УДК 519.688:004.522

СИСТЕМАТИЗОВАНИЙ ОГЛЯД ПРОГРАМНИХ ЗАСОБІВ ОБРОБКИ ЗВУКОВИХ ДАНИХ

Маркелов О.Е.[1)], Мельник М.Р.[2)], Косовський В.М.[3)]
Національний університет «Львівська політехніка»
[1)] старший викладач; [2)] асистент; [3)] магістр

І. Постановка проблеми

Інтелектуальні мовні рішення, що дозволяють автоматично синтезувати і розпізнавати людську мову, є наступною сходинкою розвитку інтерактивних голосових систем [1]. Такі технології призначені для звукового дубляжу і реагувати на людський голос. Вони мають безліч застосувань: спілкуватися з комп'ютером без клавіатури, послуги зв'язку, надання допомоги людям з обмеженими аудіо можливостями тощо [2].

ІІ. Мета роботи

Необхідно створити систему інтелектуальної обробки звукових даних, для цього потрібно знайти бібліотеки програмних засобів.

ІІІ. Огляд та короткі характеристики програмних бібліотек

Бібліотека **dsPIC30F** підтримує голосове управлянні додатками зі словником до 100 слів. Вимагає невеликих затрат ресурсів для обробки даних. Призначена для роботи з контролерами на процесорах: dsPIC30F5011, dsPIC30F5013, dsPIC30F6012 і dsPIC30F6014. Характерною особливістю бібліотеки є: обробка даних в той момент, коли виявлено закінчення слова, отримані слова ідентифікуються за допомогою прихованих моделей Маркова [3].

Sphinx − система розпізнавання мови, повністю написана на мові програмування Java™. Така система здатна розпізнавати окремі слова та неперервний набір слів. Узагальнена модульна архітектура включає в себе інструмент попередньої корекції, зважувальну функцію Хеммінга, швидке перетворення Фур'є, частотний фільтр, дискретне косинусоїдальне перетворення.

Pocketsphinx − це бібліотека яка є залежною від іншої бібліотеки під назвою SphinxBase яка забезпечує загальну функціональність усіх CMUSphinx проектів [4].

Julius - високопродуктивна бібліотека для розпізнавання звукових даних. Використовується два проходи для пошуку слів у словнику, пошуку слів здійснюється в реальному часі. Бібліотека незалежна від структурних моделей, можуть бути використані різні варіації типів прихованих моделей Маркова. Основною операційною системою є Linux, але є можливість також використання бібліотеки у операційній системі Windows. Для використання бібліотеки необхідна модель природної людської мови, а також акустичні моделі мови. Julius розпізнає акістичні моделі у форматі ASCII, HTK, ARPA [5].

iATROS − це нова реалізація попередніх методів розпізнавання мови яка була адаптована для рукописного розпізнавання тексту. iATROS має модульну структуру, яка може бути використана для створення різних систем. iATROS забезпечує стандартні інструменти для обробки і розпізнавання, он-лайн розпізнавання мови (на базі модулів ALSA). iATROS складається з двох модулів попередньої обробки, функції вилучення та основного модулю розпізнавання. Модулі попередньої обробки і функція вилучення аудіо даних забезпечують видобуток векторів ознак для визнання мовних асоціацій за допомогою прихованих моделей Маркова і мовних моделей, виконує пошук кращих гіпотез голосових даних [6].

RWTH ASR − бібліотеки розпізнавання мови з відкритим програмним кодом. Бібліотеки включають в себе інструменти для розробки акустичних моделей і декодерів, а також інструменти для навчання. Підтримуються акустичні моделі у форматі ARPA. Програмне забезпечення працює на Linux і Mac OS X. Інструментарій опублікований під відкритою ліцензією яка є похідною ліцензією від QPL [7]. Дана ліцензія надає безкоштовне використання в тому числі і для некомерційного використання [8].

