Speech Recognition and the future of artificial intelligence.

Group Members:
Sheldon Garwood
Nathan Nash
Romario Tomlin

History

- **Speech Recognition has been around since the 1960s and it has seen steady, incremental improvement over the years.**

- **Early systems were acoustic phonetics-based and worked with small vocabularies to identify isolated words.**

- **They now have very large vocabularies and continuous speech is able to be recognized.**

What Is Speech Recognition?

Speech recognition is defined as a process by which a computer takes a speech signal that is recorded using a microphone and converts it into text in real-time. With speech recognition, the aim is to get the computer to understand spoken language. This includes reacting ineptly and being able to translate the speech input into another medium such as text.

Advantages of Speech Recognition

- **Faster than hand-writing or typing**
- **Hands-free capability**
- **Helpful for persons with disabilities. (Both physical and mental)**
- **Can help organizations to reduce costs, increase productivity and limit errors.**

Challenges Faced When Designing SR Systems

- **The accent of a speaker varies between individuals.**
- **The speaker's voice is often accompanied by surrounding noise which causes accurate recognition to be difficult**
- **The speed at which a speaker speaks may cause it to be difficult for the system to recognize individual words accurately.**
- **Homonyms (for example : "There and their" and "be and bee")**

How Speech Recognition Works.

1. **The individual speaks to the software via an audio feed.**

2. **The device which was spoken to creates a wave file of the words spoken.**

3. **The wave file is cleaned by removing background noise and normalizing volume.**

How Speech Recognition Works.

4. The filtered waveform is then broken down into what are called phonemes. (Phonemes are the basic building block sounds of words and language. English for example has 44 phonemes. These consist of sound blocks such as “wh”, “th”, “ka” and “t”)

How Speech Recognition Works.

5. Each phoneme is like a chain link and through analysis of them in sequence, from the first phoneme the software uses statistical probability analysis to deduce whole words and then complete sentences.

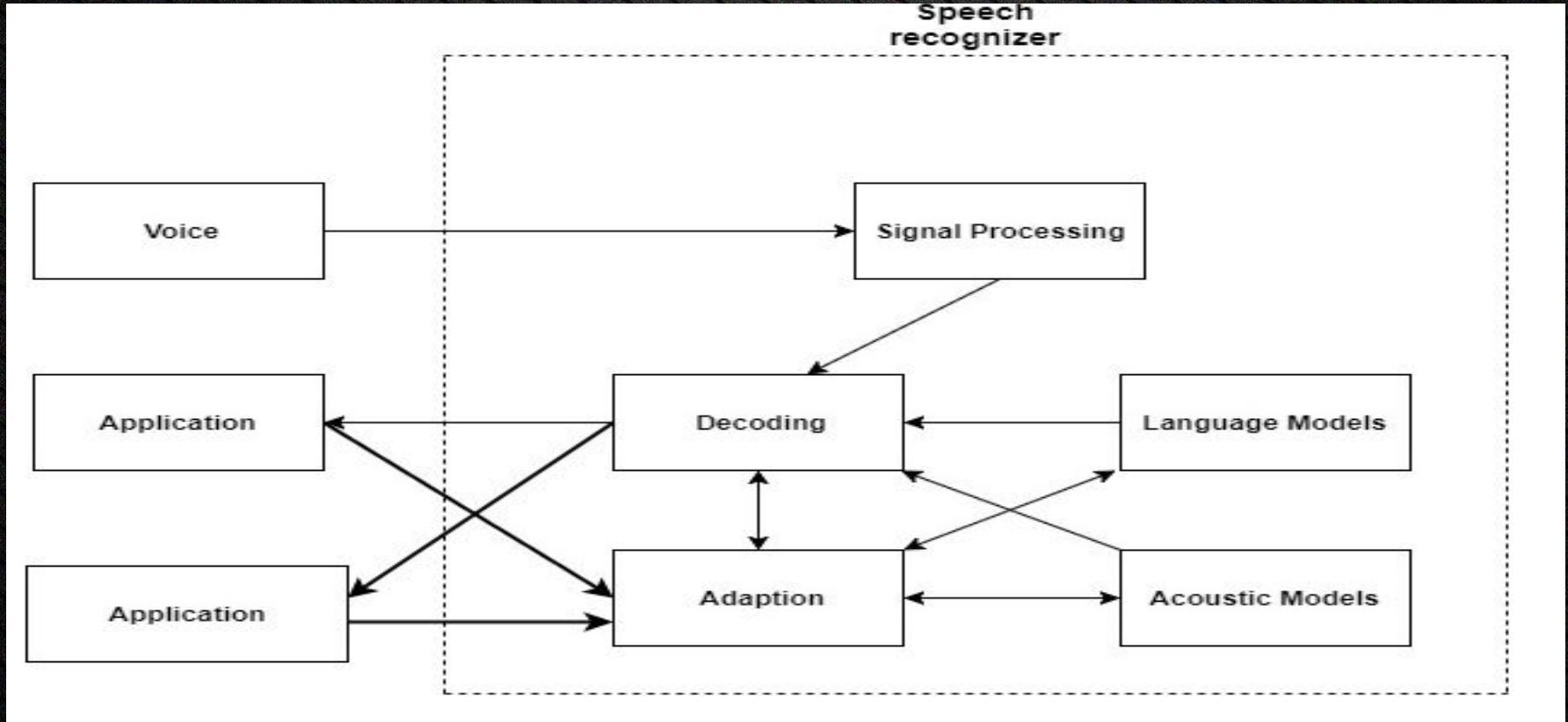
6. ASR now after it has “understood” your words, can respond to you in a meaningful way.

