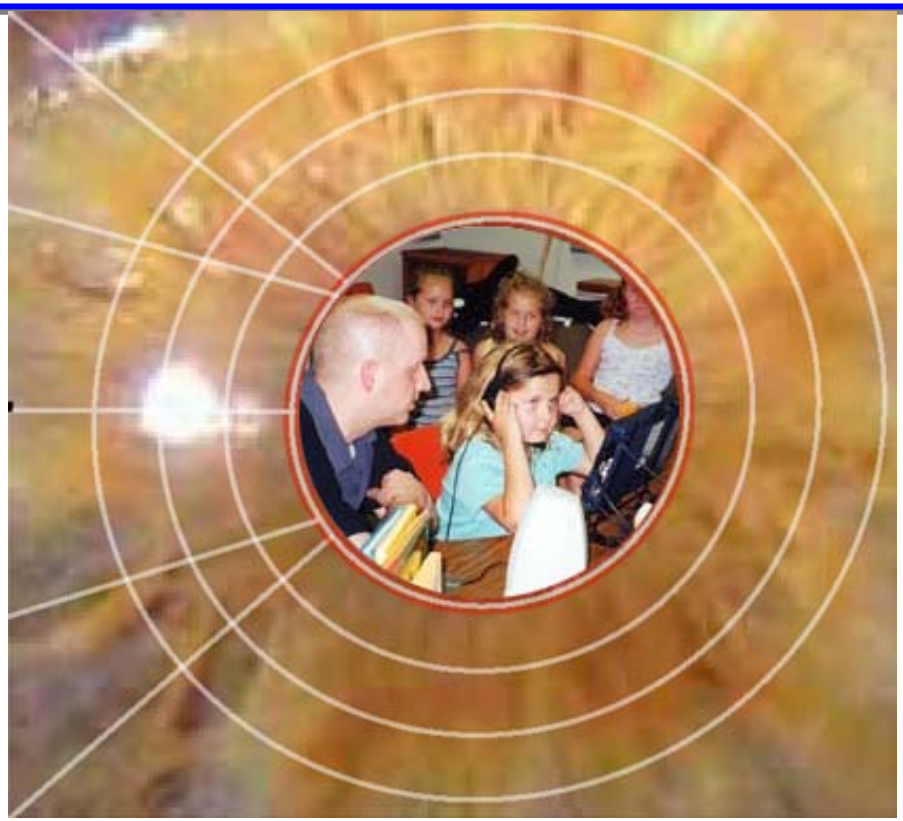
Implicit User-Adaptive System Engagement in Speech, Pen and Multimodal Interfaces



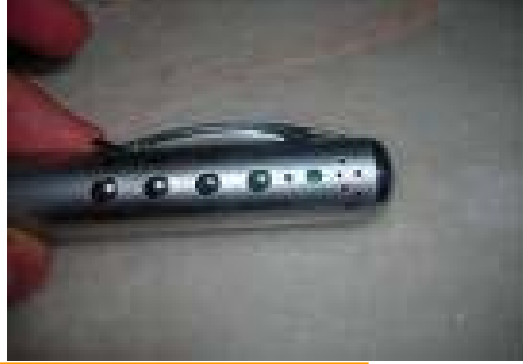
Sharon Oviatt
oviatt@incaadesigns.org

ASRU Workshop
December 2007

Introduction: Why Design Implicit System Engagement?

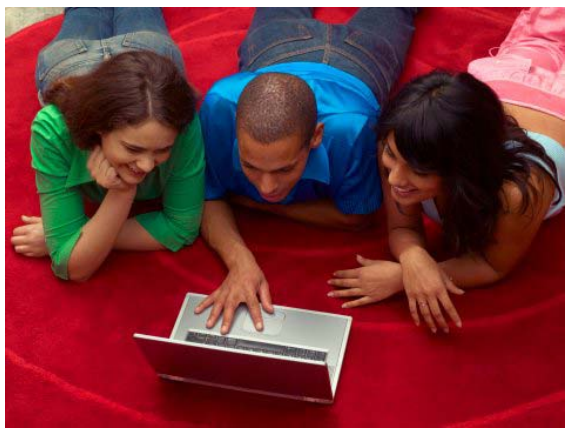
- Systems capable of *implicit engagement* minimize cognitive load due to interface so students can focus on learning activities
- Implicit engagement useful for educational & mobile tasks, where load is high & excess load extracts performance cost
- System engagement & disengagement can be large percentage of total interaction *(40% steps during voice dialing)*, so very distracting
- Goal— Prototyping of implicit user-adaptive interfaces for field use, collaborative use & educational activities

