International Journal of Computer Engineering & Technology (IJCET)
Volume 6, Issue 10, Oct 2015, pp. 66-80, Article ID: IJCET_06_10_008
Available online at
http://www.iaeme.com/IJCET/issues.asp?JType=IJCET&VType=6&IType=10
ISSN Print: 0976-6367 and ISSN Online: 0976–6375
© IAEME Publication

APPLICATION OF SPEECH WITH THEIR ANALYSIS ABOUT RECOGNITION

Manju
M. Tech student

Abhishek Bhatnagar
Asstt. Prof, IIET (Jind)

ABSTRACT

Speech recognition applications include voice user interfaces such as voice dialing, simple data entry, p reparation of structured documents, speech-to-text processing, and aircraft. The term voice recognition or speaker identification refers to identifying the speaker, rather than what they are saying. Recognizing the speaker can simplify the task of translating speech in systems that have been trained on a specific person's voice or it can be used to authenticate or verify the identity of a speaker as part of a security process.

Keywords: Single Word Error Rate, Command Success Rate, Recurrent Neural Networks(RNN's), Time Delay Neural Networks (TDNN's), *Deep Neural Networks, Matlab, Voice Box, Fourier Transformation*.

Cite this Article: Manju and Abhishek Bhatnagar. Application of Speech with Their Analysis about Recognition. *International Journal of Computer Engineering and Technology*, **6**(10), 2015, pp. 66-80.
http://www.iaeme.com/IJCET/issues.asp?JType=IJCET&VType=6&IType=10

1. INTRODUCTION

In computer science and electrical engineering, **speech recognition** (SR) is conversion of spoken words into text. It is also recognized as "automatic speech recognition" (ASR), "computer speech recognition", or just "speech to text" (STT).

Some SR systems use "speaker-independent speech recognition" while others use "instruction" where an individual speaker reads sections of text into SR scheme. These scheme examine person's specific voice and use it to modify recognition of that person's dialogue, resulting in more exact text. Systems that do not use instruction are called "speaker-independent" systems. Systems that use instruction are called "speaker-dependent" systems.

The term voice recognition or speaker identification refers to identifying speaker, rather than what they are saying.

Recognizing speaker can simplify task of translating speech in systems that have been trained on a particular person's voice or it can be used to authenticate or verify identity of a speaker as part of a security procedure.

2. INNOVATIONS

From technology perspective, speech identification has a long history with several influence of major innovations. In recent time, field has benefited from advances in deep knowledge and large data. advances are evidenced not only by course of academic papers published in field, but more importantly by world-wide industry implementation of a variety of deep learning methods in designing and deploying speech identification systems. These speech production players consist of Microsoft, Google, IBM, Baidu (China), Apple, Amazon, Nuance, IflyTek (China), many of which have publicized core knowledge in their speech identification systems being based on deep learning.

3. EXISTING IMPLEMENTATION

As early as 1932, Bell Labs researchers like Harvey Fletcher were investigating science of speech observation. In 1952 three Bell Labs researchers built a system for single-speaker digit detection. Their system worked by locating formants in power range of each speech. 1950s era technology was limited to single-speaker systems with vocabularies of around ten words.

Unfortunately, funding at Bell Labs dried up for several years when, in 1969, important John Pierce wrote an open letter that was critical of speech recognition research. Pierce's letter compared speech recognition to "schemes for turning water into gasoline, extracting gold from sea, curing cancer, or going to moon." Pierce defunded speech recognition research at Bell Labs.

Raj Reddy was first person to take on continuous speech recognition as a graduate student at Stanford University in late 1960s. Previous systems required users to make a gap after each word. Reddy's system was designed to issue spoken guidelines for game of chess. Also around this time Soviet researchers invented dynamic time warping algorithm and used it to create a recognizer capable of operating on a 200-word vocabulary. Achieving speaker independence was a major unsolved goal of researchers during this time period.

In 1971, DARPA funded five years of speech recognition research through its Speech Understanding Research program with determined end goals including a minimum vocabulary size of 1,000 words. BBN. IBM., Carnegie Mellon and Stanford Research Institute all participated in program. government funding revitalized speech recognition research that had been largely neglected in United States after John Pierce's letter. Despite fact that CMU's Harpy system met goals established at outset of program, many of predictions turned out to be nothing more than hype unsatisfactory DARPA administrators. This disappointment led to DARPA not ongoing support. Several innovations happened during this time, such as discovery of beam search for use in CMU's Harpy system. field also benefited from discovery of several algorithms in other fields such as linear predictive coding and cepstral investigation.

