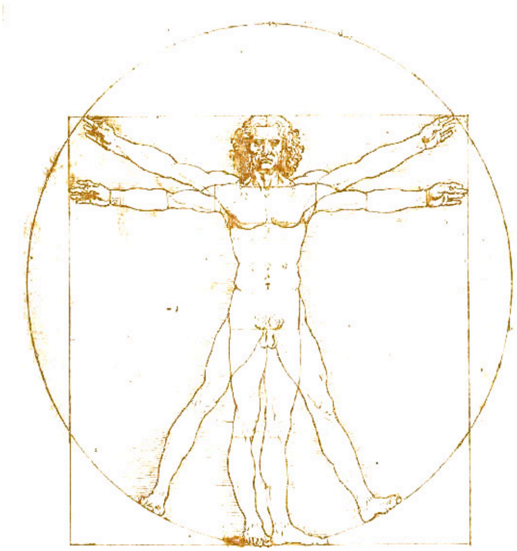
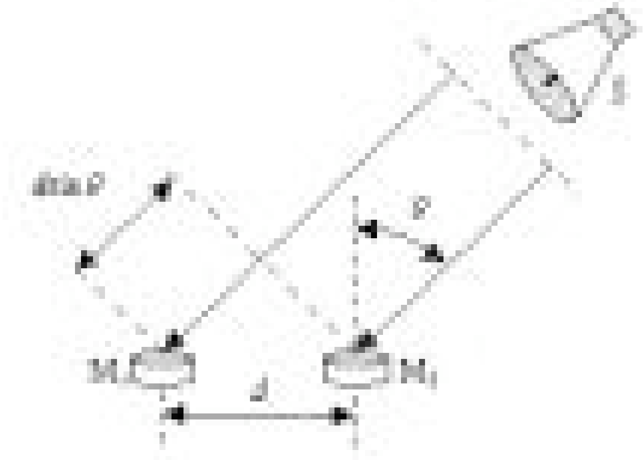
Interfacing with the Machine





Jay Desloge
SENS
Corporation

Sumit Basu
Microsoft
Research

They (We) Are Better Than We Think!

- Machine source separation, localization, and recognition are not as distant as they may seem.
- There are, in fact, already systems that achieve limited success in these areas.
- These machines provide many opportunities to investigate the interaction of machines with the human operator.

Consider: Hearing Aids

- Directional microphones can yield target-location (in front of wearer) intelligibility-weighted SNR improvements of up to 5-6 dB.
- Adaptive directional capability can yield higher SNR improvements (on the order of 8-12 dB).
- FM capability allows aid to receive signals from remote sources (TVs, remote microphones).

Perseo· BTEs – reliable beauty

The Perseo BTE family includes a full choice of sizes, colors and power levels. All feature an elegant, award-winning design with a refined, ergonomic shape. They have a patented housing construction with bonded plastic and silicon to effectively keep out debris and moisture. Environmental protection extends further with patented, disposable microphone covers to ensure consistent, reliable performance, in all conditions. With dual-microphone systems this is particularly important since electronic matching cannot successfully compensate for performance reduction as a result of debris and moisture. A comprehensive range of colors for the housing also caters for specific tastes and trends.

Original size
New small Perseo BTE 111/211

(Phonak Persio)

Consider: Tele/Video Conferencing

- Directional microphones used to identify and extract the sources from the environment.
  IW SNR improvements 5-6 dB on average.
- Active speaker is determined by microphone input.
- Voice-tracking capability can focus video camera on an active source within the environment.
  RMS loc. error < 10 deg.

(Polycom Soundpoint)

Consider: ASR State of the Art

	Type	Characteristics	WER
Meeting Room (16kHz)	Business Spontaneous	Task oriented, but includes true meetings collected in uncontrolled conditions Far-talking, but also have close-talking (head-mounted) for comparison	30% (head-mounted) 50% (distant)
Switchboard (Telephone)	Polite Spontaneous	Close-talking, relatively free of noise These are real people (with a slight bias toward females housewives and higher education), who don't know each other and have some conversation on some topic. Real data, but instrumented Conditions	15%
Broadcast News	"Planned" speech	"Found data" (exists in nature, not artificially collected) Spoken by professional speakers; not read, but speakers know what they are going to say in advance, and possibly Practice	9%
WSJ (Dictation)	Read speech	High-quality microphones, professional speakers, "Wall Street Journal" sentences (ie it's a rich, but restricted domain)	3-8%
String of Digits	Read speech	Easy task; no noise, close-talk	<0.5%

From Patrick Nguyen (MSR)

Consider: Wireless Communication, GPS

- Wireless communication links can connect team members (e.g., military, firefighter, police) and can provide clean, separated signals for each source.
- GPS can provide accurate information about the location of each source.
- Efforts have already been made to present these sources to the team members in a logical manner (e.g., spatialized audio).