Unsolved Problems: Speech Open-Microphone & Pen Moding



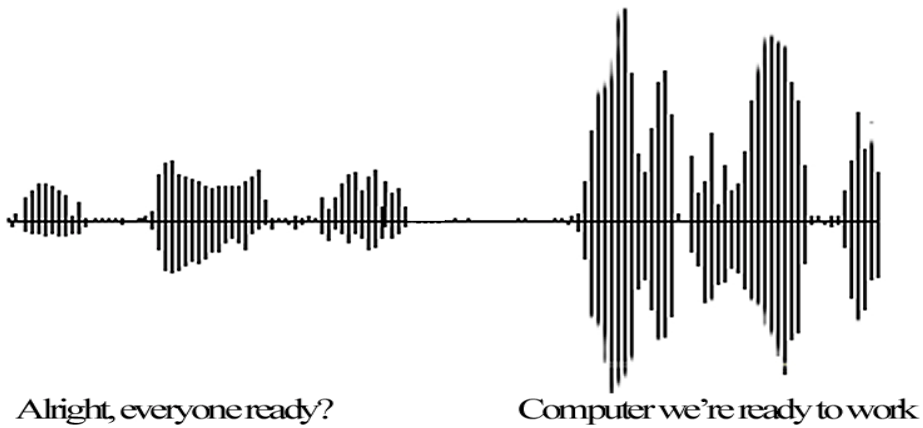
- Explicit speech push-to-talk & pen button press engagement techniques dominate commercial recognition systems

  (except pressure control of stroke thickness in drawing, but not user-adapted)
- Active research on implicit audio-visual speech engagement using combined cues *(language processing, gaze & head position, lip movement)*
- Limited success since empirical data still lacking on best information sources *(gaze often misleading, prevalence of self-talk ignored)*



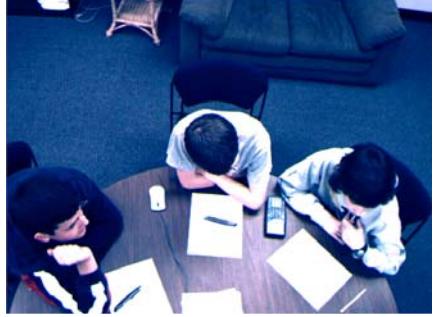
Unsolved Problems: Speech Open-Microphone & Pen Moding (cont.)

- Recent work shows amplitude can reliably distinguish when speech is self-directed, peer-directed, or computer-directed
- But, open-microphone engagement remains unsolved problem, especially for multi-person interactions & field settings
- Comparable empirical research lacking for interactive pen use



What is Human-Centered Design? Research Strategy & Philosophy

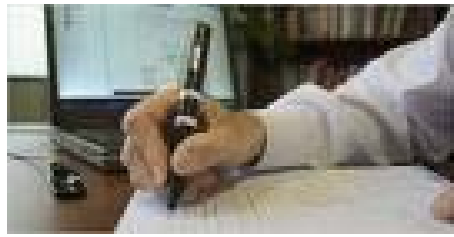
- Human-centered interfaces (HCI) are tailored to students' natural communication patterns & work practice
- Accommodating natural communication patterns reduces users' cognitive load & errors, since not under full conscious control— thereby improving usability
- Adapting interfaces to individual users improves system reliability due to large individual differences in communication patterns
- HCI provide users with valuable functionality they are motivated to achieve— e.g., *being recognized correctly by an interlocutor*



Research Goals: Questions & Hypotheses

- Do users spontaneously adapt communicative energy when addressing a human vs. computer assistant during meetings?
  - Speech amplitude & pen pressure increase to computer
- Following system failures to engage, do users further increase signal energy?
  - Uniformly & forcefully increase amplitude & pressure
- Can systems be designed that engage *entirely implicitly* based on users' natural energy adaptations?
  - Yes, with speech amplitude most reliable

Research Goals: Questions & Hypotheses (cont.)