During late 1960's Leonard Baum developed mathematics of Markov chains at Institute for Defense investigation. At CMU, Raj Reddy's student James Baker and his wife Janet Baker began using Hidden Markov Model (HMM) for speech recognition. James Baker had well-read about HMMs from a summer job at Institute of Defense investigation during his undergraduate education. use of HMMs permitted researchers to combine different sources of information, such as acoustics, language, and syntax, in a integrated probabilistic model.

4. PERFORMANCE

The performance of speech recognition systems is usually evaluated in terms of correctness and speed. Accuracy is usually rated with word error rate (WER), whereas speed is measured with real time factor. Other measures of accuracy include Single Word Error Rate (SWER) and Command Success Rate (CSR). However, speech recognition (by a machine) is a very complex problem. Vocalizations vary in terms of accent,[70] pronunciation, articulation, roughness, nasality, pitch, volume, and speed. Speech is vague by a background noise and echoes, electrical characteristics. Accuracy of speech recognition vary with following:

- Vocabulary size and confusability
- Speaker dependence vs. independence
- Isolated, discontinuous, or continuous speech
- Task and language constraints
- Read vs. spontaneous speech
- Adverse conditions

5. VOICEBOX: SPEECH PROCESSING TOOLBOX FOR MATLAB

VOICEBOX is a speech processing toolbox consists of MATLAB routines that are maintained by and mostly written by Mike Brookes, Department of Electrical & Electronic Engineering, Imperial College, Exhibition Road, London SW7 2BT, UK. Several of routines require MATLAB V6.5 or above and require (normally slight) modification to work with earlier veresions.

The routines are available as a zip archive and are made available under terms of GNU Public License. routine VOICEBOX.M contains various installation-dependent parameters which may need to be altered before using toolbox. In particular it contains a number of default directory paths indicating where temporary files should be created, where speech data normally resides, etc. You can override these defaults by editing voicebox.m directly or, more conveniently, by setting an environment variable VOICEBOX to path of an initializing m-file.

6. PREPARATIONS

1. Download Voicebox.zip from API2011 web site and save it in your home directory. Unzip file in your home directory by opening a console window andtyping: ‘unzip Voicebox.zip’. This will make a new subdirectory Voicebox with all Voicebox MATLAB routines and data files. Note that, further information on Voicebox toolkit functionality can be found on following website:
2. You are ready to start matlab. In console window type: ‘matlab &’, this will start MATLAB 7.6.0 (R2008a). After starting MATLAB a window appears with several sub-windows. If not, select <Desktop><Desktop Layout><Default> to initiate default layout. In command window you can issue all Matlab commands.
3. Under <Help><Product Help> an extensive Help-environment can be found for references, tutorials, and examples on complete functionality of MATLAB.

4. In MATLAB select <File><Set Path>. In <Set Path><Dialog> that appears you can add folders and their subfolders to MATLAB search paths. MATLAB scripts and data in these folders can directly be used from MATLAB command line.

   Click on <Add with Subfolders> button and browse to just created Voicebox directory and select it. Click <close> and <yes> to apply your changes for following MATLAB sessions.
5. This step (5) only works if sound drivers are installed. You should use example wave files that are available in Voicebox/data directory or take new samples under Windows XP. Connect your microphone and headphones. Record a wave file in API2011 folder, and check if it contains data. In console window type:

   ```
   cd API2011
   rec -r 21000 -t wav sound.wav
   play sound.wav
   ```
6. Now you are ready to start using Voicebox functions. After Step 4 they are automatically recognized by MATLAB. You can see them in left top window, if you browse to directory under Voicebox that contains them. If you need help, go to Voicebox website mentioned above. Now have a quick look at different functions present in Voicebox toolkit.
7. Let's do a small example. We will load just recorded wave file, or another example wave file from Voicebox/data directory. Then convert it to frequency domain, filter out some frequencies and transform result back to time domain.

