Real-Time Transcription of Radiology Dictation: A Case Study for TabletPCs

Wu FENG
feng@cs.vt.edu

Depts. of Computer Science and
Electrical & Computer Engineering
Virginia Tech

Biomedical & Bioinformatics Background

- This Talk
  - Making IT More Useful for Radiologists (and Hospitals)
  - Acknowledgments: IBM
- mpiBLAST (http://www.mpiblast.org)
  - A parallelized version of NCBI BLAST (Basic Local Alignment Search Tool) that delivers super-linear speed-up.

mpiBLAST reduces search time from 1346 minutes (22.4 hours) to just over 4 minutes!

# Nodes	mpiBLAST v1.0	mpiBLAST New
1	1.00	1.00
4	9.23	4.52
16	33.15	48.03
64	94.95	173.24
128	170.49	**305.49**

"The Design, Implementation, and Evaluation of mpiBLAST," *4th Int'l Conf. on Linux Clusters*, 6/2002.

"Parallel Genomic Sequence-Searching on an Ad-Hoc Grid" ACM/IEEE SC|06, 11/2006.
"How to do all-vs-all comparisons of 1000s of genomes"

Overview

- Motivation
- Approach
- Speech Recognition System
- Handwriting Recognition System
- Gesture Recognition System
- Integrated Multimodal Environment for Immediate Radiology Transcription
- Usage Scenarios
- Case Study
- Conclusion

Radiology Transcription Today?

- Radiologist dictates x-ray diagnoses into a tape recorder.
- Radiologist forwards the tape to a transcriber who types the analyses into hardcopy reports for the radiologist.
- Radiologist receives the hardcopy reports 24 to 72 hours later (i.e., after having looked at potentially hundreds of other x-ray analyses in the interim).

Motivation: Radiology

- Issues
  - Previously such transcription was performed at (or near) the hospital. Now such transcription is outsourced overseas?!
  - Slow turnaround means that the radiologist cannot be expected to remember exactly what (s)he dictated for each individual x-ray, thus increasing the liability to the hospital.
- Challenge
  - Immediate turnaround time on radiology transcription
    - Lower cost (by leveraging information technology to eliminate the need for transcribers).
    - Significantly reduced liability to the hospital.

Motivation: Information Technology

- Problems with Computers, PDAs, and PocketPCs
  - Ease of Use
    - Computers: Users have been forced to adapt to "unnatural" input devices, e.g., keyboard, mouse, trackball.
    - PDAs and PocketPCs: Relatively good due to natural metaphor that "PDA = paper" and "stylus = pen"
  - User Productivity
    - Keyboard and mouse are slower than speech input but faster than stylus input, e.g., sending e-mail.
      - ✓ Keyboard: 60-100 words per minute (wpm).
      - ✓ Stylus: 20-30 wpm.
      - ✓ Speech: 150-250 wpm.
- Is the "Tablet PC" an ideal platform?
  - Yes (and no).

Overview

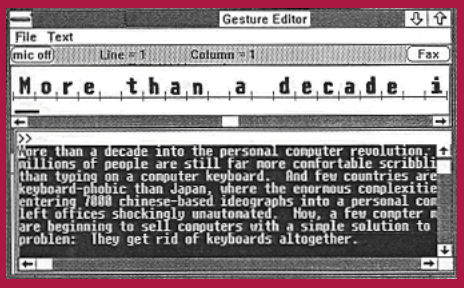
- Motivation
- Approach
- Speech Recognition System
- Handwriting Recognition System
- Gesture Recognition System
- Integrated Multimodal Environment for Immediate Radiology Transcription
- Usage Scenarios
- Case Study
- Conclusion

Approach

- Speech to improve ease of use & increase productivity.
  - Requirements
    - A speech recognizer that runs in real time and relatively error-free, e.g., > 95% recognition accuracy.
    - Tools for correcting speech-recognition errors:
      - Simple and natural to use, e.g., stylus and keyboard.
      - Seamlessly integrated with the speech recognizer.
      - Virtually error-free.
  - Necessity of Requirements
    - Ensure that productivity is significantly better than typing and mouse-ing.