- With experience using system, can people learn to further differentiate their energy when addressing a computer vs. human, so reliable system engagement is optimized?
  - Users will increase energy to computer & decrease to human, improving system engagement over time
- Can such system engagement occur without users' awareness?
  - Most users unaware of energy changes
- Can implicit engagement avoid distracting students during problem solving, so performance remains high?
  - Correct solutions maintained during >100 engagements per session

Theoretical Context: Generalizing Lindblom's H & H Theory

- Interpersonal speech varies stylistically along spectrum:
  hypo-clear *(relaxed)*  hyper-clear *(clarified)*
- Speakers assess how much explicit signal information their listener requires in given context
- Speakers adopt *hyper-clear*, *high-energy* speech when they expect or experience communication error, since it improves intelligibility
- Lindblom's theory provides basis for predicting higher energy communications when user addresses a computer
- We generalize this theory of interpersonal speech to:
  - Interactive pen modality
  - Interactive computer exchanges
  - Designation of interlocutors

Research Method

- Participants: 12 pairs of high school students
- Activity: peer tutoring on geometry problems
- Tutorial system facilitated solutions
  (displayed problems, formulas, terms, solutions, explanations)
- Students engaged system >100 times per session
- Example geometry problem & solution:

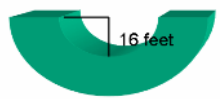
Daniel is building a half pipe so he can do bike tricks. If he wants the radius of the pipe to be 16 feet, how long should the plywood be for the curved inner area?

Answer: 50.27 feet (16π)

Example Solution:

Circumference of a circle = 2π (radius)
Circumference of 1/2 circle = π (radius)

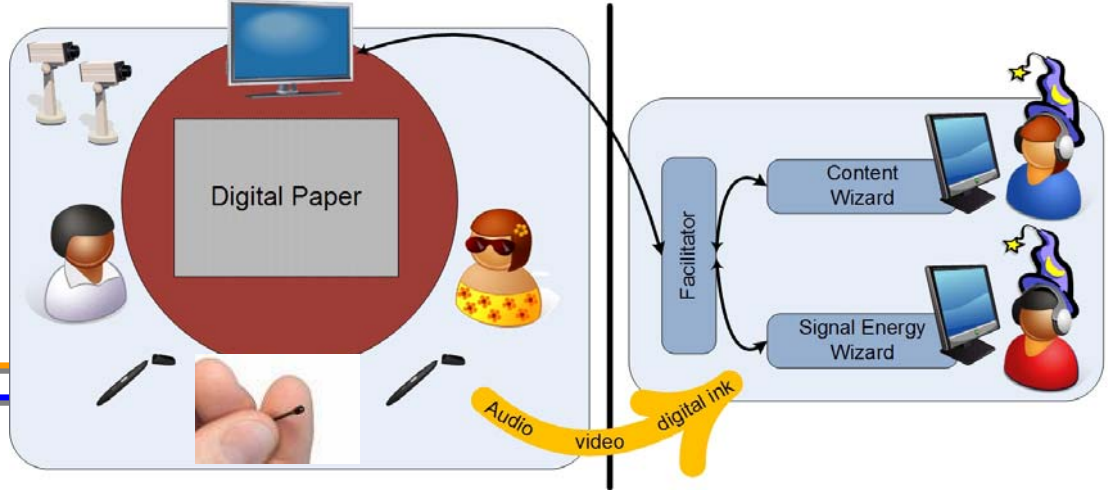
Circumference = π (16) feet
= 50.27 feet



Research Method (cont.)

- Longitudinal study: Students completed 2 sessions *(speech & digital pen input)*
- Within-subject factors:
  - Modality *(speech, pen)*
  - Intended Addressee *(computer, human peer)*
- Dual-wizard simulation: Collected audio, visual & digital ink user data during meetings *(synchronized & time-stamped)*
- Data collected: 24+ hours; >360 geometry tasks

Novel Dual-Wizard Simulation Method



- Content wizard (CW): Responded to speech or pen constructions *when semantic content was compatible with a system request*
- Signal energy wizard (SEW): Responded when energy *(amplitude, pressure)* of construction exceeded user-specific threshold
- Real-time contingent learning paradigm, with system engaging whenever signal energy met user's threshold
- Signal detection methodology: (***Hits vs. Misses*** *of system engagement attempts,* ***False alarms vs. Correct rejects*** *of interpersonal communications*)
- Wizard coordination supported by distributed agent architecture