Компанія IBM реалізувала специфікацію **JSAPI** та створила інструмент для розпізнавання звукових даних на мові програмування Java™, який заснований на ViaVoice технології, що забезпечує безперервне розпізнавання звукових даних (розпізнавання мови) і перетворення тексту в мову (синтез мови). ViaVoice включає можливість розширення словникового запасу до двох мільйонів слів. На даний час підтримуються наступні мови: англійська, португальська, французька,

німецька, італійська та іспанські мови повністю, а також японська. Інструмент працює на платформах операційних систем Windows та Linux і може бути завантажений з веб-сайту IBM AlphaWorks [9].

NICO SPEECH − інструментарій, що розроблений для автоматичного розпізнавання звукових даних. Включає набір інструментів для редагування звукових даних. NICO SPEECH може слугувати для створення фонем. Фонеми можуть бути об'єднані у файли з мітками («label files»), що містять одну фонему в основному рядку. Ці файли в свою чергу можуть бути також об'єднані в N-вихідних одиниць (одна одиниця для кожної фонеми). «LabelListFile» є списком різних фонем таких, що кількість цільових значень в слові дорівнює числу рядків в «LabelListFile». Крім того, програма повинна знати кількість зразків для кожного висловлювання, тому потрібно дати файл даних з одного довільного іншого потоку і розмір цього потоку. Каталоги і розширення вхідного файлу, мітка файлу може бути задана з різними властивостями [10].

З метою порівняння бібліотек програмних засобів, в таблиці 1 зведено їх основні характеристики.

Таблиця 1

Характеристики бібліотек розпізнавання звукових даних

Бібліотека / Характеристика	dsPIC30F	Pocket Sphinx	Julius	iATROS	RWTH ASR	ViaVoice 0.02	NICO SPEECH
1	2	3	4	5	6	7	8
Мова програмування	Асемблер	C	C	C	C++	java	C
Акустична модель, kHz	16	8; 16	16	8; 16	16	16	16
Формат вхідних даних	WAV	WAV	WAV, RAW	WAV	WAV	WAV	WAV, kth, cmu, Au, nist , binary
Тип акустичної моделі	Напівне-перервна	Напівне-перервна, зв'язна, неперервна	Напівне-перервна, зв'язна, неперервна	Напівне-перервна, зв'язна, неперервна	Напівне-перервна, зв'язна, неперервна	Неперерв-на	Зв'язна
Час розпізнавання, msec	<500	Залежить від параметрів апаратної системи	Залежить від параметрів апаратної системи	Залежить від параметрів апаратної системи	Залежить від параметрів апаратної системи	Залежить від параметрів апаратної системи	Залежить від параметрів апаратної системи
Навчання (внесення звукових еталонів)	Так	Так	Так	Так	Так	Так	Так
Розмір словника, слів	100000	2000	60000	30000	90000	280000	-
Тип ліцензії	GNU GPL	BSD	BSD	GNU GPL	QPL	LGPL	BSD
Ціна	$5	-	-	-	-	-	-
Особливості інтеграції	Статична Бібліотека	Статична бібліотека	Статична бібліотека	Динамічна бібліотека	Статична бібліотека	Динамічна бібліотека	Статична бібліотека
Операційна система	Windows	Linux, Windows, Mac OS X, iPhoneOS	Linux, Windows, Mac OS X, FreeBSD, Sun Solaris	Mac OS X	Linux, Mac OS X	Windows, Linux, Mac OS X	Linux, Windows, Sun Solaris

Формули визначення придатності бібліотеки до застосування на основі критеріальної функції: мінімізація витрат. Необхідно знайти серед елементів x, що утворюють множину X, такий елемент x*, що виражає мінімальне значення функції $f(x^*)$ заданої функції $f(x)$. Допустима множина X:

$$X = \{\vec{x} \mid g_i(\vec{x}) \le 0, i = 1,...,7\} \subset R^n, \quad \text{де} \tag{1}$$

$g_i(\vec{x})$ − обмеження цільової функції $f(x)$.

Зображення множини X в просторі R: $X \to R$

(2)

Тоді вирішити задачу:

$$f(x) \to \min_{\vec{x} \in X} \tag{3}$$

означає знайти $\vec{x}^*$, таке що:

$$\vec{x}^* \in X : f(\vec{x}^*) = \min_{\vec{x} \in X} f(\vec{x}) \tag{4}$$

Максимізація функціональності. Необхідно знайти серед елементів y, що утворюють множину Y, такий елемент y*, що виражає максимальне значення функції $l(y^*)$ заданої функції $l(y)$.