   Finally, we save file and listen to result and observe resulting data. In matlab console (rightmost window) type:

   ```
   [y,fs,wmode,fidx]=readwav('sound.wav','r',-1,0);
   ```
8. Variable <y> now contains stereo sample data, y(:,1) contains left channel. Type following to put left channel in <left> and view contents of left channel:

   ```
   left=y(:,1);
   ```
9. Now let’s have a look at sound data and frequency spectrum (we have to tell function that we used a 16000 sampling rate). Type:

   ```
   plot(left);
   figure;
   spgrambw(left,16000);
   ```
10. Observe data. Let's slice up left channel in 6 equal parts and show individual power spectra using rfft (fast fourrier transform) on all 6 parts. Type:

    ```
    frames=enframe(left, uint16(length(left)/6));
    frames=transpose(frames)
    ;
    fftdata=rfft(frames);
    fftdata=fftdata.*conj(fftdata)
    ;
    plot(fftdata);
    ```
11. 11.This results in a combined plot of power spectra of six consecutive parts of sound file. You should be able to see that power of different frequencies is different in each of parts. These kinds of differences thus characterize sound data in terms of average frequency amplitudes. If you cannot see these differences,

    plot different spectra in different windows, by typing:

    ```
    plot(fftdata(:,1))
    figure
    ```

```
plot(fftdata(:,2))
figure
etc...
```

12. It should be clear that earlier drawn figure of power spectrum (step 8) actuallyis a finer grained (and better visualized) version of different plots you have just created

7. IMPLEMENTATION OF SPEECH RECOGNIZATION

Code to recognize english voice:

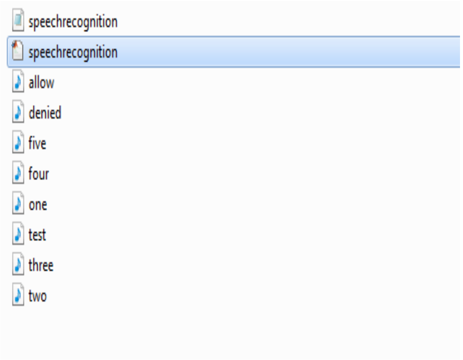
```
function speechrecognition(filename)
%Speech Recognition Using Correlation Method
%Write Following Command On Command Window
%speechrecognition('test.wav')
voice=wavread(filename);
x=voice;
x=x';
x=x(1,:);
x=x';
y1=wavread('one.wav');
y1=y1';
y1=y1(1,:);
y1=y1';
z1=xcorr(x,y1);
m1=max(z1);
l1=length(z1);
t1=-((l1-1)/2):1:((l1-1)/2);
t1=t1';
%subplot(3,2,1);
plot(t1,z1);
y2=wavread('two.wav');
y2=y2';
y2=y2(1,:);
y2=y2';
z2=xcorr(x,y2);
m2=max(z2);
l2=length(z2);
t2=-((l2-1)/2):1:((l2-1)/2);
```

```
t2=t2';
%subplot(3,2,2);
figure
plot(t2,z2);
y3=wavread('three.wav');
y3=y3';
y3=y3(1,:);
y3=y3';
z3=xcorr(x,y3);
m3=max(z3);
l3=length(z3);
t3=-((l3-1)/2):1:((l3-1)/2);
t3=t3';
%subplot(3,2,3);
figure
plot(t3,z3);
y4=wavread('four.wav');
y4=y4';
y4=y4(1,:);
y4=y4';
z4=xcorr(x,y4);
m4=max(z4);
l4=length(z4);
t4=-((l4-1)/2):1:((l4-1)/2);
t4=t4';
%subplot(3,2,4);
figure
plot(t4,z4);
y5=wavread('five.wav');
y5=y5';
y5=y5(1,:);
y5=y5';
z5=xcorr(x,y5);
m5=max(z5);
l5=length(z5);
```

```
t5=-((l5-1)/2):1:((l5-1)/2);
t5=t5';
%subplot(3,2,5);
figure
plot(t5,z5);
m6=300;
a=[m1 m2 m3 m4 m5 m6];
m=max(a);
h=wavread('allow.wav');
if m<=m1
   soundsc(wavread('one.wav'),50000)
      soundsc(h,50000)
elseif m<=m2
   soundsc(wavread('two.wav'),50000)
      soundsc(h,50000)
elseif m<=m3
   soundsc(wavread('three.wav'),50000)
      soundsc(h,50000)
elseif m<=m4
   soundsc(wavread('four.wav'),50000)
      soundsc(h,50000)
elseif m<m5
   soundsc(wavread('five.wav'),50000)
      soundsc(h,50000)
else soundsc(wavread('denied.wav'),50000)

end
```

