A Voice Recognizing Elevator System

Chapter · December 2016

DOI: 10.1007/978-81-322-2671-0_17

3 authors, including:

Shahina A.
Sri Sivasubramaniya Nadar College of Engineering
47 PUBLICATIONS **333** CITATIONS

SEE PROFILE

A Voice Recognizing Elevator System

Meenatchi.D
Department of Information Technology
SSN College of Engineering
Affiliated to Anna University
Chennai, India.
meenatchidau@gmail.com

Aishwarya.R
Department of Information Technology
SSN College of Engineering
Affiliated to Anna University
Chennai, India.
r.aishwarya5294@gmail.com

Shahina .A
Department of Information Technology
SSN College of Engineering
Affiliated to Anna University
Chennai, India.
shahinaa@ssn.edu.in

Abstract: *The Paper proposes a voice-controlled, simulated elevator system for the benefit of differently abled persons, such as those who are visually impaired or are paraplegics. The proposed approach for an eight floor elevator system uses a speaker independent automatic speech recognition system to recognize spoken words which includes the floor numbers, directions and door commands. Sphinx4 is used for this purpose. To handle emergencies, the recognition of ten digit cellular phone numbers is incorporated with a text to speech system that gives voice feedback for verification before the telephone call is placed. The mean recognition accuracy for floor numbers is 97%, while it is 90% for directions and door operations.*

keywords: elevator system, automatic speech recognition, Sphinx4, Text To Speech conversion.

I. INTRODUCTION

In this rapid world of technology where voice begins its era of domination to replace the touch screens from smart phones to huge computer systems, bringing voice in day to day affairs becomes significant. Elevators being one such system used in daily life serves this purpose of making future generations hands free which also becomes a boon for the disabled.

The basic working principle of elevator is based on the elevator algorithm, where an elevator can decide to stop based on two conditions. The first one being the direction and second one based on the current floor and destination floor. The elevator is generally made up of rotors, cables, pulleys based on traction, climbing or hydraulic model. To serve laboratory purposes it can also be designed by connecting the elevator system to a desktop or microprocessor to accept input voice.

Voice control option is attractive for several reasons. It is potentially appropriate for a large number of elevator users since the system can be used by any individual capable of consistent and distinguishable vocalization [7]. Voice control also reduces physical requirements. However, the recognition accuracy of Automatic Speech Recognition (ASR) system is a constraint in the deployment of many voice controlled system in real world application.

In this paper, a voice controlled elevator system is proposed where the input commands to stimulate the movement of the elevator system are kept convenient for the users. The commands include voice input for the floor operations, directions, elevator car's door operation and a special option to place a call of speaker's choice in case of any unexpected event that requires immediate action.

The paper is organized as follows: Section II mentions briefly a review on some of the related earlier works. Section III describes how Sphinx4 has been adapted for this project and also explains various features of speech recognition. Section IV explains the experimental setup for the proposed model. Section V discusses the experimental results that are obtained on performing the tests under laboratory conditions.

II. RELATED WORK

Speech recognition had been effectively contemplated since 1950s, however late improvements in Personal Computer (PC) and telecommunication technology have enhanced speech recognition abilities [1]. In a useful sense, speech recognition has tackled issues, enhanced benefits and bought a greater revolution in current scenario [1]. Voice control can replace the function of a push-button efficiently [2]. Speech recognition is a very complex issue. It includes numerous calculations which oblige high computational necessities [2]. The mixture of utilizations of programmed speech recognition systems, human PC interfaces, telephony, or robotics has driven the exploration of a vast academic group over a decades ago [3]. Automatic speech recognition is used in various areas ranging from medical transcription to game control, from call center dialogue systems to data recovery[4].

This automatic speech recognition process is used in most of the voice controlled systems. The voice controlled wheelchair for the physically challenged was proposed in 2002 [7]. This paper had described an experiment that compared the performance of abled and disabled people using voice control to operate a power wheelchair both with and without navigation assistance, where the navigations were assisted by the sensors to identify and avoid obstacles in wheelchair's path [7].