Допустима множина Y:

$$Y = \{\vec{y} \mid k_j(\vec{y}) \le 0, j = 1,...,7\} \subset N^n, \text{ де} \tag{5}$$

$k_j(\vec{y})$ − обмеження цільової функції $l(y)$.

Зображення множини Y в просторі N: $Y \to N$

(6)

Тоді вирішити задачу:

$$l(y) \to \max_{\vec{y} \in Y} \tag{7}$$

означає знайти $\vec{y}^*$, таке що:

$$\vec{y}^* \in Y : l(\vec{y}^*) = \max_{\vec{y} \in Y} l(\vec{y}) \tag{8}$$

Висновок

У роботі досліджено програмні бібліотеки розпізнавання звукових даних, наведено їх систематизовану порівняльну характеристику. Використовуючи формули для мінімізації витрат та максимізації функціональності визначаємо програмну бібліотеку, котра найбільш придатна для реалізації системи розпізнавання звукових даних. Такою бібліотекою є ViaVoice 0.02.

Список використаних джерел

3. 10,000+ Speech Topics Problem Solution Speech Topics [Електронний ресурс] / Speech Topics Help, Advice & Ideas. – 2012 – Режим доступу:http://www.speech-topics-help.com/index.html. – Назва з домашньої сторінки інтернету.
4. Myron D. Speech Technology [Електронний ресурс] / Myron D. – 2001. – Режим доступу:http://www.speechtechmag.com. – Назва з домашньої сторінки інтернету.
5. dsPIC30F Speech Recognition Library [Електронний ресур] : (проект) / Microchip Technology Inc. – Електрон. дан. (4 файли) – 2004-2008 - Режим доступу: http://www.microchip.com/stellent/idcplg?IdcService=SS_GET_PAGE&nodeId=1406&dDocName=en023596. – Назва з домашньої сторінки інтернету.
6. CMUSphinx [Електронний ресурс] / Carnegie Mellon University. – 2012. – Режим доступу: http://cmusphinx.sourceforge.net/wiki/tutorialpocketsphinx. – Назва з домашньої сторінки інтернету.
7. Open-Source Large Vocabulary CSR Engine Julius [Електронний ресурс] / Kawahara Lab., Kyoto University. – 1991-2011. – Режим доступу:http://julius.sourceforge.jp/juliusbook/en/. – Назва з домашньої сторінки інтернету.
8. iDoc: Interactive Analysis, Transcription and Translation of Old Text Documents [Електронний ресурс] : (проект) / Alabau V., Luján M., Pastor M. – 2006-2009. – Режим доступу:https://prhlt.iti.upv.es/page/projects/multimodal/idoc/iatros. – Назва з домашньої сторінки інтернету.
9. Trolltech AS. The Q Public License Version 1.0 [Електронний ресурс] / Trolltech AS. – 1991. – Режим доступу:http://www.opensource.org/licenses/QPL-1.0. – Назва з домашньої сторінки інтернету.
10. The RWTH Aachen University Speech Recognition System. RWTH ASR – [Електронний ресурс] / J. Lööf, C. Gollan, S. Hahn, G. Heigold, B. Hoffmeister, C. Plahl, D. Rybach, R. Schlüter, H. Ney. – 2007. – Режим доступу:http://www-i6.informatik.rwth-aachen.de/rwth-asr/. – Назва з домашньої сторінки інтернету.
11. Satish Swaroop Create applications with speech recognition and synthesis using IBM Speech for Java [Електронний ресурс] / Satish Swaroop. – 2001. – Режим доступу:http://www.ibm.com/developerworks/ibm/library/i-voice/. – Назва з домашньої сторінки інтернету.
12. Neural Inference COmputation SPEECH TOOLS [Електронний ресурс] / department for Speech, Music and Hearing at KTH, Stockholm. – 2006. – Режим доступу:http://nico.nikkostrom.com/doc/sptools.html. – Назва з домашньої сторінки інтернету.