What We Will Talk About

Given that these and other possibilities for human-machine interaction already exist, it is important to study how the humans and machines can interact in a manner that achieves the best possible performance.

We will discuss:

- Machine enhancement of human capabilities (H+)
- Human enhancement of machine performance (M+)
- Design factors in human-machine interfaces

Machines Enhancing Human Capabilities (H+)

- Despite their limitations, machines can outdo what we do

Vs.

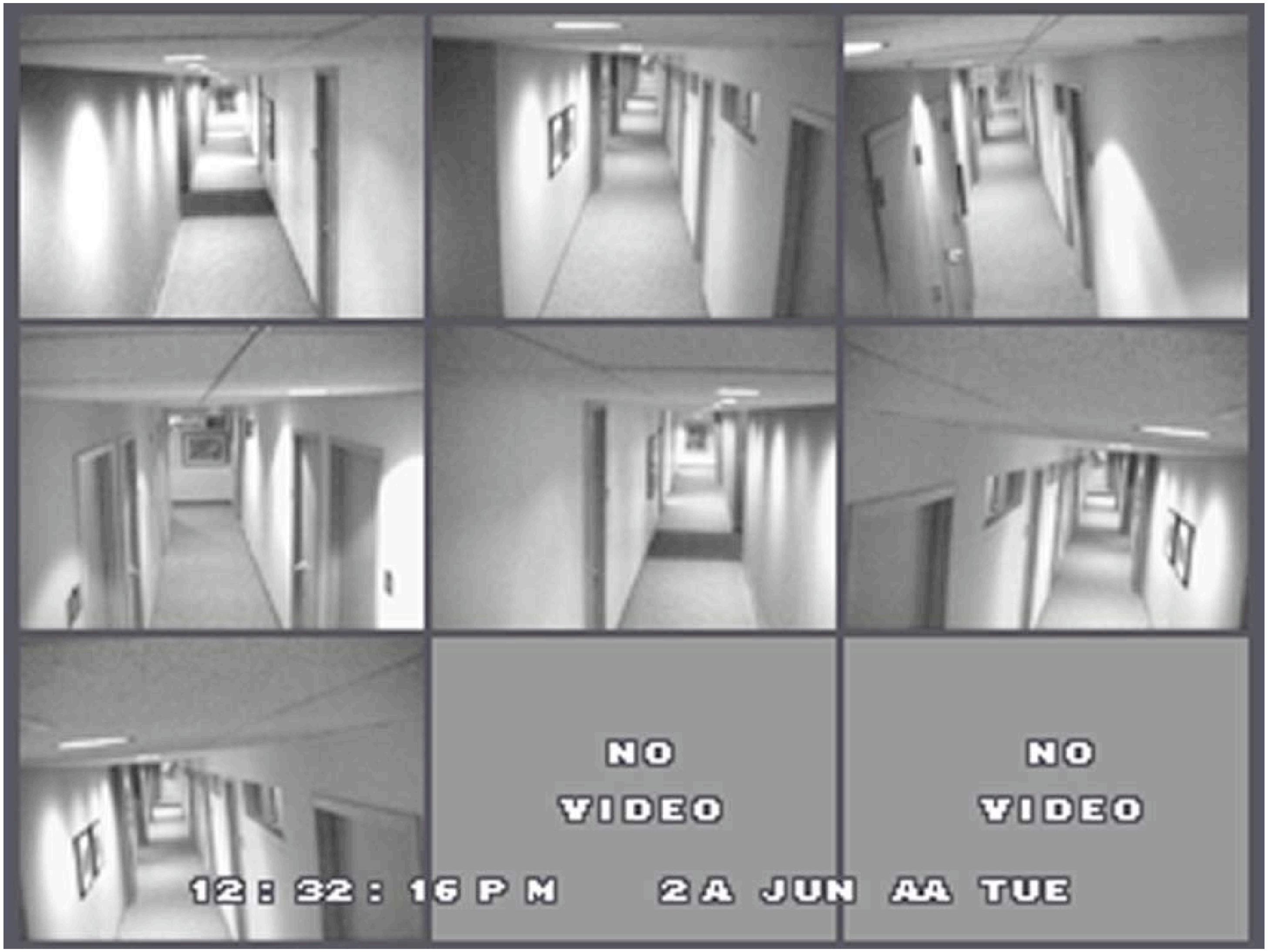
H+: Going Beyond the Human Scale

- Very large arrays:
  - – Localization for low-frequencies
  - – Localization for impulsive/wideband sounds