The intelligent lift control model that uses voice recognition was proposed in 2003 [8]. This proposed model had been controlled by voice and sensor panel [8]. The modification of the well-known DTW (Dynamic Time Warping) algorithm was used [8]. The set of voice commands for the model consisted of eight Lithuanian words [8]. The model was specifically designed for domestic use (smart houses).

In contrast to the intelligent lift control model [8], this paper proposes a simulated model that utilizes speech recognition to implement an elevator system that could be of help to visually or physically challenged person where the system generated voice gives assistance. Also it brings in the

concept of immediate safety measures for emergency. Older concepts of button and switches are rapidly being replaced with voice.

III. ADAPTING SPHINX4 INTO THIS PROJECT

A. Overview of Sphinx4

Sphinx4 is a Java speech recognition library where Speech recognition used is an Open Source recognition platform from CMU. It gives a quick API to change the speech recordings into text with the help of "CMUSphinx" acoustic models. It can be utilized on servers and as a part of desktop applications and it is highly portable. Beside speech recognition Sphinx4 serves to distinguish speakers, adapt models, allows highly flexible user interfacing and more. Sphinx4 supports US English and many other dialects.

B. Using Sphinx4 in This Project

The recognition platform of Sphinx4 helps to add library files into dependencies of the project and there are few high level recognition interfaces such as *LiveSpeechRecognizer*, *StreamSpeechRecognizer*, *SpeechAligner* and *SpeechResult*. These interfaces along with the *Acoustic model*, *Dictionary*, *Grammar/Language model* and *Source of speech* are executed for the task of speech recognition.

The recognition interfaces are explained as follows: *LiveSpeechRecognizer* uses the microphone as speech source where the start and stop recording functions performs the recognition [9]. *StreamSpeechRecognizer* utilizes InputStream as the speech source where the information can be passed from the file or network socket [9]. *SpeechAligner* aligns text with audio speech [9]. *SpeechResult* gives access to different parts of the recognition result such as recognized utterance, list of words with time stamps, recognition lattice etcetera [9]. Each of these recognition interfaces are supplied with required and discretionary attributes by the configuration manager.

The four necessary attributes of Sphinx4 are described in detail. *Acoustic model* is used as part of automatic speech recognition to represent association between an audio signal and the phonemes. There are context independent models that contain properties (feature vectors for each phone) and context dependent ones (created from phones with context) [10]. A *phonetic dictionary* has a mapping from words to phones, which is not very effective. For instance, just two to three pronunciation variants are noted in it, yet it is sufficient most of the time.The dictionary is not the only variant of mapper from words to phones. It might be possible with some complex function such as machine learning calculation [10].

A *language model* is used to limit word search and to represent measurable properties of speech [10]. It characterizes which word could take after already perceived words and serves to limit the coordinating process by stripping words that are not likely. To achieve a better recognition rate, a language model must be effective in search space restriction [10]. This implies that it should be very good at anticipating the next word. A language model normally allows the vocabulary to contain the words specific to that dialect. That is an issue for name recognition. To manage this, a language model can contain smaller chunks like sub-words or even phones. The search space restriction for this situation is generally worse and corresponding recognition accuracies are lower when compared with a word-based language model. *Source of speech* in Sphinx4 generally is open source where any speaker can pronounce any word that has been already defined in the phonetic dictionary belonging to any one of the specified language.

C. Using Text To Speech Conversion in This Project

Text-To-Speech (TTS) synthesizer is a computer-based system that reads any text aloud, whether it was introduced in the PC by an administrator or checked and submitted to an Optical Character Recognition (OCR) system [11]. It is a speech synthesis application that is used to make a spoken sound version of the text in a computer report, such as a help document or a Web page [12]. It can empower the reading of computer display information for the visually challenged person or may just be utilized to enlarge the perusing of an instant message.

