Shannon Donnelly

IT 103-007

February 29, 2012

Automatic Speech Recognition

"By placing this statement on my webpage, I certify that I have read and understand

the GMU Honor Code on http://academicintegrity.gmu.edu/honorcode/ . I am fully

aware of the following sections of the Honor Code: Extent of the Honor Code,

Responsibility of the Student and Penalty. In addition, I have received permission

from the copyright holder for any copyrighted material that is displayed on my site.

This includes quoting extensive amounts of text, any material copied directly from a

web page and graphics/pictures that are copyrighted. This project or subject material

has not been used in another class by me or any other student. Finally, I certify that

this site is not for commercial purposes, which is a violation of the George Mason

Responsible Use of Computing (RUC) Policy posted on

http://universitypolicy.gmu.edu/1301gen.html web site."

Shannon Donnelly

Introduction

"Speech recognition is the ability of a machine or program to identify words and phrases in spoken language and convert them to a machine-readable format" ("Speech recognition ," 1997). In today's world, automatic speech recognition has been put on a side burner to the point that people do not notice it how great technology is or how often we use it. Its creation was only a few decades ago and it has come this far, there is still so much more room for advancement in years to come. Today we use this software for many things such as call centers when we call companies and corporations that receive a lot of incoming business or with the blind or authors when they speak into a microphone and their words are made into a document. As we take a deeper look into the creation of automatic speech recognition we will address its possible disadvantages and advantages that will give us a better understanding of this new technology.

Background Information

As previously stated, automatic speech recognition was only launched decades ago, to be specific, in the 1950s. "Bell laboratories designed in 1952 the "Aubrey" system, which recognized digits spoken by a single voice" (Pinola, 2011). Notice that only a single voice could be used when first tested on the recognition software because the technology at that time was not advanced enough to understand more than one pitch or even different sexes. Numbers are also the most general form of communication when speaking verbally because they only represent quantities rather then expressions and feelings. Numbers can be communicated in many other forms even with quick universal hand signals of holding up your fingers. Not until 1962, did IBM create its "Shoebox machine, which could understand sixteen words spoken in English" (Pinola, 2011).

Transitioning from numbers to words was a huge advancement because those sixteen words were the first step in the human language being basic commands, which in the field of an answering service is a huge help because it can decide whom the caller specifically needs to talk to and narrow down what they need help with. A major step was in 1976 when the "U.S. Department of Defense funded the DoD's DARPA Speech Understanding Research (SUR) program…it was responsible for Carnegie Mellon's "Harpy" speech understanding system. Harpy could understand 1011, words, approximately the vocabulary of an average of a three-year old" (Pinola, 2011). At this point in time, a development on interpretation different voices was made as well as understanding fully formed sentences. The Department of Defense funded this research and program in hopes for their benefit to use the software for themselves. Their funding was obviously beneficial and brought much progress to past research. Years went by and automatic speech recognition worked its way further and further into commercial applications such as dictionaries and the business world until today where its part of our everyday activities.

Potential Benefits

Automatic speech recognition has brought organization to problems, efficiency to the business world as well as super power like abilities to people in need. With its range of skills, it can be used for, a variety of business and medical progressions that could expand upon it. As stated by Source Security, "voice recognition is non-contact, non-intrusive and easy to use" (Schelps). This software allows a representative to deal with multiple customers at once, but also be hundreds of miles away but still get the same information from the customer and be able to communicate back with them. Customers

can purchase products bringing in income for the company while using the company's automatic speech recognition decreases expenses because the company would need less on-call representatives. Instead companies can those representatives to deal with more in-depth problems and questions and use the software for simple tasks and issues that can be handled with simple statements and directions. Also the software is very direct and simple to use so people of all ages can use it and benefit from the software without needing instructions, its an automatic understanding. The elderly, who may not be technology updated, do not need someone to teach them terms and the process of automatic speech recognition, it an automatic understanding. For the blind disabled, they are able to speak into a microphone and have their words appear on a computer possibly in a document if they want to write a book or a paper for school with out the need for assistance. Where as before they would need to speak their thoughts as someone else typed them on a computer for them. Also with commands, they can tell the computer to do things with out having to see the computer itself and choose with their hands and eyes but rather with their voice. For security benefits, the ability that automatic speech recognition can understand both letters and numbers allows verbal passwords to be created with mixtures of the alphabet and numbers on top of the security system recognizing the individual's voice, which allows them to be of a much higher level of security than with just entering in digits and numbers into a key pad. Hackers are always becoming smarter and smarter and we need technology and security to keep up with them and be a step ahead of them in order to keep our selves protected.

