УДК 519.68

Ю.М. Тесля, Ю.О. Остапчук, І.М. Бахмач, О.О. Кучеренко

Київський національний університет будівництва і архітектури, Київ

СУЧАСНИЙ СТАН КОМП'ЮТЕРНИХ СИСТЕМ РОЗПІЗНАННЯ МОВЛЕННЯ

Розглянуто сучасний стан розвитку комп'ютерних систем розпізнання природної мови, виокремлено основні пріоритети розвитку та основні типи проблем з якими стикаються розробники систем розпізнання голосу на даному етапі їх розвитку.

Ключові слова: методології розпізнавання природної мови, голосове управління, комп'ютерні системи

Сучасний стан комп'ютерних систем розпізнання мови

Використання сучасних, але високо-інтелектуальних інформаційних комп'ютерних технологій у сфері людської діяльності вимагає кардинальної зміни в управлінні автоматизованими системами для більш зручного та раціонального їх використання. Необхідність в мовному спілкуванні з комп'ютером є абсолютно природною. Найбільшою мірою її стимулює не стільки бажання створити більші зручності користувачу, скільки існування специфічних областей комп'ютеризації, де голосові команди є найбільш придатними чи навіть єдиним можливим рішенням. До них можна віднести голосовий доступ до автоматичних довідкових систем, керування віддаленим комп'ютером чи портативним пристроєм, що відбувається під час руху. Створення повноцінних мовних інтерфейсів, які підтримують діалог «користувач – комп'ютер» є дуже перспективним, але надзвичайно складним напрямом розвитку сучасних комп'ютерних систем, що стикаються з велетенською кількістю проблем на шляху їх вирішення [10].

На сьогодні, під поняттям «розпізнавання голосу» прихована ціла сфера наукової та інженерної діяльності [8]. В цілому, завдання розпізнавання голосу зводиться до того, щоб виділити, класифікувати та відповідним чином відреагувати на людський голос з вхідного звукового потоку. Це може бути виконання певної дії на команду людини чи виокремлення певного слова-маркера з великого масиву телефонних розмов, чи система для голосового вводу тексту. Також всім відомі програми голосової ідентифікації користувачів, що реалізовані в деяких системах безпеки. Потенційно, сфера використання голосового розпізнання надзвичайно широка, але, на жаль, на даний момент не може бути реалізована внаслідок слабкої стійкості самих систем розпізнання мови до різних факторів.

Ознаки класифікації систем розпізнання мови

Кожна система розпізнання мови має певні задачі, які вона створена вирішувати, та комплекс методів котрий використовується для рішення цих задач [1]. Класифікація систем розпізнання мови буде проводитися згідно нового стандарту прийнятого в галузі програмування таких систем - Microsoft Speech API. Згідно з цим стандартом системи розпізнання мови розрізняються за певними ознаками.

- *Інтервал між окремими словами.* Якщо система розпізнає здільну мову, користувач може вимовляти фрази в природному вигляді, не роблячи проміжків між словами. Неперервне розпізнання має перевагу, але його реалізація більш складна та вимагає більших апаратних можливостей комп'ютерів, результатом чого є мала кількість таких систем. В системах, що працюють з дискретною мовою диктор має робити паузи між окремими словами, як правило не менше 1/4 секунди. Третім різновидом є системи, які виділяють одне слово – маркер, в певному мовному інтервалі. Навіть, якщо маркер знаходиться всередині фрази вимовленої здільно.
- *Залежність від диктора.* За визначенням система залежна від диктора призначена для використання одним користувачем, в той час, як альтернативні системи призначені для роботи з будь-яким диктором. Незалежність від диктора – складна задача оскільки під час навчання системи вона налаштовується на параметри голосу диктора, на прикладі якого вона навчається. Кількість помилок в таких системах, як правило в 4-5 разів більша, ніж в

системах залежних від диктора. Системи, що володіють відносною незалежністю від диктора, дозволяють працювати з ними без попереднього налаштування, навчання системи, однак результати все таки є кращими, за умови навчання системи. Незалежність від диктора, як правило, досягається за рахунок збереження звукових еталонів для всіх найбільш типових голосових носіїв даного типу, що в результаті ставить більші апаратні вимоги до таких систем. Процес навчання, налаштування під диктора, як правило, займає від 30 хв. до кількох годин. Саме цей факт є головною незручністю для користувачів. Третім різновидом за даною ознакою є системи, що автоматично налаштовуються на голос диктора в процесі їх експлуатації. У систем такого типу є дві особливості: їм необхідно знати чи зробив користувач помилку, вимовляючи те чи інше слово (інакше навчання буде не вірним); після налаштування на конкретного диктора, ці системи стають менш надійними при роботі з іншим диктором.