Overview of Proposed Solution

- Integrated environment that seamlessly integrates different modes of input: speech, handwriting, and gestures.



  - Dictate x-ray analysis into a speech recognizer.
  - Correct any recognition errors (or even re-organize the dictation) using handwriting and gesture recognizers.
  - Print and sign hardcopy of the radiology transcription.

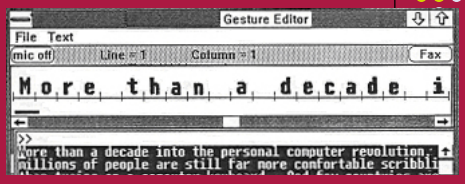
Overview

- Motivation
- Approach
- Speech Recognition System
- Handwriting Recognition System
- Gesture Recognition System
- Integrated Multimodal Environment for Immediate Radiology Transcription
- Usage Scenarios
- Case Study
- Conclusion

Speech Recognition System

- Requirements
  - Real time and relatively error-free, e.g., 95% accurate.
- Currently Available Systems
  - IBM ViaVoice & Dragon Systems Naturally Speaking.
    - Continuous-speech processing produces pseudo-real-time and error-prone, i.e., typically 70-75%, recognition.
- Our System
  - Reduce computational complexity; improve recognition.
    - Customized speech profile for each individual.
    - Discrete speech processing (i.e., talking like a robot).
    - Result: Real-time and relatively error-free recognition, i.e., 95% on average.

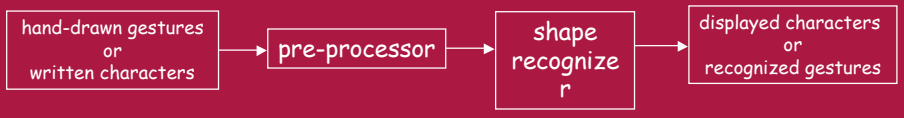
Handwriting Recognition System



- Requirements
  - Virtually error-free.
- Current Solutions
  - Palm Pilot Graffiti
    - Requires learning a new "alphabet" and is error-prone.
  - Virtual keyboard (as a back-up)
- Our System
  - Discrete block letters from the regular alphabet (with each letter confined to a virtual box).
  - Virtual keyboard as a back-up.

Gesture Recognition System

- Handwriting and Written-Gesture Recognition
  - Use the same software architecture



  - Capture temporal information as the user writes.
    - Number of strokes, order of strokes, direction of strokes, and speed of strokes.

- Major Difference
  - Gesture recognition is more general than handwriting recognition. Why?

Complications with Gesture Recognition

- Handwriting
  - Depends only on shape consistency.
    - The difference between the same symbol produced at different times is less than the difference between different symbols.

- Gestures
  - Symbols routinely violate shape consistency but are still recognizable as the same symbol by the human eye.

Gestural Variations

Size

Non-Linear Scaling

Orientation

Direction Reversal

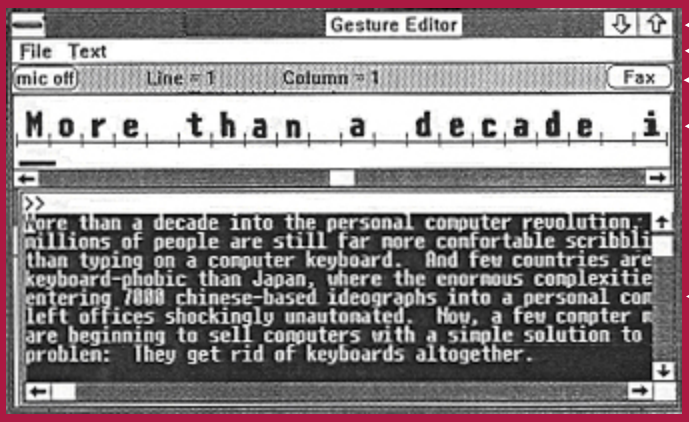
W. Feng, feng@cs.vt.edu

Defining a Gesture Alphabet

	Select the region encompassed by the oval.
[ ]	Select the linear region encompassed by the brackets.
⌜ ⌟	Select the vertical region bounded by the corners, e.g., complete lines of text.
	Delete selected region.
	Move selected region.
—	Delete "crossed out" linear region.
//	Split a line (or object).
?	Display alternate word list.
>	Playback speech.
∪	Undo.