Sequence of Events during Two Sessions

User Baseline

Main Session: Either Pen or Speech

4 Practice Problems

Triad 1
1-3

Triad 2
4-6

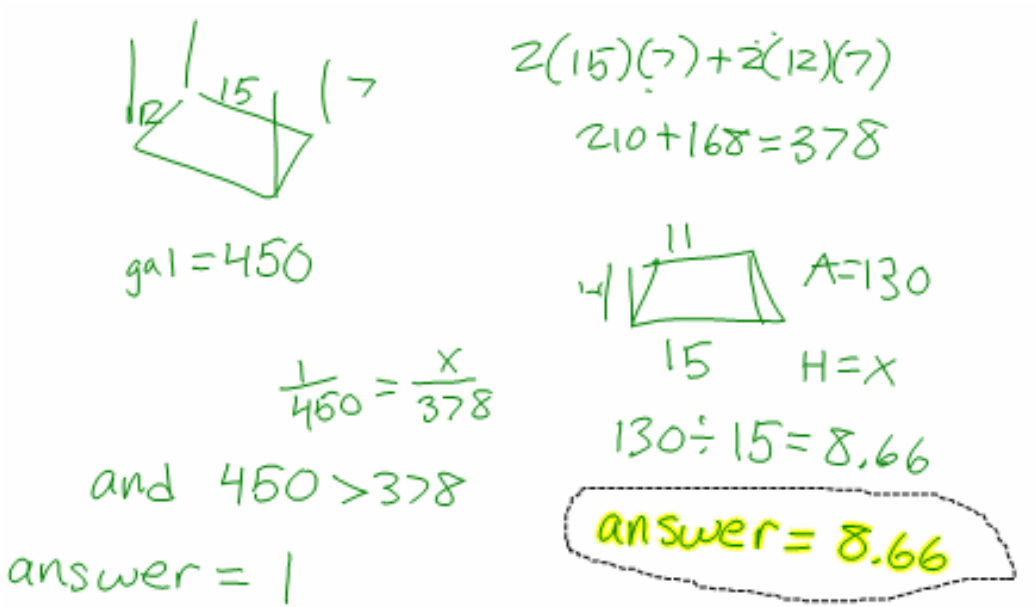
Triad 3
7-9

Triad 4
10-12

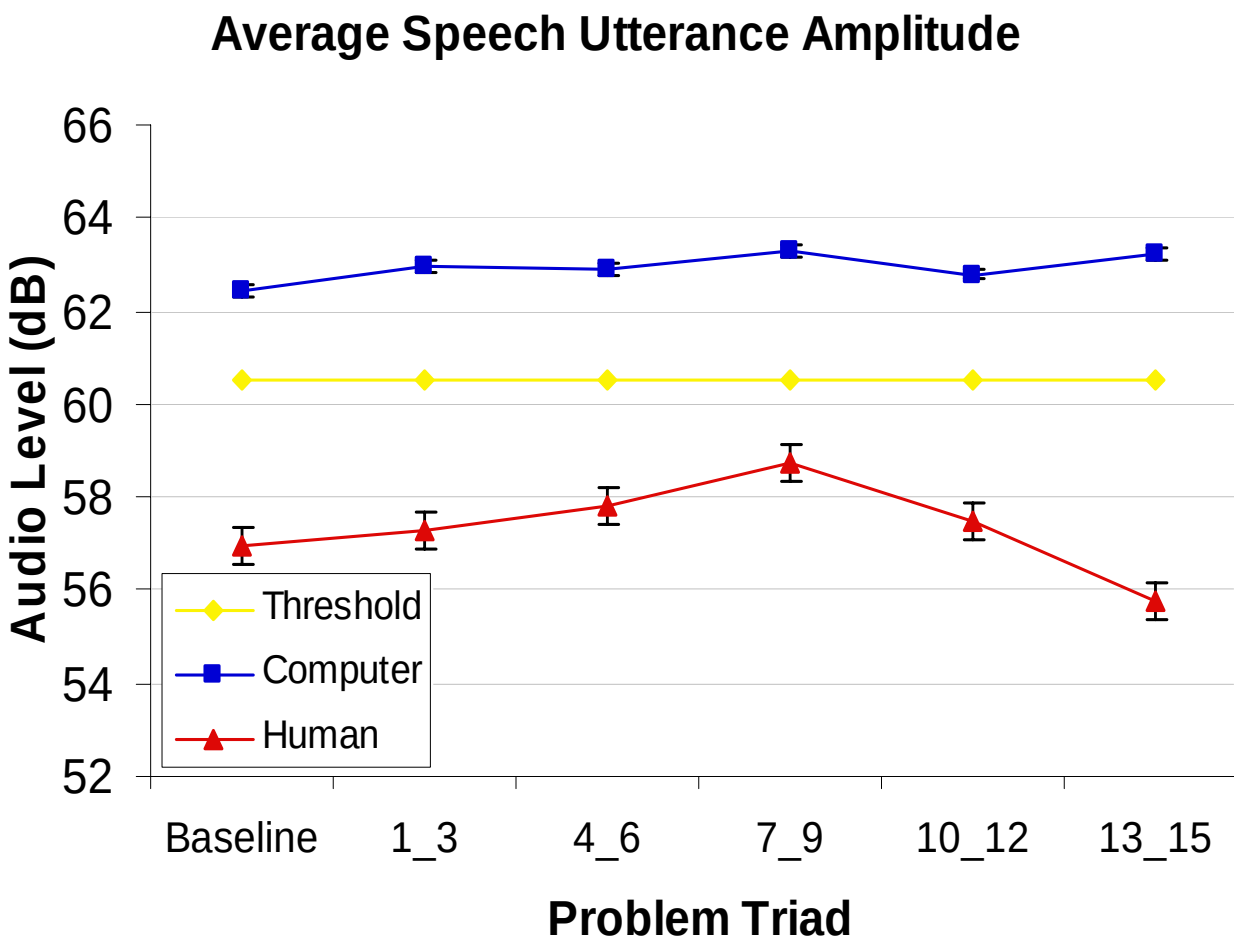
Triad 5
13-15

Example of Student Pen Input during Problem

- Wizards saw real-time streamed digital ink from student pens while they worked & could:
  - -Pan, rotate & zoom their displays
  - -Encircle ink constructions to calculate pen pressure