Potential Problems

All new advancements come with their flaws and in the case of automatic speech recognition, there is quite a few. First of all, everyone's voices are different that is what makes us unique but it also generates problems. "The fundamental frequencies of speech sounds uttered by woman are about an octave higher than those of men, and those of children are even higher" (Ainsworth, 1939). Not only does voice differ person to person but it also depends on our age and our sex. When a toddler, they are just learning to speak so they do not pronounce sounds and words properly, and then children go through puberty where if a boy, their voice will squeak and change dramatically over a short period of time until settled into their so called set voice for life. How is voice recognition suppose to understand all these different stages of growth all at one time on top of accents and mumbling, it cannot. Also, with so much migration there are so many different accents;

> People from different parts of the country and different social an economic
> phenome from another in certain words, and changes in the rhythm and inonation
> of the utterance. This variability is even greater with people speaking a second
> language. The competence with which it is spoken depends on the motivation,
> intelligence, and perceptual and motor skills of the speaker, and also of the age at
> which the second language was learned (Ainsworth, 1939).

With the example that Bostonians do not pronounce their "R's" and that New Yorker's replace their "E's" for "A's," its merely impossible for technology to comprehend accents of Americans on top of other countries if real live people have a hard time understanding each other. In addition, white noise is an issue, which is "background noise as the caller interacts with the system," that may give false triggers to the recognition or

the software that may not recognize at all what the speaker is saying (Schelps). There are still many flaws to be adjusted to the program to make it a more credible software.

Conclusion

Technology is constantly advancing day to day and the more it advances, the more humans deal with technology rather than with humans themselves. Technology is able to deal with minor tasks and problems allowing humans to deal with and spend more quality time on bigger, more important issues. Automatic speech recognition allows a cheap, efficient way for businesses and commercial products to communicate with their customers without spending money on representatives to communicate harmless information like business hours. Any new technology will have flaws but automatic speech recognition is definitely a positive advancement and will continue to benefit the task of multitasking and aid the handicap. Flaws will continue to decrease and improvements will continue to increase just like in any new in invention.

Reference Page

Ainsworth, W. A. (1939). Speech Recognition by Machine. (p. 51). London United

 Kingdom: Peter Peregrinus Ltd.

This Source is Valid because the booked is dedicated entirely to the process of automatic

speech recognition and how it is made from the noises that come from our mouth to the

responses the computer makes when it hears the noises. The book addresses pro and cons

of speech recognition which I also talk about in my research paper.

Pinola, M. (2011, November 02). Speech recognition through the decades: How we

 ended up with siri. PCWorld, Retrieved from

 http://www.pcworld.com/article/243060/speech_recognition_through_the_decade

 s_how_ we_ended_up_with_siri.html

February 23, 2012

PCWorld is a online magazine which specializes in technology, they actually sell

products themselves so they obviously know what they are talking about. This magazine

discusses other similar topics such as tablets, phones, and laptops. The website has

privacy policy and "contact us" section to reach them for questions and concerns.

Schelps, D. (n.d.). Voice recognition – benefits and challenges of this biometric

 application for access control. Retrieved from

 http://www.sourcesecurity.com/news/articles/co-3108- ga.4100.html

February 23, 2012

Source Security talks specifically about security details and the security market. They sell

different types of security packages and discuss why their security is great which is where

they discuss automatic speech recognition and it advantages and what it does. They are a well known and established industry.

Speech recognition . (1997, June). Retrieved from

http://searchcrm.techtarget.com/definition/speech-recognition

February 29, 2012

Tech Target is a valid source because is discusses my particular website down to the definition and it also gives tutorials about other technological problems. It has expert advice as well as blogs for outside opinions and advice. This website gives direct definitions for terms used for my topic and for research paper. These are valid because by the definitions they shows, "people are read" articles to relate similar topics.

Speech interfaces are ready to listen. (2001, October 22). CNN. Retrieved from

http://articles.cnn.com/2001-10-22/tech/speech.interfaces.idg_1_speech-recognition-   speech-interfaces-phonemes?_s=PM:TECH

February 24, 2012

CNN is a credible source because they are a well know news source. They site their sources when they use outside information, and it is known around the world making it an international credible source. CNN is constantly updating their stories and following new topics with the media.