

АНАЛІЗ СИСТЕМ РОЗПІЗНАВАННЯ МОВИ

Використання комп'ютерних систем набагато розшириться, якщо стане можливим використання людської мови при роботі безпосередньо з комп'ютером, і зокрема стане можливим управління машиною звичайним голосом в реальному часі, а також введення і виведення інформації у вигляді звичайної людської мови.

Голосове управління засноване на технології розпізнавання мови: система отримує інформацію про коливання повітря через мікрофон, порівнює отримані дані з командами, які записані в системі і, в разі збігу, виконує запропоноване дію.

Існуючі технології розпізнавання мови не мають поки достатніх можливостей для їх широкого використання, але на даному етапі досліджень проводиться інтенсивний пошук можливостей вживання коротких багатозначних слів (процедур) для полегшення розуміння. Задача розпізнавання принаймні 20 тисяч слів природної мови - залишається поки недосяжною. Ці можливості поки недоступні для широкого комерційного використання. Зараз повсюдного поширення голосового управління заважає недостатня обчислювальна потужність процесорів і проблема наявності зовнішніх шумів.

В даний час роботи по розпізнаванню мови не тільки не втратили актуальності, а й розвиваються широким фронтом, знаходячи для себе безліч областей для практичного застосування. Зараз можна виділити 4 порівняно ізолюваних напрямки в області розвитку мовних технологій:

1. Розпізнавання мови - тобто перетворення мовного акустичного сигналу в машинний ланцюжок символів і слів. Ці системи можуть бути охарактеризовані по ряду параметрів. Перш за все це обсяг словника: малі обсяги до 20 слів, великі - тисячі і десятки тисяч. Кількість дикторів: від одного до довільного. Стиль проголошення: від ізолюваних команд до злиганої промови і від читання до спонтанної мови. Коефіцієнт розгалуження, тобто величина, яка визначає кількість гіпотез на кожному кроці розпізнавання: від малих величин ($10 < 15$) до великих ($100 > 200$). Співвідношення сигнал/шум від великих (> 30 дБ) до низьких (< 10 дБ). Якість каналів зв'язку: від високоякісного мікрофону до телефонного каналу. Якість роботи систем розпізнавання мови зазвичай характеризується надійністю розпізнавання слів, або, що те ж саме, відсотком помилок.

2. Визначення індивідуальності мовця. Ці системи поділяються на два класи: верифікація мовця (тобто підтвердження його особистості) і ідентифікація мовця (тобто визначення його особистості з заздалегідь обмеженого числа людей). Обидва ці класу далі можуть бути розділені на тексто - залежні і тексто - незалежні. Основний характеристичний параметр - обсяг паролної фрази. Два інших (як і в розпізнаванні мови): співвідношення сигнал/шум і якість каналу зв'язку. Якість роботи систем верифікації / ідентифікації мовця характеризується двома величинами: ймовірністю не впізнання «свого» диктора і ймовірністю прийняття «чужого» диктора за свого.

3. Синтез мови. Практично існує два класи: відтворення записаного в тій чи іншій формі обмеженого числа повідомлень та синтез мови по тексту. Синтезатори характеризуються за наступними параметрами: розбірливість (словесна або складова), природність звучання, стійкість перед завадами.

4. Компресія мови. Основна (і єдина) класифікаційна ознака цих систем, це ступінь компресії: від низької (32-16 кбіт/сек) до високої (1200-2400 кбіт/сек і нижче). Якість роботи систем компресії мови характеризується, перш за все, розбірливість компресованої мови. Додатковими характеристиками дуже важливими в ряді програм є впізнаваність голосу мовця і можливість визначення стресового рівня мовця.

Бібліотека розпізнавання голосових команд становить ядро системи голосового управління. З її допомогою розробники можуть додати голосове управління в створювані ними додатки.

Розробка власних засобів голосового управління вимагає спеціальних знань. Вона може відняти у розробників додатків чимало часу і сил. Однак скориставшись бібліотекою розпізнавання голосових команд, розробники можуть легко додати в додатки наступні функціональні можливості: управління обладнанням за допомогою голосу; виконання мовних запитів до баз даних через мікрофон або навіть по телефону; пошук за ключовими словами в звукових файлах. Бібліотека дозволяє вбудовувати голосові функції не тільки в звичайні програми для персональних комп'ютерів, але і в автономні пристрої, обладнані цифровими сигнальними процесорами DSP. Алгоритми, реалізовані в бібліотеці розпізнавання голосових команд, мають високу швидкість, невибагливі до обсягу оперативної пам'яті і здатні адаптуватися до шумів. Бібліотека забезпечує розпізнавання команд, вимовлених будь-яким голосом і на будь-якій мові. При цьому є можливість структурування для практично необмеженого словника.