W. Feng, feng@cs.vt.edu

Integrated Multimodal Environment

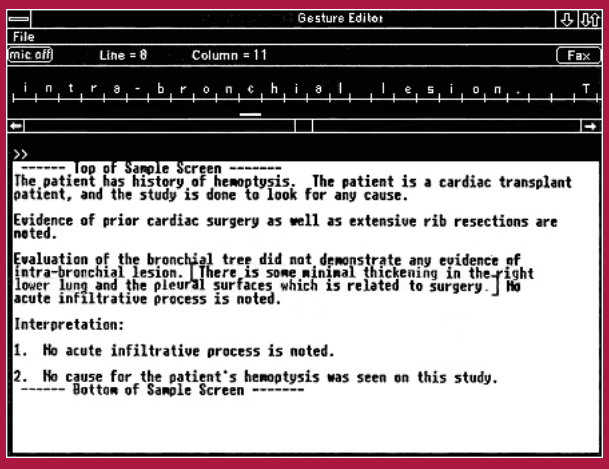


Overview

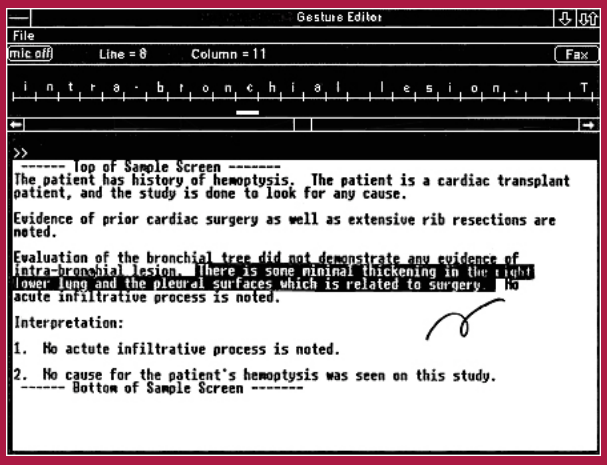
- Motivation
- Approach
- Speech Recognition System
- Handwriting Recognition System
- Gesture Recognition System
- Integrated Multimodal Environment for Immediate Radiology Transcription
- Usage Scenarios
- Case Study
- Conclusion