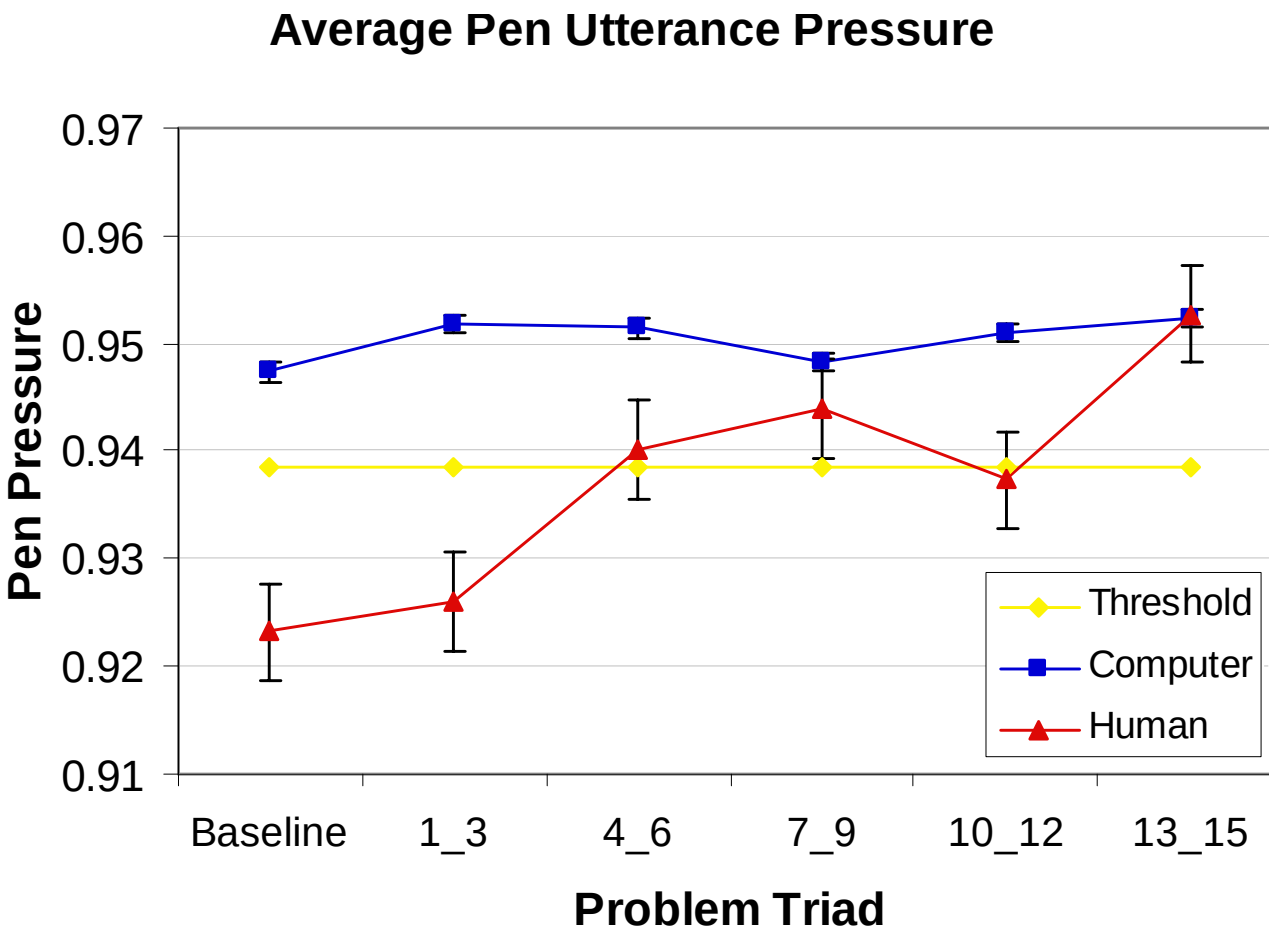


Average Speech Amplitude & Changes over Time during Computer- versus Human-Directed Input



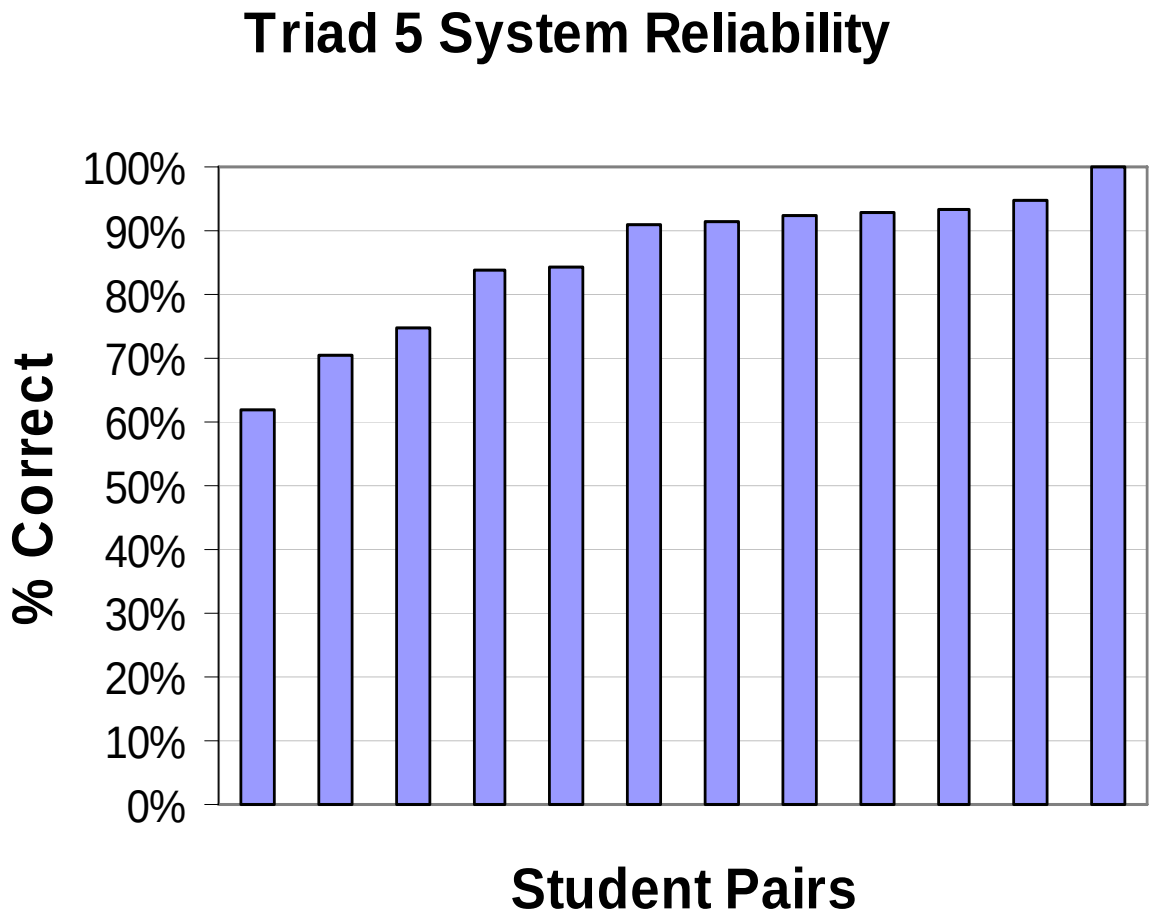
- Baseline amplitude higher to C than H
- Amplitude to human dropped over session
- Amplitude to computer increased marginally over session
- Amplitude differential between C & H expanded over session