    - Silverman, Patterson, and Flanagan, “The Huge Microphone Array,” IEEE Concurrency, October, 1998.
    - Pregliasco and Martinez, “Gunshot Localization through Recorded Sound,” Journal of Forensic Science, 2002.

H+: Augmenting Ears

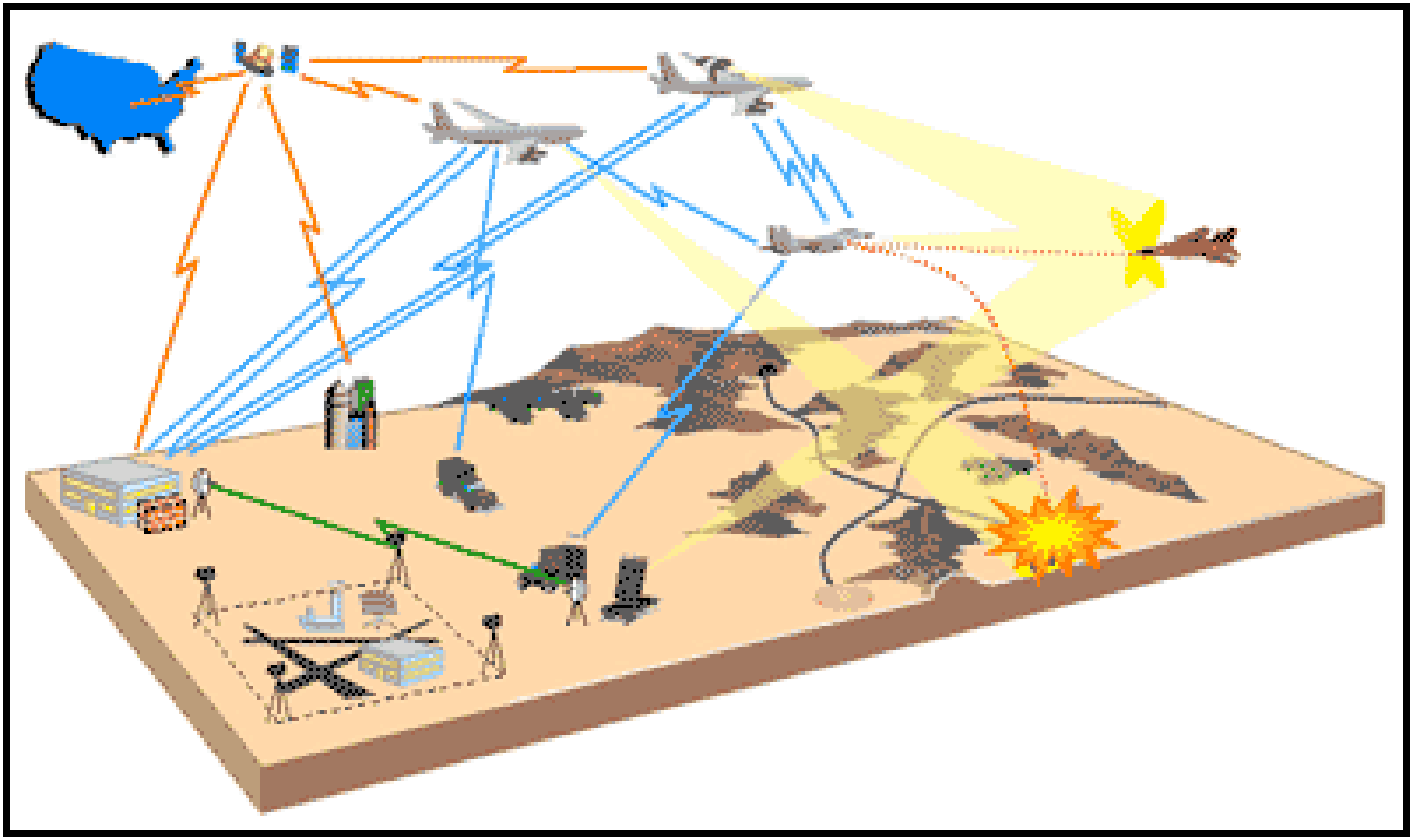
- The strength of numbers:
  - As a localizer or recognizer, machines may be at about half human performance
  - With 100 sensors => 50 humans worth!!
  - But what good is a fractional human?
- State of the Art in General Sound Recognition
  - Speech detection
    - Everybody and their Uncle Joe, “My Novel Method for Speech Detection,” 1960-2004.
  - Everything else

