4-18-2016

Potential of Dedicated Language Processing Units in Computer Voice Interaction

Justin Tolman

Potential of Dedicated Language Processing Units in Computer Voice Interaction

Abstract

This research explores the possibility of using a combination of a phonetic feature based binary encoding format for phonemes and dedicated coprocessors to improve computer voice interactions. While both speech synthesis and speech recognition have made great strides recently current performance still leaves much to be desired. Dedicated graphics cards and binary encoding have had a huge impact on computer graphics in the last two decades, and could do the same for voice interactions. Sources for this research consist primarily of course text books and documentation from open source software projects. At the time that this research was conducted it was purely speculative, but since then advances such as Google's TensorFlow AI, and NVIDIA's CUDA development kit make experimental research practical. This research indicates that it would be worthwhile to conduct experimental research on dedicated

The Phonemic Language Processing Unit:

Potential Benefits & Practicality of Experimental Research on Dedicated Hardware for Computer Voice Interaction

Justin Tolman
MDS400 – Capstone In Multidisciplinary Studies – Fall 2015
Advisor: H. Greg Wilson

BOISE STATE UNIVERSITY
SERVICE-LEARNING

BOISE STATE UNIVERSITY
The Mary Ellen Ryder Linguistics Lab

Objective

Determine viability of experimental research

Assess the viability and complexity of experimental research on a phonemic language processing unit using current technologies. Estimate when such a system might become practical.

Hypothesis

LPU research beneficial but not yet practical

The use of a phonemically based data storage format, combined with a dedicated language processing unit would likely result in significant improvements in accuracy and speed in computer voice recognition and synthesis.

Due to technological limitations testing will remain impractical for several years. (This portion of the hypothesis proved incorrect.)

Methodology

This research looks at current voice interaction systems, parallels in fields such as computer graphics, and hardware principals like Moore's law to estimate when and how dedicated hardware may benefit computer voice interaction.

Sources

"Arpabet." Wikipedia. Wikimedia Foundation. Web. 19 Oct. 2015.

Brandl, Georg. "The CMU Pronouncing Dictionary." The CMU Pronouncing Dictionary. Open SPHINX. Web. 19 Oct. 2015.

"Coprocessor." Wikipedia. Wikimedia Foundation. Web. 19 Oct. 2015.

Halverson, Zoann. Interview

Hayes, Bruce. Introductory Phonology. Malden, MA: Wiley-Blackwell, 2009. Print.

"IPA Chart." Full IPA Chart. International Phonetic Association. Web. 19 Oct. 2015.

Mattys, Sven, Ann R. Bradlow, Matthew H. Davis, and Sophie Scott. Speech Recognition in Adverse Conditions: Explorations in Behaviour and Neuroscience. Hoboken: Taylor and Francis, 2013. Internet resource.

O'Grady, William. Contemporary Linguistics: An Introduction. U.S. ed. New York: St. Martin's, 1989. Print.

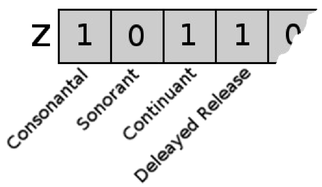
Payne, Thomas Edward. Exploring Language Structure: A Student's Guide. Cambridge, UK: Cambridge UP, 2006. Print.

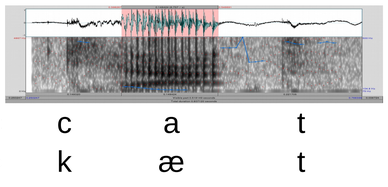
Background

The problems

- **Accuracy vs. speed**
  Modern speech recognition systems frequently produce inaccurate results.
- **Quality**
  Speech synthesis sounds mechanical and at times is difficult to understand.
- **Security**
  Some systems now send speech data to external servers for precessing leading to privacy and security concerns.
- **Data loss & reconstruction**
  Many (if not most) systems use text based encoding and storage. Text doesn't have a one to one relationship with human speech. This has parallels to using extremely lossy compression. It speeds processing and reduces storage space at the cost of detail and accuracy.

  Using a phonetic encoding improves the situation, but doesn't completely resolve it. For example water can be pronounced: [watəɹ], [waɾɹ], or even [wɔɾa].





Proposed solution

In the field of computer graphics addressed the need for rapid complex calculations by using a dedicated coprocessor (GPU).

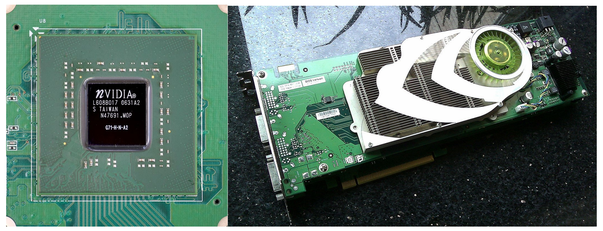
Once speed is addressed we can move on to accuracy. Most distinctive features of phoneme are either present or not. The remainder such as pitch and duration can be represented by small numbers, making phonemes easily converted to binary.

This system would also require mapping layers similar to keyboard mappings. A phoneme based lexicon, a language profile containing a phonemic inventory and allophony rules, and for some advanced applications an index of speech sounds.

Conclusion

LPU experimentation is practical now

The result of this research was a pleasant surprise. It does indicate that there may be significant benefits gained by using dedicated linguistic hardware and binary encoding. However thanks to recent developments (Google's release of TensorFlow, which integrates with NVIDIA's CUDA technology), experimental research on a phonemic language processing unit is currently viable with no hardware development, comparatively little programming, and relatively low hardware costs.



Graphics Processing Unit

Definitions

Linguistic definitions

Phone:	Any distinct spoken sound.
Phoneme:	The mental (or in our case computational) representation of a contrastive sound.
Allophone:	A set of multiple phones that represent a single phoneme.
Lexeme:	The smallest meaningful unit of language. (Often word roots & affixes.)
Lexicon:	A collection, index, or database of a Language's lexemes.

Computational definitions

Bit:	A single unit of data storage with an on or off state. Often represented as 0 or 1.
Coprocessor:	A processor used to supplement the CPU.