Usage Scenarios

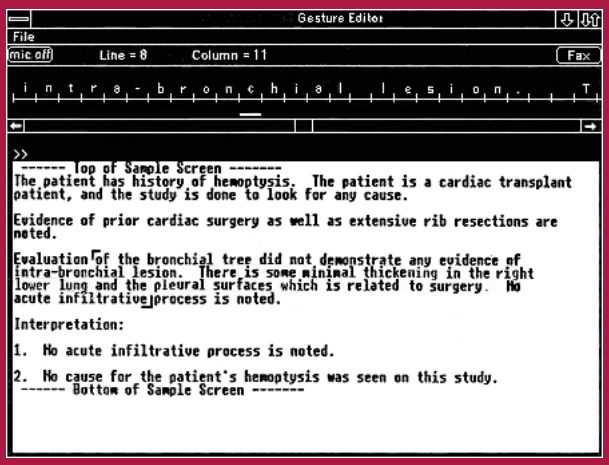
Brackets: Select Linear Text



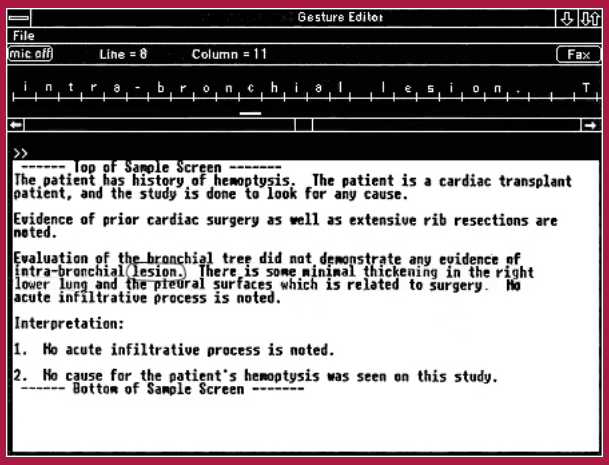
Delete Gesture: Removes Selected Region



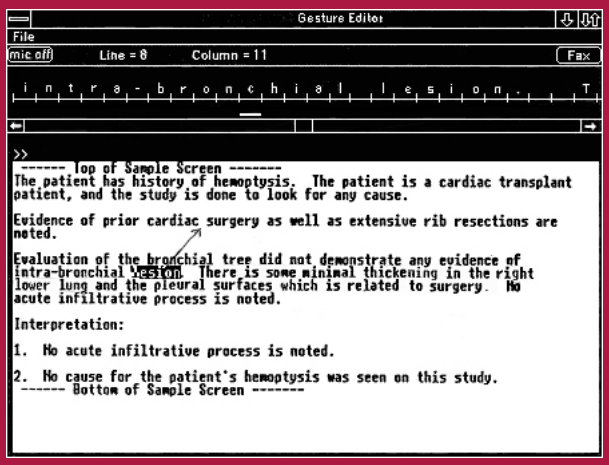
Corner Gestures: Select Row(s) of Text



Oval: Select a Region of Text



Arrow: Moves Selected Text



Informal Case Study

- Setting: A leading research hospital.
- Test Subjects: Five
  - Four native English speakers, one European.
  - All male.
- Duration of Case Study: One Working Week
  - One day set-up and training.
  - Three days of use.
  - One day of gathering information and feedback.

Empirical Data

- Speech Recognition Rates
  - 91% for the European (with a thick accent).
  - 94%, 95%, 97%, and 97% for the native English speakers.
  - Recognition rates improved over the course of three days.
- Handwriting & Written-Gesture Recognition Rates
  - Given the infrequency of use, not enough data to really make any conclusive statements.
  - Handwriting
    - Initially below 90% for everyone. Improved significantly over the three days. Clear preference for virtual keyboard.
  - Gesture
    - Virtually 100% due to distinctness of gestures.

What Test Subjects Said

- The Negative
  - Speaker profiles
    - Remembering to load speaker profile a priori.
    - Training the system to create a custom speaker profile.
  - Turning speech recognition on and off.
  - Having to talk in discrete speech.
    - Slower than dictating into a tape recorder.
  - Requiring too much precision in handwriting.
    - Virtual keyboard preferred (*a la* what PDA market found out about thumbpads).
- The Positive
  - Very usable.
  - Very promising and innovative use of technology for immediate transcription.

Conclusion

- (Traditional) Graphical User Interface of a PC
  - Enhances ease of use and increase productivity.
  - Problem: Keyboard and mouse are not natural input devices to the novice user.
- PDA & PocketPC Interface
  - Enhances ease of use but reduce productivity.
  - Problem: Stylus writing is 2x-5x slower than keyboarding and 8x-10x slower than talking.
- Tablet PC Interface
  - (Arguably) enhances ease of use with marginal increase in productivity (lack of seamless integration of input devices).

Conclusion

- Integrated Multimodal Environment
  - Speech
    - Natural interface to enhance ease of use.
    - Productivity increase of 2x-4x over keyboarding and 8x-10x over writing with a stylus.
  - Gestures and Handwriting
    - Natural interface to enhance ease of use.
    - Supplements the speech recognition environment.
- Re-visiting the Challenge …
  - Immediate turnaround time on radiology transcription
    - Lower cost (by leveraging information technology to eliminate the need for transcribers).
    - Significantly reduced liability to the hospital.

W. Feng, feng@cs.vt.edu