SESSION 7 Speech Recognition & Synthesis for Human-Robot Interaction







Desirable Speech Characteristics

Naturalness

Intelligibility





Why is :+

Difficult?



SSUES

Pronunciation: -Ambiguous words

-Punctuation

- Emotion Content



Peter Piper picked a peck of pickled peppers; A peck of pickled peppers Peter Piper picked; If Peter Piper picked a peck of pickled peppers, Where's the peck of pickled peppers Peter Piper picked?

SSUES

Pronunciation: -Ambiguous words







Voice

Speech Recognition

State of the art







IC Module

Speech Open SDK













Google Cloud Platform



Speech Integrated Hardware

LD3320 非特定语音识别(不需要录音训练)

科大讯飞 XFS5152 模块

<u>DFRobot 语音合成 Shield</u>

SYN7318 语音识别和合成













What should we use?

(inter REAL SENSE TECHNOLOGY

Intel RealSense SDK







Command and Control

Dictation

Process

Identify an Audio Source

Configure the Module

Set the mode

Execution Flow

Internal Speaker

- ・sr.StopRec(); //停止识别

RealSense built-in Microphone

·pinfo.language=rss.LanguageType.LANGUAGE_US_ENGLISH; // 设定语言

・sr.SetDictation(); // 听写模式 ・sr.SetGrammar(1); //命令控制模式

・sr.StartRec(source, handler); //开始识别

Language

LANGUAGE_US_ENGLISH LANGUAGE_GB_ENGLISH LANGUAGE_DE_GERMAN LANGUAGE_US_SPANISH LANGUAGE_LA_SPANISH LANGUAGE_FR_FRENCH LANGUAGE_IT_ITALIAN LANGUAGE_JA_JAPANESE LANGUAGE_CN_CHINESE LANGUAGE_BR_PORTUGUESE Portuguese (ptBR)

US English (enUS) British English (enGB) German (deDE) US Spanish (esUS) French (frFR) Italian (itIT) Japanese (jaJP) Mandarin (zhCN)

Configure

TimeStamp

The recognition time stamp, in 100 ns.

Scores

The top most likely voice recognition results. Unused scores have a zero confidence level.

Grammar

The grammar identifier (for the command and control mode.)

Duration

The duration of the recognized command, in 100 ns.

Commands

String[] cmds=new String[3]{ "One", "Two", "Three" }; sr.BuildGrammarFromStringList(1, cmds, null);

Data

Label : The label of the recognized command.
Confidence: The recognition confidence level, from 0 to 100.
Sentence: The recognized text from dictation or command and control.
Tags: Reserved.

Alert

ALERT_VOLUME_HIGH The volume is too high for recognition. **ALERT_VOLUME_LOW** The volume is too low for recognition. **ALERT_SNR_LOW** The signal-to-noise ratio is too low for recognition. **ALERT_SPEECH_UNRECOGNIZABLE** The speech is detected but not recognized. **ALERT_SPEECH_BEGIN** The speech is detected and starting. **ALERT_SPEECH_END** The speech is detected and ended. **ALERT_RECOGNITION_ABORTED** The recognition is aborted due to error such as the engine internal errors or audio source failure.

ALERT_RECOGNITION_ENDED The recognition is completed. The audio source no longer provides any data

Process

// Synthesize the text
string tts.BuildSentence(1, "Speak this");
// Retrieve the synthesized speech
int nbuffers=tts.QueryBufferNum(1);
for (int i=0;i<nbuffers;i++) {
// send audio to the audio output device ...
PXCMAudio audio=tts.QueryBuffer(1, i);}
// Clean up
tts.ReleaseSentence(1);</pre>

Configuring Synthesis Quality

volume (音量): Set the volume level. The value is between 0 (mute) and 100 (the maximum volume). The default value is 80. pitch (音调): Scale the voice pitch with a factor between 50 (half the default pitch) and 200 (two times the default pitch). The default value is 100. rate (语速): Set the speaking rate. The value is between 50 (half the default rate) and 400 (four times the default rate). The default value is 100. eosPauseDuration (暂停延时): Set the end of sentence pause duration to a value between 0 and 9.

eosPauseDuration (暂停延时): Set the end of sentence pause duration to a value between 0 and 9. The pause is about 200 milliseconds multiplied by the value.

ESC Control Codes

All the control sequences follow this general syntax notation:

<ESC>\<parameter> = <value> \

Configuring Synthesis Quality

Inserting a pause Guiding text normalization Inserting a bookmark Changing the speaking rate Changing the volume Setting the end-of-sentence pause duration Setting the spelling pause duration

- Controlling end-of-sentence detection
- Controlling the read mode
- Changing the voice
- Labeling text for language identification
- Indicating a paragraph break
- Identifying a word to accent within a sentence
- Resetting control sequences to the default

size the speech e_t maxLength = 256; entence[maxLength] = {}; ce(hWnd, sentence, sizeof(sentence) / sizeof(pxcCHAR)); nlen_s(sentence, maxLength) <= 0)</pre>

geBox(hWnd, L"String should not be empty", L"Empty string!", MB_OK); /_s(sentence, L"String should not be empty");

cf(sentence, L"大哥\x1b\\pause = 1000\\ 你玩摇滚?"); cf(sentence, L"\x1b\\voice=samantha\\ Hello, this is Samantha. \x1b\\voice=tom\\ Hello, this is Tom. "); f(sentence, L"I am the person. I\x1b\\nlu=PRM:0\\ am the person. I\x1b\\nlu=PRM:1\\ am the person. I\x1b' f(sentence, L"hello\x1b\\pause = 500\\ mike"); :f(sentence, L"\x1b\\tn = time\\10:00\x1b\\tn = normal\\"); [5] = 27;sentence, L"The address reads \x1b\\tn=address\\ 2111 NE 25th Ave, Hillsbo. \x1b\\tn=spell\\Hillsbo."); :f(sentence, L"\x1b\\sent_accent\\John is coming tomorrow. John is coming \x1b\\sent_accent\\tomorrow.");

.ldSentence(1, sentence);

Error List

• 4 ×

按照如下说明接线,然后下载程序测试即可;

NE DOODO

1.SYN7318语音模块供上电(5V,模块的5V与GND对应接到Arduino UNO主控板的5V与GND引脚上); 2.插上小喇叭到SYN7318语音模块(模块对应S+,S-); 3.在void loop()中调用功能函数实现对应的功能; 4.下载本程序到Arduino中; 5.杜邦线连接RX(0)-TX、TX(1)-RX(下载完程序之后再接线)