Models Used in Speech Recognition

Language Model - This is a file used by any speech recognition software to recognize speech. The file would consist of a large amount of words and the probability that a word will occur. It is the main decoder.

Acoustic Models - An acoustic model file consists of statistical representations of each distinct sound that makes up a word. These are called phonemes. This is how speech recognition adapts to different language accents.

Speech Recognition System Architecture



How Does Speech Recognition Differ From Voice Recognition?

Speech recognition is often used interchangeably with voice recognition but they are two different concepts.

Voice recognition can be used to identify individuals like with a face scan or fingerprint.

Speech recognition on the other hand can be used to control and maneuver with a computer.

Therefore with speech recognition it can be said that it matters WHAT was spoken whereas with voice recognition it matters WHO it was spoken by.

Voice Recognition vs. Speech Recognition

Therefore, with speech recognition it can be said that it matters WHAT was spoken whereas with voice recognition it matters WHO it was spoken by.

Applications of Speech Recognition Systems

1. **Health Care**
2. **Military- High Performance Aircrafts**
   -Air Traffic Control Systems
3. **Telephony - Smart-phones, Customer Helpline Services**
4. **Personal Computers**

What is Natural language processing?

- Another major area that has begun to work hand in hand with speech recognition is natural language processing.

- Natural language processing (NLP) is the ability of a computer program to understand human language that has been spoken. NLP is a component of artificial intelligence.

Ref: http://searchbusinessanalytics.techtarget.com/definition/natural-language-processing-NLP

Speech Recognition and Natural Language Processing

Natural language processing is the future direction of Speech Recognition technology.

It is designed to loosely simulate how we understand speech and respond appropriately.

Typical Natural Language processing vocabulary consists of more than 60 thousand words. That gives over 215 trillion possible word combinations.

Speech Recognition and Natural Language Processing

Thus it is impractical for an NLP SR system to scan all of its vocabulary for each word and process them individually. Natural language systems are designed to react to a much smaller list of “tagged” words that have been chosen which give context to requests.

System uses contextual clues to narrow what is being said so it can react appropriately and adequately.

Eg: If you say “I’d like to pay bills” tagged keywords may be “bills”

System would then use this word to find the context of other words used and not commit errors.

The Future Of Speech Recognition

Wearable Computers - Companies have already started to develop wearable technology that utilizes speech recognition such as Google Glasses. The use and production of these types of technology is surely set to increase in the years to come.

Home Automation - There is plenty of interest in the use of speech recognition in domestic appliances. This includes ovens, refrigerators, air conditioning units and washing machines.

The Future Of Speech Recognition

- **Accuracy will increase**
- **There will be more use of “intelligent systems” which may attempt to guess what the speaker intends/intended to say, rather than what was actually said, as persons often speak and make unintentional mistakes.**
- **Sound systems and microphones will be designed to adapt more quickly to changing background noise levels and different environments. There will be better recognition of external sounds to be ignored.**

Conclusion

It is observable that speech recognition will revolutionize the way individuals conduct business over the internet.

In the near future people will be using their home and business computers by speech and not by keyboard. Home automation will be mainly based on speech recognition systems.

References

Automatic speech recognition. (n.d.). [SlideShare slide]. Retrieved from https://www.slideshare.net/richiebmthimmaiah/automatic-speech-recognition-4721204?qid=d8136dd3-3619-4494-a911-6f8942e11041&v=&b=&from_search=2

Jones, M. (2016, March 4). Natural language processing: A financial and technological perspective. [Blog post]. Retrieved from https://www.plumvoice.com/resources/blog/natural-language-processing-financial-technology-perspective/

References contd.

Sahu, T. n.d. Speech recognition systems. [SlideShare slide].Retrieved from https://www.slideshare.net/diptimayasarangi/speech-recognition-system-seminar?qid=ab10779c-ac0c-419f-92f1-3550c34d4dcd&v=&b=&from_search=7

Zajechowski, M. (2014). Automatic speech recognition (ASR) software – An introduction. UsabilityGeek. Retrieved from https://usabilitygeek.com/?s=automatic+speech+recognition