Average Pen Pressure & Changes over Time during Computer- versus Human-Directed Input

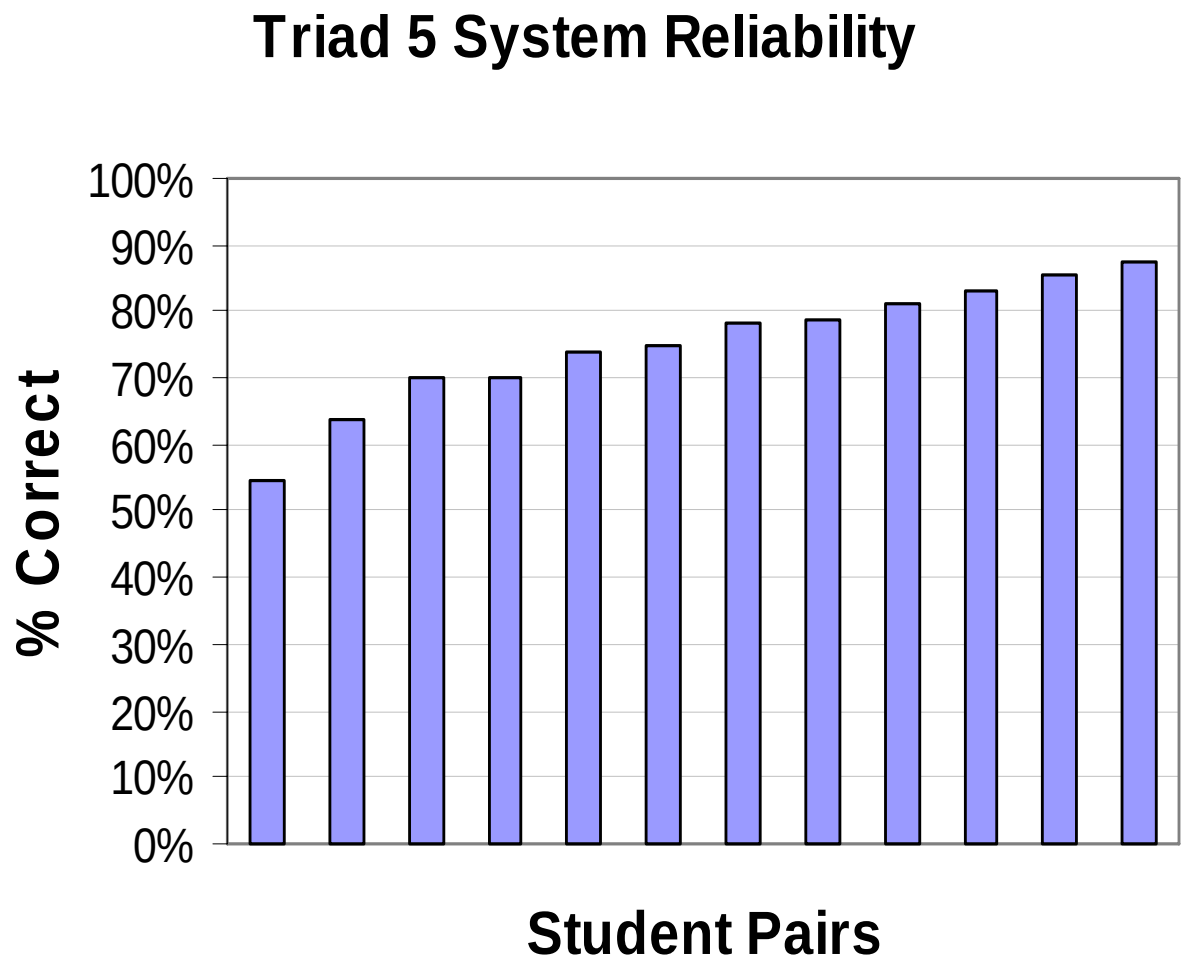


- Baseline pressure higher to C than H
- Pressure to computer increased over session
- No increased pressure differential between C & H over session