H+: Multiplying Ears



…because there may be too many things to listen to…

H+: Multiplying Ears



…too many sounds in too many places…

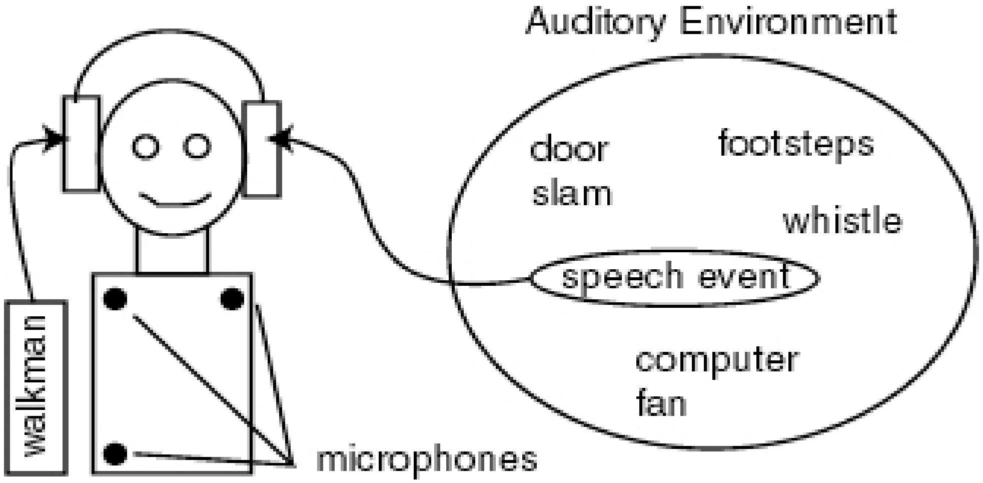
H+: Distant Ears

…because we can’t be everywhere at once…

H+: Replacing Ears

…because we may have limited hearing capabilities…

H+: Augmenting Ears



…because we're not always paying attention…

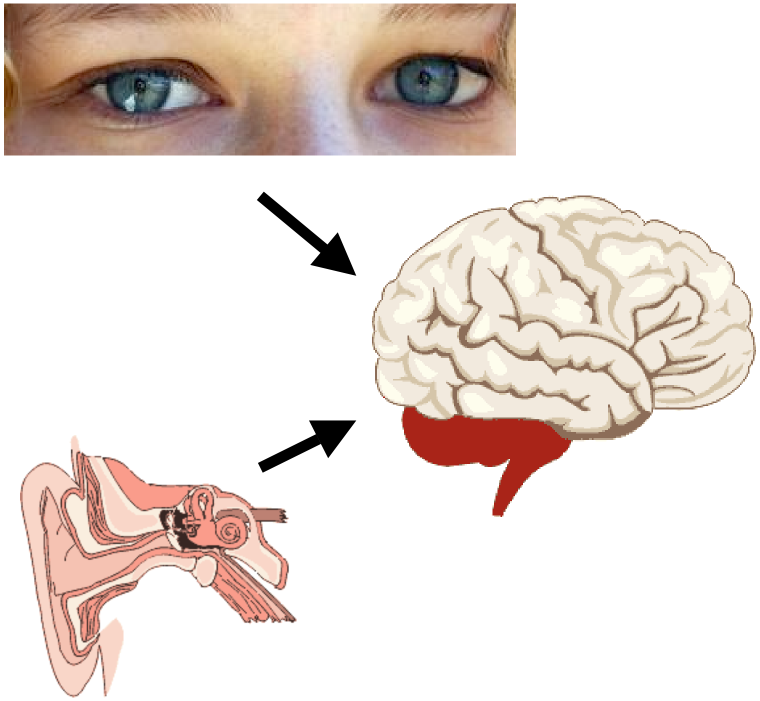
H+: The Sixth (Seventh, etc.) Sense

- We can apply existing techniques to frequency ranges/senses we don't have
  - Ultrasound
  - Microwave

Humans Enhancing Machine Performance (M+)

- Despite impressive machine computational capability, there are still certain tasks that the human can do faster and more reliably.

vs.