List of files



speechrecognition
speechrecognition
allow
denied
five
four
one
test
three
two

Figure A [List of english files]

RUNNING CODE

```
>> speechrecognition ('one.wav')
```

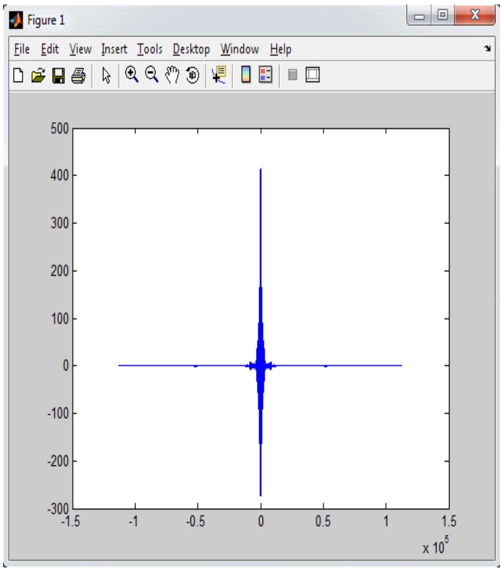


Figure 1[one.wav file]

```
>> speechrecognition ('two.wav')
```

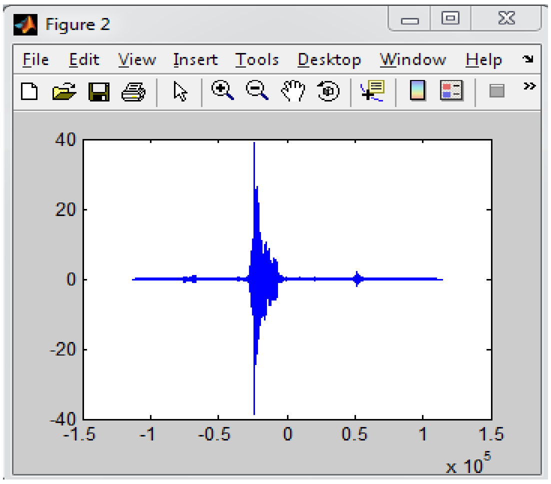


Figure 2[two.wav file]

>> speechrecognition (‘three.wav’)

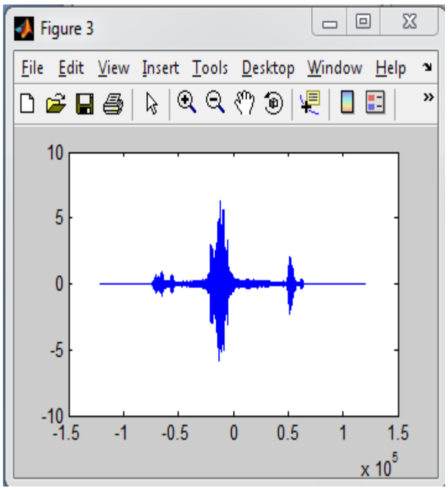


Figure 3 [three.wav file]

>> speechrecognition (‘four.wav’)

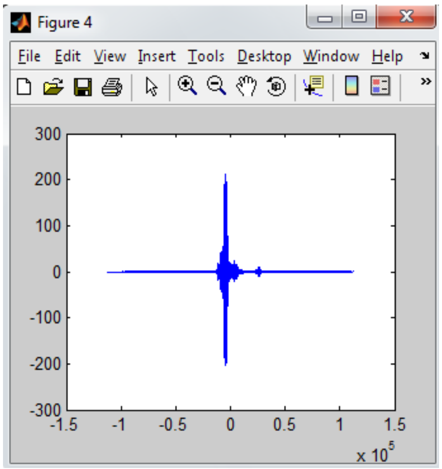


Figure 4 [four.wav file]

