

LING 120: Language and Computers

Semester: FALL 2017

Instructor: Sowmya Vajjala

Iowa State University, USA

28 Aug 2017

Class Outline

- ▶ Quick recap of last week
- ▶ Encoding speech - continued
- ▶ Relating written and spoken language - an overview
- ▶ Wrapping up Topic 1!

Encoding Text: Recap from last week

- ▶ ASCII - 7 bit encoding
- ▶ Extending ASCII to accomodate new languages
- ▶ Unicode
- ▶ 3 different ways of Unicode conversion: UTF8, UTF16, UTF32
- ▶ How different writing systems look different on browser if you change the encoding.

Review question on encoding text

What is "America" (English letters) in UTF-8, UTF-16, UTF-32 - will it be same, or similar, or totally different?

Review question on encoding text

What is "America" (English letters) in UTF-8, UTF-16, UTF-32 - will it be same, or similar, or totally different?

Useful website (you can also use this for Assignment 1 instead of the other one):

<http://macchiato.com/unicode/convert.html>

Encoding Speech: Recap from last week

What speech properties are interesting?

- ▶ speech rate (fluency, number of pauses etc)
- ▶ Loudness/amplitude
- ▶ What sound frequencies correspond to different characters in human speech?
- ▶ How can we tell sounds apart with this frequency information?
- ▶ Pitch - how high or low is a sound (useful especially for identifying vowels)
- ▶ Intonation - rise and fall of pitch

Question from Friday's class

The following phrases/sentences represent some mishearings of songs and possible errors that a speech recognition software can also make. Try to guess an alternate version and post your responses on Canvas forum for today. That is your attendance for today:

- ▶ Example: "How to wreck a nice beach" - "How to recognise speech"
- ▶ "Secret agent man"
- ▶ "when the rainbow shaves you clean, you'll know"
- ▶ "with my knee on my mind"
- ▶ "language interpreters"
- ▶ "synthetic meditation"

Note: People who did not answer did not get attendance for friday.

Answers

- ▶ "Secret agent man" - secret asian man
- ▶ "when the rainbow shaves you clean, you'll know" - when the rain washes you clean you'll know
- ▶ "with my knee on my mind" - with money on my mind
- ▶ "language interpreters" - language and computers
- ▶ "synthetic meditation" - syntactic annotation

(First 3 are popular Mondegreens, last 2 are speech recogniser output for my speech).

Some of your responses from Friday

- ▶ Will you grab my peas - heard as "will you grab my keys"
- ▶ "grey chair" for "great share"
- ▶ "The slicer for the next class" - "The slides for phonetics class"
- ▶ "language interpreters" - languish under orders, or vintage interloper? "Synthetic meditation." - "Should take medication." / "send the invitation"

note: Some popular Mondegreens can be seen at:

<http://www.uh.edu/~mbarber/mondegreens.html>. You can see more on youtube.

How do we understand speech signals on a computer?

- ▶ Oscillogram: shows time on X-axis and changes in signal amplitude on y-axis
- ▶ Spectrogram: shows frequency on X-axis and time on Y-axis
- ▶ Darkness of a spectrogram: loudness of a sound.
- ▶ Sound frequencies, change of darkness, (and several other such measures) help us measure speech and understand words in the speech.

How do we understand speech signals on a computer?

- ▶ Oscillogram: shows time on X-axis and changes in signal amplitude on y-axis
- ▶ Spectrogram: shows frequency on X-axis and time on Y-axis
- ▶ Darkness of a spectrogram: loudness of a sound.
- ▶ Sound frequencies, change of darkness, (and several other such measures) help us measure speech and understand words in the speech.
- ▶ Let us see an example using one software that analyses speech
- Praat

How do we understand speech signals on a computer?

- ▶ Oscillogram: shows time on X-axis and changes in signal amplitude on y-axis
- ▶ Spectrogram: shows frequency on X-axis and time on Y-axis
- ▶ Darkness of a spectrogram: loudness of a sound.
- ▶ Sound frequencies, change of darkness, (and several other such measures) help us measure speech and understand words in the speech.
- ▶ Let us see an example using one software that analyses speech
- Praat

Note: Read "Under the Hood 1" in Chapter 1 if you are curious about Spectrograms. Some of it is a part of Assignment 1 too!

Praat

- ▶ Praat is a free software package to analyse speech signals.
- ▶ You can record sounds, visualize the spectrogram and oscillograms for these sounds with Praat.
- ▶ You can pull out a smaller section of these files for further analysis
- ▶ You can alter the speech files and create new files
- ▶ You can even measure things like: creakiness, nasality in voice etc!

More information: <http://savethevowels.org/praat/UsingPraatforLinguisticResearchLatest.pdf>

Quick Demo of Praat

How do three ways of saying the same word differ in their spectrogram images?

Quick Demo of Praat

How do three ways of saying the same word differ in their spectrogram images?

... You will need to work with this tool for your Assignment 1!

Relationship between written and spoken language

- ▶ Automatic speech recognition: converting a speech sample into text representation
- ▶ Text to speech: converting text into speech representation

Relationship between written and spoken language

- ▶ Automatic speech recognition: converting a speech sample into text representation
- ▶ Text to speech: converting text into speech representation

According to you, which is easier? Why?

Automatic Speech Recognition

Some major issues in achieving this:

- ▶ Converting those recordings into individual sounds, and words.
- ▶ Identifying word boundaries (remember: there are no punctuation markers or white spaces as in text!)
- ▶ Different people pronounce differently - different accents.
- ▶ The issue of multiple languages occurs here too!

More on this when we talk about dialog systems.

Text to Speech Synthesis

Seems like a trivial task?

- ▶ Isn't it just recording individual sounds (called phones) and stitching them together by looking at text and form speech automatically???

Text to Speech Synthesis

Seems like a trivial task?

- ▶ Isn't it just recording individual sounds (called phones) and stitching them together by looking at text and form speech automatically???
- ▶ Yes, but different phones also sound differently based on their neighboring sounds - we need context for right pronunciation of a word and its phones!

Text to Speech Synthesis

Seems like a trivial task?

- ▶ Isn't it just recording individual sounds (called phones) and stitching them together by looking at text and form speech automatically???
- ▶ Yes, but different phones also sound differently based on their neighboring sounds - we need context for right pronunciation of a word and its phones!
- ▶ Further, the synthesized sound should feel human, not look robotic!
- ▶ Question: how many of you use Waze? Did you try changing voices of the speakers?

Writing Text on computer

We saw how to render a text on a computer (using encodings) so that you can read. We also saw a little bit about how to represent speech. What about writing?

Writing Text on computer

We saw how to render a text on a computer (using encodings) so that you can read. We also saw a little bit about how to represent speech. What about writing?

Three common methods exist:

- ▶ Using keyboard layouts
- ▶ Using soft keyboards
- ▶ Using phonetic mappings

Useful link: https://en.wikipedia.org/wiki/Input_method

Next Class

- ▶ Topic 2: Writing aids - introduction

Attendance for Today

If I record you speak, use ASR and convert it to text, and then use TTS to convert it again to speech - do you think I will get your speech sample back? Write a short explanation for your answer (few sentences). You can either write it now and give it, or post on Canvas forum for today and get attendance.