IV. ELEVATOR EXECUTION

A. Databases

The experiment (refer section v) is performed under laboratory conditions with the use of an Audio Technica pro37 condenser microphone. The experimental conditions are impervious to superfluous noise. One speaker at a time is allowed to give any of the following input commands.

TABLE I. LIST OF VOICE COMMANDS

	Commands	Description
Floor numbers	*One, two, three, four, five, six, seven, eight.*	The simulated elevator moves to the corresponding floor number.
Direction	*Up, down.*	The simulated elevator moves along the corresponding directions.
Door operations	*Open, close*	The opening and closing of the elevator car's door is controlled.
Cellular number	Sequence of ten digits*	Call for help is placed to this number

*also includes *double, triple, nought (0).*

The system generated voice greets the user, guides the user about the current floor and also instructs them to feed in the input voice commands.

B. Block Diagram

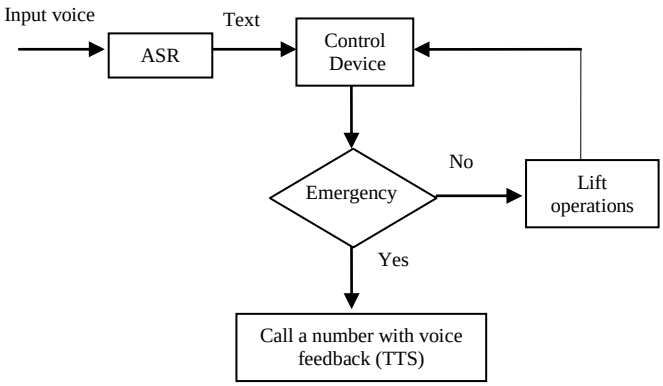

Figure 1: Block diagram of voice controlled lift

The input voice to the Automatic Speech Recognition (ASR) system deduces a command for the control device. This device controls entire elevator operations including its speed and traction. It also handles any emergency situation. In the event of unexpected situations, to avoid panicking, calling methods has been improvised where the user can verify the number using a TTS system. Any cellular number of speaker's choice can be called, Otherwise the normal lift operations are resumed.

The different parts of the block diagram are explained in the following sub sections C (ASR), D (Emergency Situations) and E (Control Device).

C. Automatic Speech Recognition

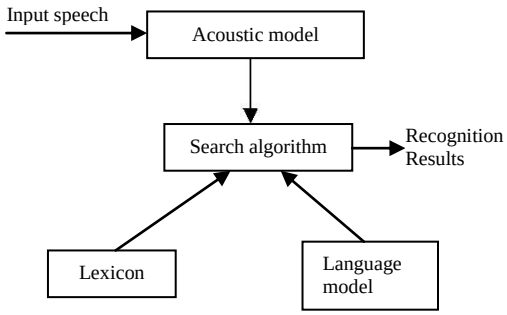

Figure 2: Block diagram of ASR

The acoustic model detects the relationship between the input speech signal and phonetic units in the language. The speech output utilizes the search algorithm to provide the recognition results. The search takes place by mapping the words to the phones in the lexicon (dictionary), where the mapped words undergo search space restriction of the language model to filter out the results.

The automatic recognition of the input speech through microphone and its related functions used in the elevator simulation are described in figure 3.

```
public class voiceRecognizer implements Runnable {
private Microphone microphone;
private Recognizer recognizer;
private List voiceListeners = new ArrayList();
publicvoiceRecognizer() throws IOException {
try
{
     URL
url=this.getClass().getResource("hellodigits.config.xml");
if (url == null)
{
throw new IOException("Can't find hellodigits.config.xml");
}
ConfigurationManager cm =
newConfigurationManager(url);
recognizer = (Recognizer) cm.lookup("recognizer");
microphone = (Microphone) cm.lookup("microphone");
}
catch (PropertyException e) {
            e.printStackTrace();
throw new IOException("Problem configuring
voiceRecognizer " + e);}
catch (InstantiationException e) {
throw new IOException("Problem creating voiceRecognizer
" + e);}
   }

public void microphoneOn() {
new Thread(this).start();
   }

public void microphoneOff() {
microphone.stopRecording();
   }

public void startup() throws IOException {
recognizer.allocate();
   }

public void shutdown() {
microphoneOff();
if (recognizer.getState() == RecognizerState.ALLOCATED)
       {
recognizer.deallocate();
       }
   }
public void run() {
microphone.clear();
microphone.startRecording();
       Result result = recognizer.recognize();
microphone.stopRecording();
if (result != null) {
          String resultText =
result.getBestFinalResultNoFiller();
else {
fireListeners(null);
          }
}
}
```

