

# 基于 Spark SDK 的连接数字语音识别

韩丹, 姚佩阳, 陈玲

(空军工程大学 电讯工程学院, 陕西 西安 710077)

**摘要:** Spark SDK 是 L&H 公司推出的可进行二次开发的语音识别软件, 它提供了一种快速简便的方法, 使开发人员可以将语音部分识别功能添加到他们的应用程序中。在雷达情报口报录取器中, 利用 Spark SDK 实现连接数字语音识别, 并通过编写识别专业词汇的函数以及修改 Spark SDK 的语法规则等方法提高了语音识别软件的识别率。

**关键词:** Delphi; Spark SDK; 连接数字语音识别

**中图分类号:** TN912.3    **文献标识码:** A    **文章编号:** 1009-3516(2002)04-0022-04

Spark SDK 是一个功能强大、性能优良的汉语普通话语音命令开发工具包<sup>[1]</sup>, 由于其可扩展性、可重用性和可再开发性等开发优势, 越来越多地为人们所应用。本文将 Spark SDK 应用于雷达情报口报录取器的研制中, 语音信号经预处理<sup>[2]</sup>, 通过语音识别软件识别出用于表示雷达情报的连接数字语音, 再经过情报整编, 完成口报雷达情报的识别。但是 Spark SDK 是针对语音命令而开发的工具包, 用于连接数字语音识别时, 还存在不能识别专业词汇, 不能对可变长数字语音以及连续无间断语音进行识别等困难。为解决这一问题, 我们采用 Delphi 语言, 编写识别专业词汇的函数 detect(), 修改 Spark SDK 的语法规则, 并对 Windows API 函数进行调用, 以实现小词汇量、连接数字语音识别。该方法较好的解决了 Spark SDK 连接数字语音识别, 同时, 该程序所用到的许多技巧在其它的程序编制中也有着相当的使用价值。

## 1 程序设计思路

该程序必须能准确识别雷达情报口报员的专业化词汇发音(例如: 1 读“邀”, 2 读“两”, 7 读“拐”, 9 读“勾”, 0 读“洞”, 将口报员的发音与 detect() 函数进行比较, 识别专业词汇。在程序开始之前先定义语法规则, 程序通过窗口消息向语音识别内核发出控制命令(开始和停止), 对语音识别内核进行性别、口音和用户信息等初始化(SPKInitialize), 准备开始识别(SPKSRStartRecognition), 识别过程中调用 Windows API 函数完成连接数字语音的识别。

## 2 具体实现

### 2.1 语法管理、初始化和清除

#### 2.1.1 语法管理

语音识别引擎依靠语法来定义引擎能够识别的词或短语, 语法内容由汉字字符串或拼音字符串组成。引擎利用关键词定位的方法, 将雷达操作员输入的语音与语法文件中的词或短语进行匹配, 程序获取匹配的结果进行识别。语法文件的格式如下:

```
% BEGIN < qingbao >  
< digit > + < * >  
% END < qingbao >
```

收稿日期: 2001-06-27

作者简介: 韩丹(1976-), 女, 吉林东丰人, 硕士生, 主要从事语音识别研究。

```
% BEGIN < digit >
洞 | 邀 | 两 | 三 | 四 | 五 | 六 | 拐 | 八 | 勾 //关键词
% END < digit >
```

Spark SDK 提供 SPKSRNewCommandGrammarEx 和 SPKNewGrammarFromFileEx 两个接口函数创建语法和命令短语列表,根据需要可利用其它接口函数进行语法的动态创建和复杂语法的修改。

### 2.1.2 初始化和清除

应用程序要使用 Spark 的语音识别功能,必须首先调用 SPKInitialize 函数对语音识别引擎进行初始化。初始化的目的是为了指定用户性别、装载说话人口音代码以及用户的母语脚本等,使语音识别内核适应用户。在结束使用 Spark 的语音识别服务之后,应用程序则应调用 SPKDestroy 函数以释放所有由 Spark 占用的内存和系统资源,防止内存泄漏。

程序实现如下:

```
procedure TForm1. InitSpark ;
.....

begin
m_IsInit := True;
if(m_IsMale = 0) then gender := GENDER_MALE
else gender := GENDER_FEMALE;
lReturn := SPKInitialize( INIT_SPEECH_VC Or DA_BEIJING Or gender );//初始化语音识别内核
lReturn := SPKSetNotification( Handle );
SPKNewGrammarFromFileEx( '. \MyData\Start. rtn', GRAMMAR_FLAG_REJECTION, m_Grammars[0] );
SPKSRAAddCommandGrammarItem( m_CommandGrammar, 0, ' KAI1 - SHI3', 1 );
end; //指定语法文件

procedure TForm1. FormDestroy( Sender:TObject );
begin
SPKSRStarRecognition(0);
SPKDestroy(); //释放内存
end;
```