Average System Reliability in Simulated Speech Amplitude Engagement System for Different Individuals (Mean 86%)



Average System Reliability in Simulated Pen Pressure Engagement System for Different Individuals (Mean 75%)



Increase in Signal Energy Before & After Computer Misses

	Speech Amplitude (dB)				*Pen Pressure*			
Pair	**Pre**	**Post**	**Diff**	**% Energy Increase**	**Pre**	**Post**	**Diff**	**% Energy Increase**
1	57.3	59.1	1.8	**22.5%**	.922	.929	.007	**3.2%**
2	54.8	57.7	2.9	**39.3%**	.920	.929	.009	**4.0%**
3	60.6	63.6	2.9	**40.3%**	.923	.934	.011	**4.9%**
4	60.9	64.0	3.1	**42.8%**	.923	.938	.014	**6.5%**
5	59.8	63.0	3.2	**45.2%**	.924	.950	.026	**12.2%**
6	59.6	63.0	3.4	**48.2%**	.921	.948	.027	**12.4%**
7	60.8	64.9	4.1	**59.8%**	.922	.953	.032	**15.0%**
8	61.4	66.5	5.1	**80.0%**	.924	.963	.039	**18.7%**
Mean	*59.4*	*62.7*	*3.3*	***46.4%***	*.923*	*.943*	*.021*	***9.5%***

- All users increased speech & pen signal energy during repairs
- Speech amplitude mean increase 46.4%
- Pen pressure mean increase 9.5%

Students' Self-Awareness of Their Signal Energy Adaptations

- Students' self-reported awareness of using speech amplitude to successfully engage system:
  - 42% mentioned spontaneously
  - 50% mentioned when prompted
- Students' self-reported awareness of using pen pressure to successfully engage system:
  - 0% mentioned spontaneously
  - 8% mentioned when prompted
- Awareness of signal energy adaptations very limited!
- Greater awareness of speech amplitude to engage system, compared with pen pressure

Students' Ability to Maintain Performance Level

- When using speech amplitude engagement, math solutions 78% vs. 80% correct on 1st & 2nd half
- When using pen pressure engagement, math solutions 67% vs. 72% correct on 1st & 2nd half
- No deterioration in performance over session
- Performance significantly higher with speech amplitude engagement (79%) than pen pressure (70%)
- Better speech performance may be due to 11% greater reliability

Main Conclusions

- Students spontaneously, reliably, and substantially adapted communicative energy when using speech and pen modalities to designate & repair an intended interlocutor during computer-mediated meetings
- During baseline, both amplitude & pressure were higher to computer than human partner
- Users uniformly & forcefully increased signal energy when repairing an intended interlocutor *(46% & 10% relative amplitude & pressure increases)*
- Using speech, the amplitude differential between computer vs. human partners expanded by 2dBs over a session— yielding a 24.3% relative reduction in engagement error rate
- Pen pressure only partially adapted *(increased to C over session)*

Main Conclusions (cont.)

- System engagement accuracies ranged 75-86%, with amplitude engagement more reliable than pressure *(6 of 7 correct engagements)*
- Students had limited awareness of their adaptations
  (0% & 42% spontaneously mentioned using pressure or amplitude, respectively)
- In spite of >100 system engagements, implicit methods enabled students to maintain their math performance over extended time
- Comparing *same students solving same problems*, amplitude engagement supported 9.2% higher math problem correctness, perhaps due to substantially lower error rate

Interpretations & Future Directions

- Implicit engagement systems can be implemented effectively, while not requiring user awareness or undermining performance
- When interface is adapted to natural communication, system functioning is more transparent & users can learn to improve system reliability
- This work generalizes Lindblom's theory to different communication modes, human-computer interaction & designation of intended interlocutors
- Future directions:
  - Engagement methods based on combined cues
  - Application of language processing & machine learning in implemented systems
  - Integration of visual feedback techniques