Figure 3: code snippet of automatic speech recognition

The inbuilt functions *startRecording()* and *stopRecording()* of the microphone class is used for recording speech. The recognizer class uses the function *microphoneOn()* to initiate and microphoneOff() to complete the recording and the received input signals are processed to remove the fillers. The extracted results from result class are sent to the control device.

D. Control Device

The general movement of the elevator between different floors including the acceleration and time of travel is determined by the control device. The directions sent to the control device are from ASR or the feedback from normal lift operations. The operation done here includes the aspect of safety, performance and perfect coordination.

E. Emergency situations

The effective and unique handling of emergency situations is a significant part of the proposed model. The ten digit phone number of speaker's choice is accepted. The TTS conversion system replays the number that is ratified by the user. The call is placed to that number which can bring immediate help.

V. EXPERIMENTAL EVALUATION

The simulated model with eight floors, elevator car and request pool is shown in the implementation. The elevator car movement is controlled either by the user at a specific floor who can voice the direction or the user inside the elevator car who can voice the destination floor number. The door opens as the elevator car reaches the respective floor and allows people to enter the lift (figure 4).

The lift operations proceed unless emergency situations arise, when the user is permitted to call a cellular number. This number is echoed back by the system for the user to verify (figure 5).

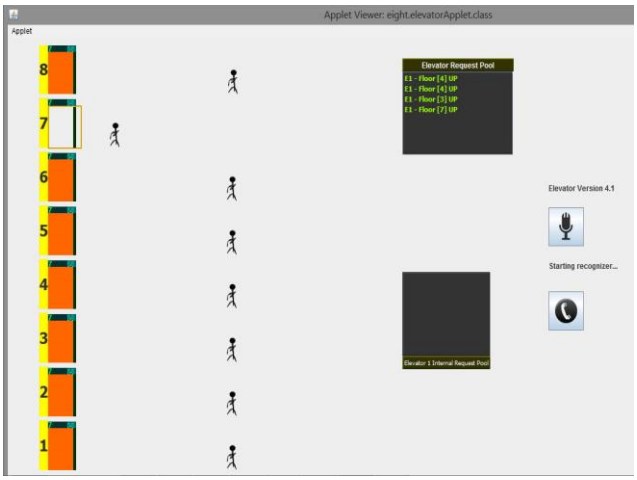

Figure 4: Implementation of the voice controlled lift (normal lift operations)

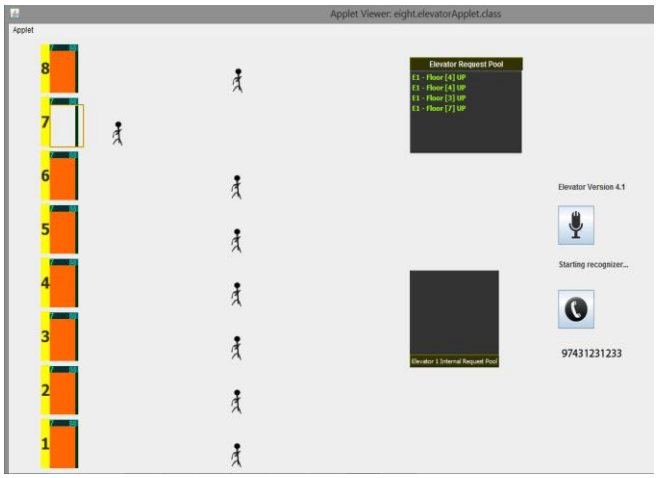

Figure 5: Implementation of the voice controlled lift (emergency situations)