- *Ступінь деталізації при задаванні еталонів.* Розрізняють алгоритми, в яких за еталони приймають цілі слова та алгоритми, що використовують в якості еталонів частини слів. Порівняння цілих слів дає більшу точність, швидкість, але при цьому вимагає більшого обсягу пам'яті. Алгоритми порівняння елементів слів (фонем, складів і т.д.) доводиться використовувати у випадку великих словників, оскільки об'єм необхідної пам'яті пропорційний кількості цих еталонних слів та не залежить від об'єму словника.
- *Розмір словника.* Системи розпізнання можуть використовувати як великі, так і маленькі словники. Системи, що працюють з маленькими словниками (близько 50 слів), дозволяють користувачу давати комп'ютеру прості команди. Для диктування текстів необхідний великий словник (десятки тисяч слів). Очевидно, що чим більший розмір словника, котрий закладено в систему розпізнання, тим більша частота помилок під час роботи системи. Наприклад, словник із 20 слів може бути розпізнано майже без помилок, тоді як частота помилок при роботі зі словником в 1000 слів може досягати 45%. З іншого боку, навіть розпізнання невеликого словника може дати велику кількість помилок, якщо слова в даному словнику дуже схожі одне на одне.

Не дивлячись на те, що в теорії можлива будь-яка комбінація даних характеристик, на практиці найбільш популярними є системи голосового управління комп'ютером та систем дискретного диктування тексту.

Різновиди методів розпізнання голосу

У процесі створення системи розпізнання голосу потрібно обрати рівень абстракції адекватний поставленій задачі. Параметри звукової хвилі мають використовуватися для розпізнання та методів розпізнання цих параметрів [5]. Можна виокремити таку основну різницю в структурі і процесі роботи різноманітних систем розпізнання голосу:

- *За типом структурної одиниці.* У процесі аналізу голосу, як базові одиниці можуть бути обрані окремі слова чи частини вимовлених слів: фонеми, ди- чи трифони, аллофони. Залежно від того, яка структурна частина обрана, змінюється структура, універсальність та складність словника елементів, що розпізнається.
- *За виділенням ознак.* Сама послідовність відрізків тиску звукової хвилі – надмірно збиткова для систем розпізнавання звуків та містить багато зайвої інформації, яка для розпізнання не потрібна чи навіть шкідлива. Таким чином, для представлення голосового сигналу з нього слід виокремити усі параметри, що адекватно представляють даний сигнал для розпізнання.
- *За механізмом функціонування.* В сучасних системах широко використовуються різноманітні підходи до механізму функціонування розпізнавальних систем. Імовірнісно-мережевий підхід полягає в тому, що голосовий сигнал розбивається на певні частини (кадри або за фонетичною ознакою), після чого імовірнісна оцінка того, до якого саме елементу словника, що розпізнається має відношення дана частина і (чи) весь вхідний сигнал. Підхід, оснований на рішенні зворотної задачі синтезу звука, полягає в тому, що за вхідним сигналом визначається характер руху артикулярів мовного каналу та за спеціальним словником відбувається визначення вимовлених фонем.

Для кращого розуміння особливостей задач розпізнання мови слід відмітити, що основна маса систем працюють практично однаково, використовуючи переважно одні й ті ж методи та алгоритми [7]. Різниця полягає в манері диктування голосу, розмірі словника, ступені фільтрації вхідного сигналу, обумовлена лише специфікою задачі та наявними технічними можливостями. Якщо спробувати представити спрощено процес розпізнання голосу, то він може бути описаний в послідовності таких кроків:

- фільтрація шуму та виокремлення необхідного сигналу;
- перетворення вхідного голосового сигналу в набір акустичних параметрів;
- приведення акустичної форми сигналу до внутрішнього алфавіту еталонних елементів;
- розпізнання послідовності фонем та перетворення їх на слова.

