Наявність голосового інтерфейсу надає "інтелектуальності" автомобільним бортовим системам, побутовій електроніці - диктофонам, телефонам, відеосистемам, дитячим іграшкам і навіть вимикачам.

Програми розпізнавання усної мови для ПК

Найпопулярнішим на сьогоднішній день програмним продуктом для персональних комп'ютерів залишається DragonDictate від Dragon Systems. Невибагливий до апаратних ресурсів, цей високоінтегрований програмний продукт з нескладним інтерфейсом підтримує 6 мов, включно з російською. Як і для інших подібних програм, необхідно застосовувати низькошумові гостроспрямовані мікрофони та стежити за чіткою вимовою слів. Оскільки для здійснення неперервного надиктовування необхідна не тільки значна адаптація до вимови та інтонації диктора, але й урахування контексту, додатково випускаються спеціалізовані модулі словників для різних застосувань - для офісу та банків, радіологічної та екстремальної медицина, телефонії. Для розробників напрацьовано багатий програмний та апаратний інструментарій для створення мовного інтерфейсу для прикладних систем (навігація по комп'ютерному інтерфейсу, ввід данних, телефонні транзакції тощо).

Апаратні рішення для розпізнавання мови

Voice Control Systems (VCS) розробила власний метод фонемного розпізнання, над яким працює вже понад 17 років. Він є достатньо надійним для керування автомобільним обладнанням чи роботи з телефоном "без рук". Базові DSP – TI-C3X та C4X.

Для порівняння, остання версія методу неперервного розпізнавання усної мови працює на 1/2 33МГц TMS320C31. Реалізовано метод для сотового та безпровідного зв'язку, включено розпізнавання чисел на 15 мовах. Метод базується на модифікованому алгоритмі Маркова, за базову мовну одиницю прийнято фонему. Програма реалізації методу коштує від $500 в роздріб до $5 оптом.

Розроблено також фонетичний словниковий розпізнавач, де за базову мовну одиницю прийнято звукові прототипи фонем кожної мови. Наприклад, для американського діалекту англійської мови це 43 одиниці. Словник для розпізнавання складається з моделей слів, побудованих з цих одиниць, тому він може бути легко розширений. Набір фонем може бути також використаний для визначення меж слів. Програмне забезпечення працює на DSP TMS320C30, на частоті 55МГц може одночасно працювати два розпізнавачі.

Для прикладних задач з високими вимогами надійності (автомобільне обладнання, телефонний зв'язок) розроблено метод розпізнавання окремо вимовлюваних слів у дикторонезалежному, дикторозалежному та адаптивному варіантах. Є готові словники для більш як 45 мов, можливе створення розробниками власних словників. Ця технологія застосовується в розширювачах клавіатур та інтерактивній мультимедії, в телефонії - для автоматизації операторних функцій та голосового набору, в комп'ютерній телефонії - для голосового термінального доступу до персональних комп'ютерів, в побутовій електроніці - голосове

Таб. 1. Порівняльні характеристики апаратних та програмних систем розпізнавання голосових команд та мови.

Розмір словника	Підтримка злитного мовлення	Багато-дикторність	Точність розпіз-навання	Аналіз грама-тики	Працездатність в шумному середовищі	Виробник, ресурси, вартість
Інтегровані схеми та портативні пристрої						
40	Ні	Так			Так	Summa Group, HM2007, $16
25	Ні		97%		Так	OKI VRP6679, $20
50	Так	Так			Так	VCS, Голосові номеронабирачі 2060 на TMS320C5X
14-25	Так	Так	96%-99%		Так	Sensory Inc., RCS-164, до $10
300-10000	Так	Ні			Ні	Verbex Voice Systems, Speech Commander - портативний пристрій для ПК
Програмні продукти						
62000 … 230000	Так, 160 слів/хв	Адаптивне	95%-98%	Так	Ні	Dragon Systems, Dragon NaturallySpeaking, $60
20-100000	Ні	Так	95%-99%	Ні	Ні	IBM VoiceType, $100
20000-60000	Так, 140 слів/хв	Так	95%	Так	Ні	L&H VoiceXpress, $200
500-20000	Так	Так		Так	Ні	Cognitive Technologies

керування для іграшок, відеомагнітофонів, телевізорів. Цей метод реалізовано для широкого спектру мікропроцесорів та мікроконтролерів – Intel-X86, TI-C5X, C3X, C4X та C2X, OKI 6679, NEC-V20 та V30. Для порівняння, на 486DX-33 може одночасно працювати 8 розпізнавачів. Вартість - від $500 у роздріб до $1 великим оптом.

Голосовий номеронабирач 2060 від VCS - автономне обладнання для дикторонезалежного пофонемного розпізнавання окремих слів та компресії сигналів по CELP на базі DSP TMS320C5X. Модель 2060 розпізнає 50 імен та набирає відповідний телефонний номер. Інтерфейс розпізнає команди "call", "program", "list" та має голосовий зворотній зв'язок.

Фірма Sensory Inc пропонує рішення на базі RISC-мікроконтролерів без використання DSP. Застосування апарату нейроних мереж, на відміну від статистичних методів, дозволяє ефективніше вирішити проблему багатодикторності, діалектів мови, акценту чи навіть стану диктора - додатковим стендовим тренуванням системи, а не збільшенням обчислювальних ресурсів, додатковою адаптацією до голосу диктора чи модифікацією алгоритмів.

Використання узагальнених нейроних мереж для розпізнавання мови дає надто малу точність - до 80%, але застосування мережи спеціальної архітектури дозволяє підняти точність до 95% і навіть до 99% для спеціальних застосувань.

Головна перевага методу нейроних мереж - мінімальні вимоги до обчислювальних ресурсів та низька вартість апаратних реалізацій.

Інструментарій для розробників

Користувачам систем розпізнавання мови вже не треба "винаходити велосипед", в Інтернеті доступні вільно розповсюджувані програмні засоби для проведення експериментів та дослідження різноманітних алгоритмів. Доступні базові алгоритми розпізнавання за моделлю Маркова, рекурентними та нерекурентними нейронними мережами, програми для дослідження спектрограм слів та побудови словників фонем тощо.

Для ефективної інтеграції мовних технологій в різні операційні системи розроблені та активно поширюються серед розробників відповідні інтерфейси:

- ASAPI: Advanced Speech API (AT&T),
- SAPI: Microsoft Windows Speech API,
- SRAPI: Speech Recognition API,
- TAPI: Microsoft Windows Telephony API.

Підсумок

Загалом, розпізнавання мови досягло значних успіхів у певних застосуваннях, проте більшість важливих застосувань не може бути реалізовано на базі сьогоднішніх технологій. Розробляються напрямки нотування діалогів, синхронного усного перекладу, формування вимови для тих, хто вивчає іноземні мови. На жаль, переважна більшість розробок стосується у першу чергу англійської мови та вимагає значної адаптації або фундаментальної переробки для інших мов.

Важливим фактором для успішної реалізації мовних проектів залишаються проведення фундаментальних досліджень мовних сигналів та створення великого структурованого та фонетично розміченого тестового та навчального мовного матеріалу. Існуючі рішення для слов'янських мов ще не забезпечують точність більше ніж 95%, щоб отримати широке розповсюдження. Відсутність суттєвих досягнень за останні роки вказує на необхідність додаткової підтримки з боку фундаментальних наук, що досліджують людську мову.