Five different voices have been used for giving the input commands and the recognition accuracy varies for each speaker. The recognition of both digits and words varies according to the pronunciation. The results are discussed as follows:

TABLE II. LIST OF VOICE RESPONSES FOR EMERGENCY SITUATIONS

Case	Phone Number	Voice response
A	805603480	Please say a ten digit phone number
B	9789941254	Nine seven eight nine nine four one two five four
C	9441424821	Nine four four one four **eight** four eight two one
D	74166947211	Please say a ten digit phone number

Table II gives the list of possibilities of the system response for different sequence of telephone numbers. When a number with less than ten digits is uttered the system responds by asking to repeat the ten digit number (Case A). When an exact ten digit phone number is uttered the system replays the number (Case B). The misinterpretation of the number is also realized by means of voice response (Case C). When a number with more than ten digits is uttered the system responds by asking to repeat the ten digit number (Case D).

TABLE III. CONFUSION MATRIX FOR FLOOR NUMBER RECOGNITION

Digits	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	9	0	0	0	1	0	0	0
2	2	6	0	0	0	0	0	2
3	0	0	7	0	0	0	0	3
4	0	0	0	8	0	1	0	1
5	1	0	0	0	8	0	0	0
6	0	0	0	2	0	8	0	0
7	1	0	0	0	0	0	9	0
8	0	0	0	0	0	0	0	10

Table III shows the confusion matrix for the recognition of the floor numbers. For a scale of one to ten, the number *eight* is recognized all the ten times as *eight*, whereas the number *two* and *three* are recognized as themselves for six and seven times respectively.

TABLE IV. CONFUSION MATRIX FOR LIFT OPERATION RECOGNITION

Lift Operations	Open	Close	Up	Down
Open	8	1	1	0
Close	0	7	3	0
Up	0	0	10	0
Down	0	0	1	9

Table IV shows the confusion matrix for the recognition of the door operations and directions. For a scale of one to ten, the word *up* is recognized all the ten times as *up*, whereas the word *close* is recognized as *close* for seven times.

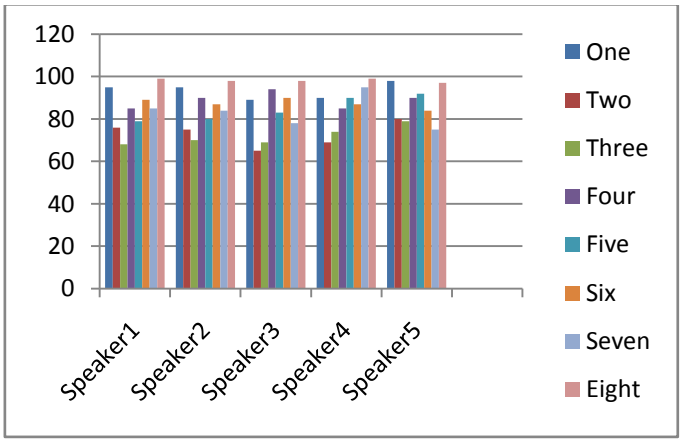

Figure 6: Recognition accuracy in percentage for floor numbers *(digits one – eight)* depending upon the speaker

The graph in figure 6 shows the percentage accuracy of the digits for each of the five speakers. The numbers *one* and *eight* have greater accuracy (95.5%). The numbers *four*, *five*, *six* and *seven* have better accuracy (86.2%) compared to the numbers *two* and *three* (72.5%) (Table V).