>> speechrecognition (‘five.wav’)

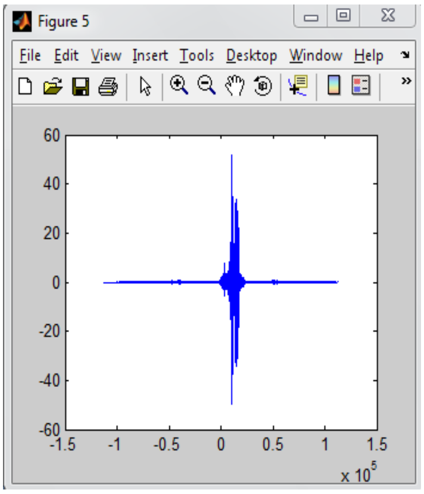


Figure 5[five.wav file]

Code to recognize hindi voice

```
function speechrecognition(filename)
%Speech Recognition Using Correlation Method
%Write Following Command On Command Window
%speechrecognition('test.wav')
voice=wavread(filename);
x=voice;
x=x';
x=x(1,:);
x=x';
y1=wavread('ek.wav');
y1=y1';
y1=y1(1,:);
y1=y1';
z1=xcorr(x,y1);
m1=max(z1);
l1=length(z1);
t1=-((l1-1)/2):1:((l1-1)/2);
t1=t1';
%subplot(3,2,1);
plot(t1,z1);
y2=wavread('do.wav');
y2=y2';
y2=y2(1,:);
y2=y2';
z2=xcorr(x,y2);
m2=max(z2);
l2=length(z2);
t2=-((l2-1)/2):1:((l2-1)/2);
t2=t2';
%subplot(3,2,2);
figure
plot(t2,z2);
y3=wavread('teen.wav');
```

```
y3=y3';
y3=y3(1,:);
y3=y3';
z3=xcorr(x,y3);
m3=max(z3);
l3=length(z3);
t3=-((l3-1)/2):1:((l3-1)/2);
t3=t3';
%subplot(3,2,3);
figure
plot(t3,z3);
y4=wavread('char.wav');
y4=y4';
y4=y4(1,:);
y4=y4';
z4=xcorr(x,y4);
m4=max(z4);
l4=length(z4);
t4=-((l4-1)/2):1:((l4-1)/2);
t4=t4';
%subplot(3,2,4);
figure
plot(t4,z4);

%y5=wavread('five.wav');
%y5=y5';
%y5=y5(1,:);
%y5=y5';
%z5=xcorr(x,y5);
%m5=max(z5);
%l5=length(z5);
%t5=-((l5-1)/2):1:((l5-1)/2);
%t5=t5';

%subplot(3,2,5);
```

```
%figure
%plot(t5,z5);
%m6=300;
a=[m1 m2 m3 m4];
m=max(a);
h=wavread('allow.wav');
if m<=m1
   soundsc(wavread('ek.wav'),50000)
      soundsc(h,50000)
elseif m<=m2
   soundsc(wavread('do.wav'),50000)
      soundsc(h,50000)
elseif m<=m3
   soundsc(wavread('teen.wav'),50000)
      soundsc(h,50000)
elseif m<=m4
   soundsc(wavread('char.wav'),50000)
      soundsc(h,50000)
else soundsc(wavread('denied.wav'),50000)

end
```

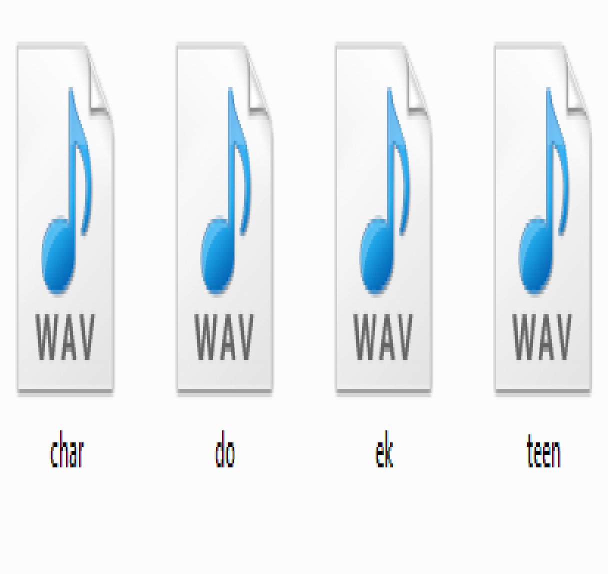


Figure B List of hindi files]