Класичний вид системи розпізнання голосу

Розпізнання голосу – це багаторівнева задача розпізнання образів, в якій акустичний сигнал аналізується та структурується в ієрархію структурних елементів, наприклад, фонем, слів, фраз та речень [4]. Кожен рівень ієрархії може передбачати деякі часові константи, наприклад, можливі послідовності слів чи відомі види вимовляння, які дозволяють зменшувати кількість помилок розпізнання на більш низькому рівні. Чим більше ми знаємо апріорної інформації про вхідний сигнал, тим якісніше ми можемо його опрацьовувати та розпізнавати. Якщо спробувати представити класичний варіант системи розпізнання голосу, то він може мати такий вигляд:

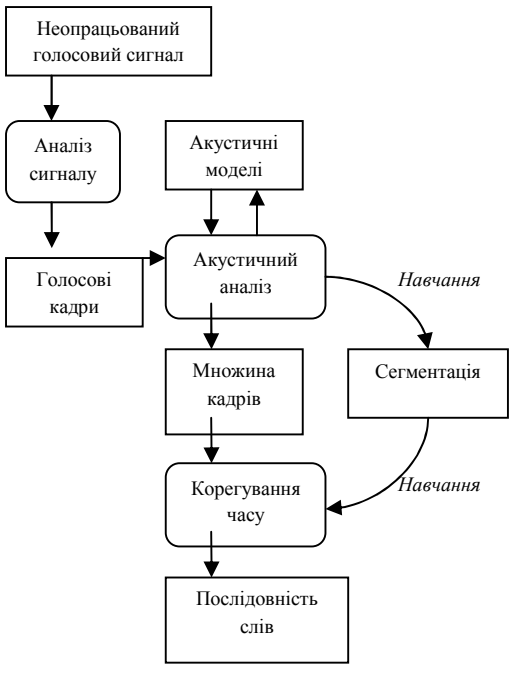

Рис.1. Модель системи розпізнання голосу

Неопрацьований голосовий сигнал. Як правило, це потік звукових даних, записаний з високою дискретизацією (20 кГц при записі з мікрофона чи 8 кГц при записі з телефонної лінії).

Аналіз сигналу. Сигнал, що надходить має бути початково трансформований та стиснений, для полегшення подальшого опрацювання. Є різноманітні методи для виокремлення корисних параметрів та стиснення початкових даних в десятки разів без втрати корисної інформації. Найбільш популярні: аналіз Фурьє, лінійне прогнозування мови, кепстральний аналіз.

Голосові кадри. Результатом аналізу сигналу є послідовність голосових кадрів. Зазвичай, кожен голосовий кадр – це результат аналізу сигналу на невеликому відрізку часу (близько 10 мс), що містить інформацію про дану ділянку.

Акустичні моделі. Для аналізу складу голосових кадрів потрібен набір акустичних моделей. Найбільш розповсюдженими з них є:

- *Шаблонна модель.* Як акустична модель виступає яким-небудь чином збережений приклад розпізнаної структурної одиниці (слова, команди). Варіативність розпізнання такою моделлю досягається шляхом збереження різноманітних варіантів вимовляння одного й того ж елементу (перелік дикторів багато разів повторюють одну й ту ж команду). Використовується переважно для розпізнання слів, як єдиного цілого (командні системи).
- *Модель стану.* Кожне слово моделюється, як послідовність станів, що вказують на набір звуків, які можна почути в даній ділянці слова, базуючись на імовірнісних правилах. Цей підхід використовується в більш масштабних системах.

Акустичний аналіз. Полягає у зіставленні різноманітних акустичних моделей до кожного кадру голосу та видає матрицю зіставлення послідовності кадрів та множини акустичних моделей. Для шаблонної моделі ця матриця являє собою Евклідову відстань між шаблонами і відстанями кадрів (тобто вираховує як сильно відрізняється отриманий сигнал від записаного шаблону й знаходиться шаблон, який найбільш підходить до отриманого сигналу). Для моделей основаних на стані, матриця складається з ймовірності того, що даний стан може згенерувати даний кадр.

Коригування часу. Використовується для опрацювання часової варіативності, виникаючої під час вимовляння слів (наприклад, «розтягуванні» чи «ковтанні» звуків).

Порядок слів. В результаті роботи, система розпізнавання голосу виділяє послідовність (чи декілька імовірних послідовностей) слів, котра, найбільш ймовірно, відповідає вхідному потоку голосу.