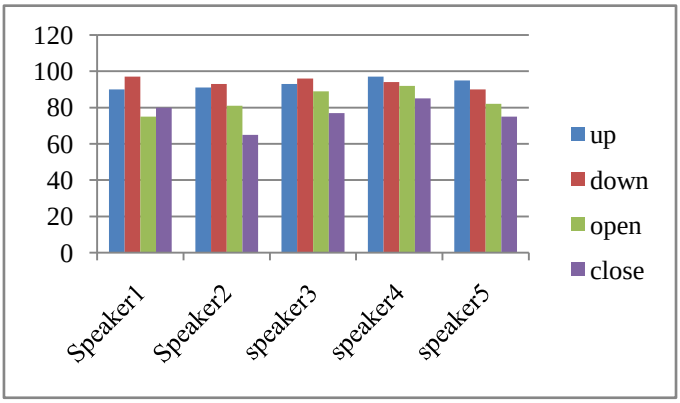

Figure 7: Recognition accuracy in percentage for directions *(up, down)* and door operations *(open, close)* depending upon the speaker

The graph in figure 7 shows the percentage accuracy of the words for each of the five speakers. The words *up* and *down* have greater accuracy (93.6%). The word *open* has better accuracy (83.8%) compared to the word *close* (76.5%) (Table VI).

TABLE V. PERCENTAGE ACCURACY FOR FLOOR NUMBERS

Commands	**% Accuracy**
One	93.4
Two	73
Three	72
Four	88.8
Five	84.8
Six	87.4
Seven	83.4
Eight	98.2

TABLE VI. PERCENTAGE ACCURACY FOR DIRECTIONS AND DOOR OPERATIONS

Commands	**% Accuracy**
Up	93.2
Down	94
Open	83.8
Close	76.4

VI. CONCLUSION

This paper explains how voice control can become a boon in the future in everyday life by means of an elevator simulation. The proposed work shows the feasibility of developing an elevator system based on voice control. It also incorporates a voice feedback system in case of placing emergency calls, which helps the user to verify the correctness

of the number. In future, the size of the experiment can be improvised to make this model a real time system. Also the recognition accuracy can be improved by making the speech recognition system speaker dependent and by including the aspect of robustness to noise.

VII. REFERENCES

[1] K. Michael, M. K. Anna, T. Justin, andY. Tony, "Development of a Voice Recognition Program", May 2002.

[2] P. Martin,P. Dušan, and F. Peter, "Speech control for car using the TMS320C6701 DSP", 10th International Scientific Conference, Bratislava, pp. 97-100, 2000.

[3] A. V. Nefian, L. Luhong, P. Xiaobo, X. Liu, C. Mao, and K. Murphy, "A coupled HMM for audio-visual speech recognition", IEEE Proceedings (ICASSP '02) on Acoustics, Speech, and Signal Processing, Volume 2, pp. 2013-2016, 2002.

[4] H. Thomas, E. H. Asmaa, N. W. Stuart, andW. Vincent, "Automatic speech recognition for scientific purposes – webASR".

[5] T. Q. Muhammad and A. A. Syed,"Voice controlled wheelchair Using DSK TMS320C6711", 2009 International conference on signal acquisition and processing, pp. 217-220.

[6] 2012 International Conference on Computer Science and Electronics Engineering, Youhao YuDepartment of Electronics and Information Engineering ,Putian University,Putian, Fujian, 351100,China "Research on Speech Recognition Technology and Its Application".

[7] C. S. Richard and P. L. Simon, "Voice Control of a Powered Wheelchair", IEEE transactions on neural systems and rehabilitation engineering, vol 10, no 2, pp. 122-125, June 2002.

[8] P.Cernys, V.Kubilius, V.Macerauskas, and K.Ratkevicius, "Intelligent Control of the Lift Model ",IEEE lntemetional Workshop on Intelligent Data Acquisition and Advanced Computing System: Technology and Applications

[9] Sphinx4 dependencies URL:http://cmusphinx.sourceforge.net/wiki/tutorialsphinx4

[10] Sphinx4 attributes URL:http://cmusphinx.sourceforge.net/wiki/tutorialconcepts

[11] Text To Speech(TTS) URL:http://searchmobilecomputing.techtarget.com/definition/text-to-speech

[12] Text To Speech(TTS) URL:http://freetts.sourceforge.net/docs/index.php#what_is_freetts