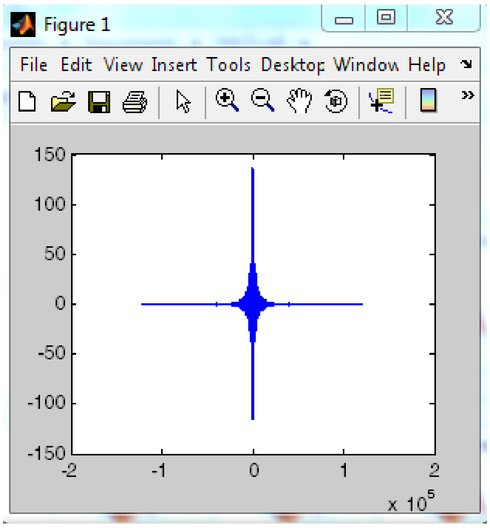


Figure 6 [ek.wav file]

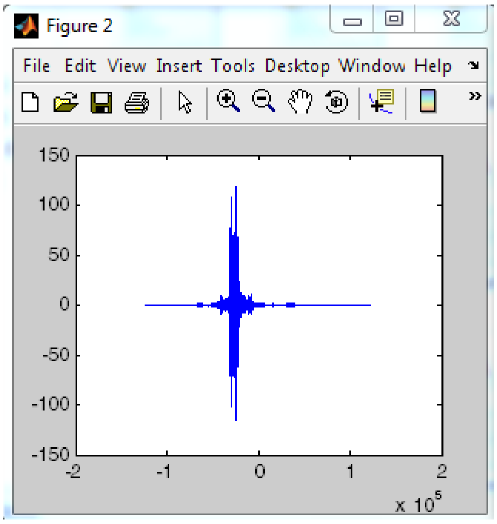


Figure 7 [do.wav file]

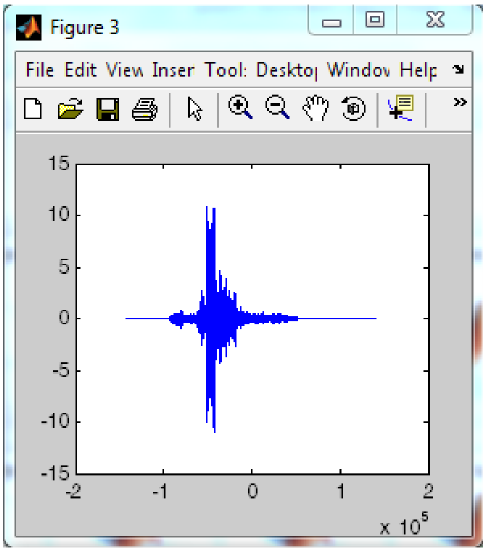


Figure 8 [teen.wav file]

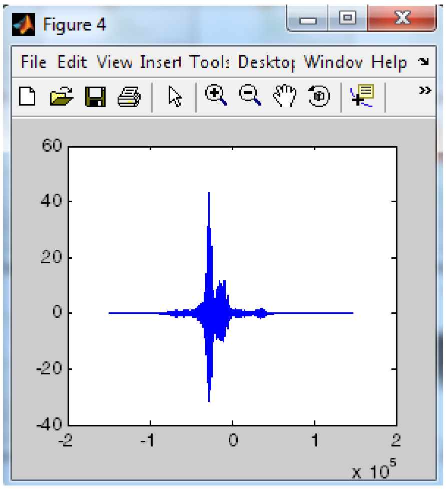


Figure 9 [char.wav file]

8. APPLICATIONS

In-car systems

Typically a manual control input, for example by means of a finger control on steering-wheel, enables speech recognition system and this is signaled to driver by an audio prompt.

Medical documentation

In health care sector, speech recognition can be implemented in front-end or back-end of medical documentation process.

Front-end speech recognition is where provider dictates into a speech-recognition engine, recognized words are displayed as they are spoken, and dictator is responsible for editing and signing off on document.

