

*Фірта Д.О., магістрант, гр. АТ-22-2м,
Воронова Т.С., асистент кафедри,
Ченюк Л.О., канд. техн. наук, доц. кафедри
Державний університет «Житомирська політехніка»*

СИСТЕМИ РОЗПІЗНАВАННЯ 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 задану раніше дію.

В даний час наукове співтовариство вкладає гігантську кількість грошей в науково-дослідні розробки для вирішення проблем автоматичного розпізнавання і розуміння мови. Це стимулюється практичними вимогами, пов'язаними зі створенням системи військового та комерційного призначення. Не торкаючись першого з них, можна вказати, що тільки в європейській спільноті обсяг продажів систем цивільного призначення становить кілька мільярдів доларів. При цьому слід звернути увагу на те, що в практичному використанні відсутні системи, метою яких є переклад мови в відповідний письмовий текст.

Якщо розглядати класичну схему «наука - технології - практичні системи», то, перш за все, треба визначити ті умови, в яких буде працювати практична система автоматичного розпізнавання або розуміння мови.

Найбільш серйозні проблеми виникають при умовах:

- довільний користувач;
- спонтанна мова, супроводжувана аграматизмами і мовним «сміттям»;
- наявність акустичних перешкод і спотворень, в тому числі змінних;
- наявність мовних перешкод.

З іншого боку необхідно визначити важливість завдання, її наукову і прикладну фундаментальність, зв'язок з іншими областями знань. При цьому необхідно враховувати стан науково-промислового потенціалу, його можливості.

Зупинимося на можливостях і недоліках відповідних систем автоматичного розпізнавання мови (анонсуються сьогодні можливістю розпізнавання сотень і навіть тисяч слів з надійністю до 98%).

Від користувача потрібна попередня настройка системи на його голос від декількох десятків хвилин до декількох годин попереднього ви-

мовлення текстів. Час обробки введеного відрізка мови в таких системах може займати хвилини. Можна виділити 4 порівняно ізольовані напрямки в області розвитку мовних технологій:

1. Розпізнавання мови - тобто перетворення мовного акустичного сигналу в ланцюжок символів, 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 ізольованих команд до зливої промови і від читання до спонтанної мови. Коефіцієнт розгалуження, тобто величина, яка визначає кількість гіпотез на кожному кроці розпізнавання: від малих величин ($<10 \div 15$) до великих ($> 100 \div 200$). Відношення 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 промовця. Ці системи діляться на два класи: верифікація промовця (тобто підтвердження його особистості) і ідентифікація промовця (тобто визначення його особистості з задалегідь обмеженого числа людей). Обидва ці класу далі можуть бути розділені на тексто-залежні і тексто-незалежні. Наступний характеристичний параметр - обсяг пароліної фрази. Два інших (як і в розпізнаванні мови): відношення сигнал / шум і якість каналу зв'язку. Якість роботи систем верифікації / ідентифікації промовця характеризується двома величинами: не впізнання «свого» диктора і ймовірністю прийняття «чужого» диктора за свого.

3. Синтез мови. Практично існує два класи:

1) Відтворення записаного в тій чи іншій формі обмеженого числа повідомлень;

2) Синтез мови по тексту. Синтезатори характеризуються за наступними параметрами: розбірливість (за словом або за складовими слова), природність звучання, стійкість перед перешкодами.

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 вище). Якість роботи систем компресії мови характеризується, перш за все, розбірливістю зкомпресованої мови. Додатковими характеристиками дуже важливими в ряді програм є впізнаваність голосу промовця і можливість визначення стресового рівня промовця.