M+: What Do We Optimize?

- Finding the right objective function is hard
  - SNR vs. intelligibility
  - Listening comfort
  - Particularly true if a human will be listening to the output
- Example: Hearing Aids

Perseo™ BTEs – reliable beauty

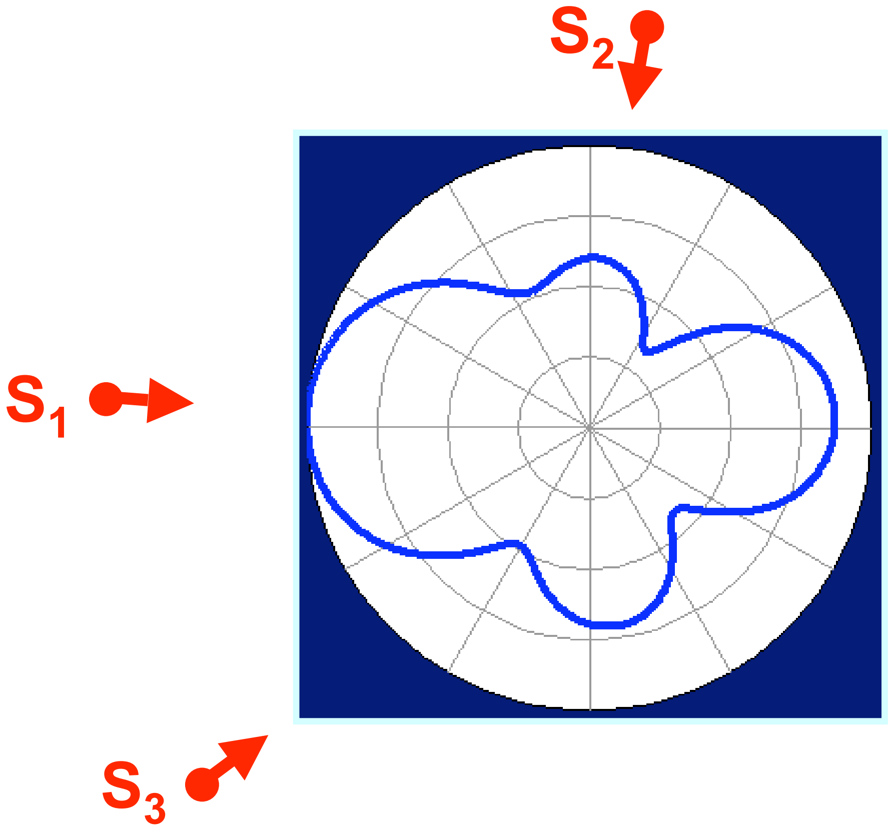
The Perseo BTE family includes a full choice of sizes, colors and power levels. All feature an elegant, award-winning design with a refined, ergonomic shape. They have a patented housing construction with bonded plastic and silicon to effectively keep out debris and moisture. Environmental protection extends further with patented, disposable microphone covers to ensure consistent, reliable performance, in all conditions. With dual-microphone systems this is particularly important since electronic matching cannot successfully compensate for performance reduction as a result of debris and moisture. A comprehensive range of colors for the housing also caters for specific tastes and trends.

Original size
New small Perseo BTE 111/211

(Phonak Persio)

M+: System focus

- Where are the sources?



M+: Environmental Conditions

- The human is often better at scene analysis
- Can drive system to optimize for varying conditions
  - Low Reverb? High Reverb?
  - Few, localized sources? Many sources?

M+: Calibration

- Some systems (e.g., conventional array processing) require knowledge of physical arrangement of microphones.
- Portable/body-mounted systems in particular must be configured and calibrated for proper operation.

Discussion and Teaser: Designing the Interactive System

- Input from the user:
  - How can we use direct manipulation and implicit manipulation to control the machine's abilities
- Output to the user
  - How do we decide what information is relevant to the user and how much they can handle?
  - How do we consolidate information into concise visuals/auralizations?
  - How can we display multiple auditory/visual streams to the user?