Back-end or deferred speech recognition is where provider dictates into a digital dictation system, voice is routed through a speech-recognition machine.

High-performance fighter aircraft

Substantial efforts have been committed in last decade to test and evaluation of speech recognition in fighter aircraft. Of particular note is U.S. program in speech recognition for Advanced Fighter Technology Integration (AFTI)/F-16 aircraft (F-16 VISTA), and a program in France installing speech recognition systems on Mirage aircraft, and also programs in UK dealing with a variety of aircraft platforms.

9. CONCLUSION

Dramatic advances have recently been made in speech recognition technology. Large-vocabulary talker-independent recognizers provide error rates that are less than 10% for read sentences recorded in a quiet environment. Machine performance, how-ever, deteriorates dramatically under degraded conditions. For example, error rates increase to roughly 40% for spontaneous speech and to 23% with channel variability and noise. Human error rates remain below 5% in quiet and under similar degraded conditions. Comparisons using many speech corpora demonstrate that human word error rates are often more than an order of magnitude lower than those of current recognizers in both quiet and degraded environments. In general, the superiority of human performance increases in noise, and for more difficult speech material such as spontaneous speech. Al-though current speech recognition technology is well suited to many practical commercial applications, these results suggest that there is much room for improvement.

10. FUTURE SCOPE

Comparisons between human and machine error rates suggest the need for more fundamental research to improve machine recognition performance. This research could focus on four areas where past studies demonstrate the most dramatic differences between human and machine performance. First, results obtained with limited context suggest that human listeners perform more accurate low-level acoustic-phonetic modeling than machines. We can accurately recognize isolated digit sequences and spoken letters, we can recognize short segments extracted from spontaneous conversations, and we can accurately recognize words in nonsense sentences that provide little contextual information. These results suggest that one important direction for future research with machine recognizers is to improve low-level acoustic phonetic analysis. Second, human recognition results obtained with channel variability and noise demonstrate that we can easily recognize speech with normally occurring degradations. Past studies have also demonstrated that we can understand speech with no training when highly unnatural distortions are applied.

REFERENCES

[1] "Speaker Independent Connected Speech Recognition- Fifth Generation Computer Corporation". Fifthgen.com. Retrieved 2013-06-15.

[2] "British English definition of voice recognition". Macmillan Publishers Limited. Retrieved February 21, 2012.

[3] "voice recognition, definition of". WebFinance, Inc. Retrieved February 21, 2012.

[4] "The Mailbag LG #114". Linuxgazette.net. Retrieved 2013-06-15.

[5] "Speaker Identification (WhisperID)". *Microsoft Research*. Microsoft. Retrieved 21 February 2014. When you speak to someone, they don't just recognize what you say: they recognize who you are. WhisperID will let computers do that, too, figuring out who you are by way you sound.

[6] Huffman, Larry. "Stokowski, Harvey Fletcher, and Bell Labs Experimental Recordings". www.stokowski.org. Retrieved February 17, 2014.

[7] Juang, B. H.; Rabiner, Lawrence R. "Automatic speech recognition–a brief history of technology development" (PDF). p. 6. Retrieved 17 January 2015.

[8] Pierce, John (1969). "Whither Speech Recognition". *Journal of Acoustical Society of America*. doi:10.1121/1.1911801.

[9] Benesty, Jacob; Sondhi, M. M.; Huang, Yiteng (2008). *Springer Handbook of Speech Processing*. Springer Science & Business Media. ISBN 3540491252.

[10] BLECHMAN, R.O.; BLECHMAN, NICHOLAS (June 23, 2008). "Hello, Hal". New Yorker. Retrieved 17 January 2015.

[11] http://www.sarasinstitute.org/Audio/JimBaker (2006).mp3. Retrieved 23 March 2015. Missing or empty |title= (help)

[12] Huang, Xuedong; Baker, James; Reddy, Raj. "A Historical Perspective of Speech Recognition". Communications of ACM. Retrieved 20 January 2015.

[13] Juang, B. H.; Rabiner, Lawrence R. "Automatic speech recognition–a brief history of technology development" (PDF). p. 10. Retrieved 17 January 2015.

[14] "History of Speech Recognition". Retrieved 17 January 2015.

[15] www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html.