### 2.2 调用 Windows API 函数,实现连接数字语音识别

程序通过接收并处理 Windows 消息来获取识别结果。首先调用 SPKSRStartRecognition 函数通知内核开始进行识别,一旦有信号通过音频设备进来,Spark 将把一条叫做 SPK\_MSG\_SPEECH\_START\_VB 的消息发送出去,告诉识别器应该使用的语法(文本识别或命令识别)。

当语音识别引擎完成识别之后,向程序发送一条消息,程序通过处理该消息来获取结果。按语法类型的不同,发出的消息分两种:SPK\_MSG\_COMMAND\_VB 消息和 SPK\_MSG\_RESULT\_VB 消息。对于前者,我们调用 Windows API 函数 sndPlaySound() 和 PlaySound() 代替命令识别,以便雷达情报口报员直接进行口报语音输入,完成连接数字语音识别。

处理 SPK\_MSG\_COMMAND\_VB 消息句柄程序实现如下:

```
procedure TForm1. OnmsgHasCommand( var Msg: TMessage );
.....

begin
theCommand := p_command( LongInt( Msg. WParam ) );
if( theCommand^. grammarID = m_CommandGrammar )
then
begin
m_CommandID := theCommand^. commandID;
if( m_CommandID > = 0 ) and ( m_CommandID < kNUMBER_OF_CMDS )
begin
m_Status := STATUS_PERMIT_FILLFORM;
```

```

    m_HilitField := true;
    SPKSRGetConfidence( theCommand^. grammarID, confidence );
    v := confidence;
    s := v;
    Label8. Caption := s;
    SetActiveItem;
    end;
end;
end;

```

在处理 SPK\_MSG\_RESULT\_VB 消息句柄时,程序要调用专业词汇检测函数 detect(),该函数用来处理专业化词汇的发音,它将识别结果从雷达情报的专业化词汇转换成标准化词汇,以便进行显示。

```

function detect(l:string):char;

begin
    if l = '洞' then r := '0'
    else if l = '邀' then r := '1'
    else if l = '两' then r := '2'
    else if l = '三' then r := '3'
    else if l = '四' then r := '4'
    else if l = '五' then r := '5'
    else if l = '六' then r := '6'
    else if l = '拐' then r := '7'
    else if l = '八' then r := '8'
    else if l = '勾' then r := '9';

    result := r;
end;

```

处理 SPK\_MSG\_RESULT\_VB 消息句柄程序如下:

```

procedure TForm1. OnmsgHasResult( var Msg: TMessage );
.....
begin
    theSentenceResult := presult( LongInt( Msg. WParam ) );
    StrCopy( Str, PChar( theSentenceResult^. text. strings[0] ) );
    for i := 0 to 1023 do
    begin
        if ( str[i] = #0 ) then break;
    end; //判断一次输入数字串的长度
    case i of
    4: begin
        s2 := str[0] + str[1];
        t1 := detect( s2 ); //detect() 是识别专业化词汇发音的函数
        s2 := str[2] + str[3]; //一个汉字是由两个音节构成的
        t2 := detect( s2 );
        t := t1 + t2;
    end; //识别字符串长度为 2
    6: begin
        s2 := str[0] + str[1];
        t1 := detect( s2 );

```

```

s2 := str[2] + str[3];
t2 := detect(s2);
s2 := str[4] + str[5];
t3 := detect(s2);
t := t1 + t2 + t3; //识别字符串长度为3
end;
.....//依此类推
Edit1.Text := Str;
m_HilitField := False;
m_Status := STATUS_PERMIT_COMMAND;
SPKDeleteResult(theSentenceResult);
sndPlaySound('f:\pihao.wav', SND_FILENAME or SND_SYNC);
PlaySound('f:\pihao.wav', hInstance, SND_RESOURCE or SND_SYNC); //替代命令识别
SetActiveItem;
end;

```

实验结果表明,基于 Spark SDK 语音识别内核,经过编写专业化词汇检测函数、修改语法规则以及调用 Windows API 函数可实现专业化词汇的连接、数字语音识别。目前,识别率可达到 98% 以上。

#### 参考文献:

- [1] Steve Teixeira Xavier Pacheco. Delphi 5 开发人员指南[M]. 任旭钧,王永生,冯泽波,等. 北京:机械工业出版社,2001.  
 [2] 易克初,田 斌,付 强. 语音信号处理[M]. 北京:国防工业出版社,2000.

(编辑:门向生)

## Connected Digit Speech Recognition Based on Spark SDK

HAN Dan, YAO Pei - yang, CHEN Ling

(The Telecommunication Engineering Institute, Air Force Engineering University, Xi'an, Shaanxi 710077, China)

**Abstract:** The speech recognition software Spark SDK, which can be developed again, is designed by Lernout and Hauspie company. And it is designed to provide a fast and easy way for the developers to add voice command recognition function into their application program. This paper introduces the method of realizing connected digit recognition by editing the function of professional words recognition and modifying the grammar rules of Spark SDK in radar intelligence oral - reporting system. The result shows that the recognition ratio is greatly improved.

**Key words:** Delphi; Spark SDK; connected digit speech recognition