Проблеми та перспективи їх рішення

Беручи до уваги все викладене, можна виокремити проблеми, які стоять перед розробниками систем розпізнання голосу.

Проблема подолання стаціонарних та нестаціонарних перешкод [2]; [3]. Наявні на даний момент системи голосового керування комп'ютером і диктування тексту практично не використовують в своїй роботі алгоритми подолання шумів. Це пов'язано з тим, що дані системи використовуються, як правило, в умовах дому чи офісу, де рівень зовнішніх шумів мінімальний. Відсутність подолання шуму в комп'ютерних голосових системах відбивається на кількості помилок під час розпізнання.

Проблема переходу до розпізнавання неперервного голосу. Ця проблема обумовлена недоліками технічних характеристик персональних комп'ютерів, що робить на даний момент системи диктування здільної мови занадто дорогими, тому непопулярними.

Проблема аналізу контексту. На сьогодні для врахування контексту (синтаксису та семантики) при відновленні хронології вимовлених слів використовують, як правило, мінімальний набір правил [6]. У подальшому слід очікувати ускладнення граматичних підходів пов'язаних зі специфікою певної мови.

Проблема пошуку нових звукових параметрів. На сьогодні для розпізнання голосу в основному використовують спектральні параметри голосу – швидке перетворення Фурьє, спектр лінійного прогнозування, кепстральні коефіцієнти [9]. Ці параметри мають як ряд переваг, так і недоліків (залежність спектральних параметрів від голосу диктора).

Проблема пошуку нових алгоритмів відновлення звукової черги. На сьогодні наявні алгоритми відновлення черги вимовлених звуків практично вичерпали свій потенціал збільшення точності розпізнавання голосу, тому в найближчому майбутньому слід очікувати створення нових підходів до рішення даної проблеми.

Список літератури

1. Информационное Агентство "Алгоритм". Распознавание речи: еще один тупик. AlgoNet. [З мережі] http://www.algonet.ru/?ID=180615.

2. Ализар, Анатолий. Незаметная смерть распознавания речи. Хабрахабр. [З мережі]4 травень2010р.http://habrahabr.ru/blogs/artificial_intelligence/92771/.

3. Курочкин С.Н., Бродин А.Г. Проблеми создания многоуровневой системи распознавания речи. Официальный сайт МГТУ "Станкин". [З мережі]1997р. http://magazine.stankin.ru/arch /n_02/automation/art05.html.

4. Савенкова О.А., Карпов О.Н. Технология построения интеллектуальной системы распознавания речи. Національна бібліотека України імені В. І. Вернадського. [З мережі] 17.08.2008р. http://www.nbuv.gov.ua/portal/natural/ii/2008_4/JournalAI_2008_4/Razdel9/00_Savenkova_Karpov.pdf

5. Веренич И.В. Анализ методов построения систем распознавания речи на основе гибрида скрытой марковской модели и нейросети. Портал магистров ДонНТУ. [З мережі] 2008 р. http://masters.donntu.edu.ua/2008/fvti/verenich/diss/index.htm .

6. Галунов В.И., Соловьев А.Н. Современные проблемы в области распознавания речи. Портал магистров ДонНТУ. [З мережі] http://masters.donntu.edu.ua/2008/fvti/verenich/library/darkness.htm.

7. Гребнов С.В. Аналитический обзор методов распознавания речи в системах голосового управления. ИГЭУ. [З мережі] 2009р. http://www.ispu.ru/files/%2083-85.pdf.

8. Мазуренко И.Л. Компьютерные системы распознавания речи. Интелектуальные системы. [З мережі]1998р. http://www.intsys.msu.ru /magazine /archive/v3 (1-2)/mazurenko.pdf.

9. Фролов А.В., Фролов Г.В. Синтез и распознавание речи. Современные решения. Электронная библиотека книг братьев Фроловых. [З мережі] 2003 р. http://frolov-lib.ru/books/hi/index.html.

10. Интернет-портал "История компьютера". История компьютера - Распознавание речи. История компьютера. [З мережі] http://chernykh.net/component/option,com_joomap/Itemid,63/.

Стаття надійшла до редколегії 21.10.2011

Рецензент: д-р техн. наук, проф. С.Д.Бушуєв , Київський національний університет будівництва і архітектури, Київ.