Designing Interactive Systems 2

Lecture 10: Multimedia & Multimodal Interfaces

Prof. Dr. Jan Borchers
Media Computing Group
RWTH Aachen University

hci.rwth-aachen.de/dis2

CHAPTER 32

Multimedia & Multimodality

using more
than one

Multimedia

text, images, music,
video, animation, …

using more
than one

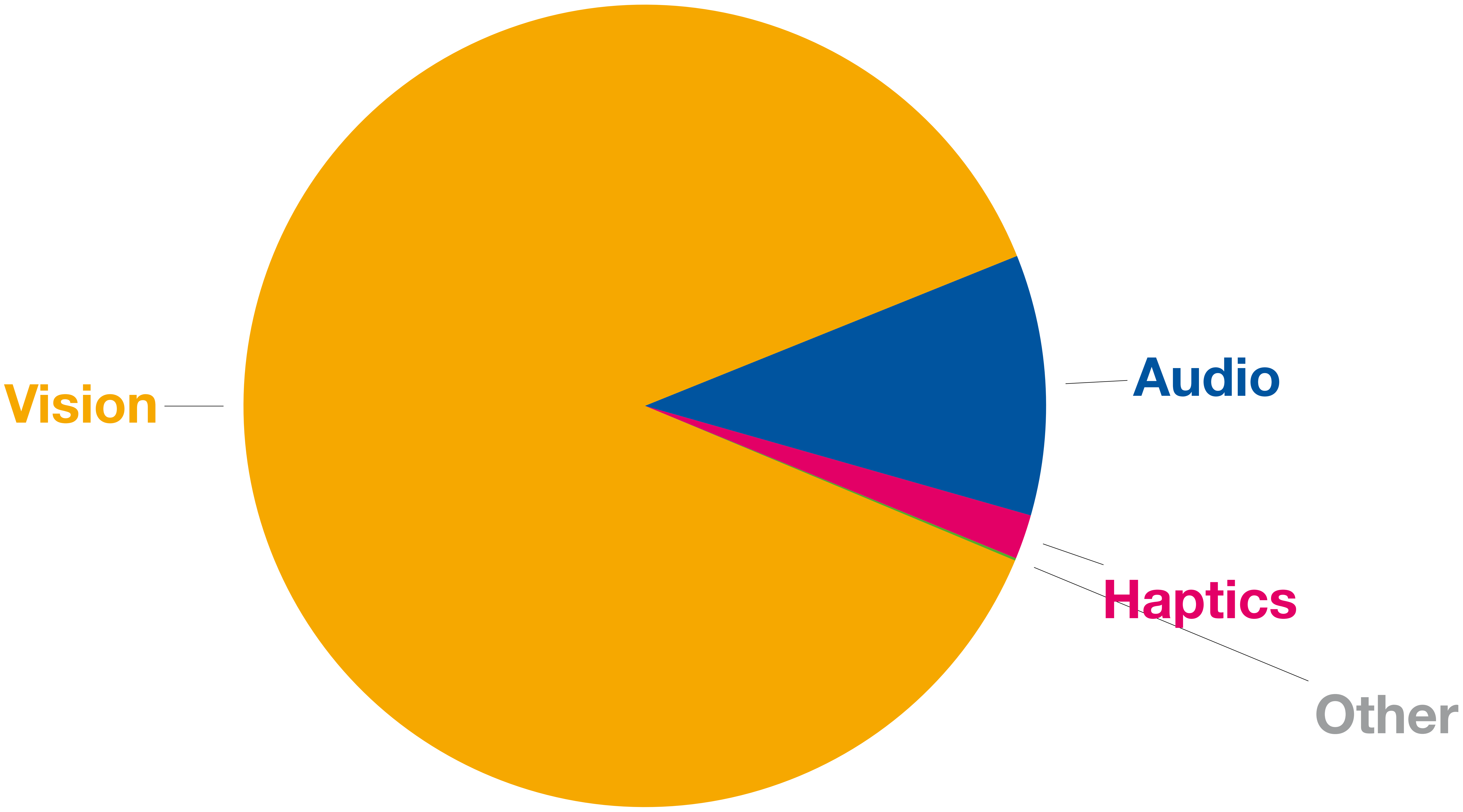
Multimodality

type of communication channel
used to convey or acquire
information

Multimodal interactions are **natural**.

Put That There

Usage of Modalities in Computer Systems



Advantages of Multimodal Systems

Input

- Fallback input techniques
  - Increased usability
  - Increased accessibility

Output

- Redundancy
- Synergy effects
- Increased bandwidth
- Realism

CHAPTER 33

Audio

Audio Output

- **Advantages**
  - Undirected
  - Needs no screen space
  - No visual contact necessary
  - Gets attention even when distracted
  - Reaches entire group
- **Disadvantages**
  - Undirected
  - Attention-grabbing
  - Annoying user and others
  - Transient
  - Dictates speed
  - Cannot overlap easily

Audio Output: Types

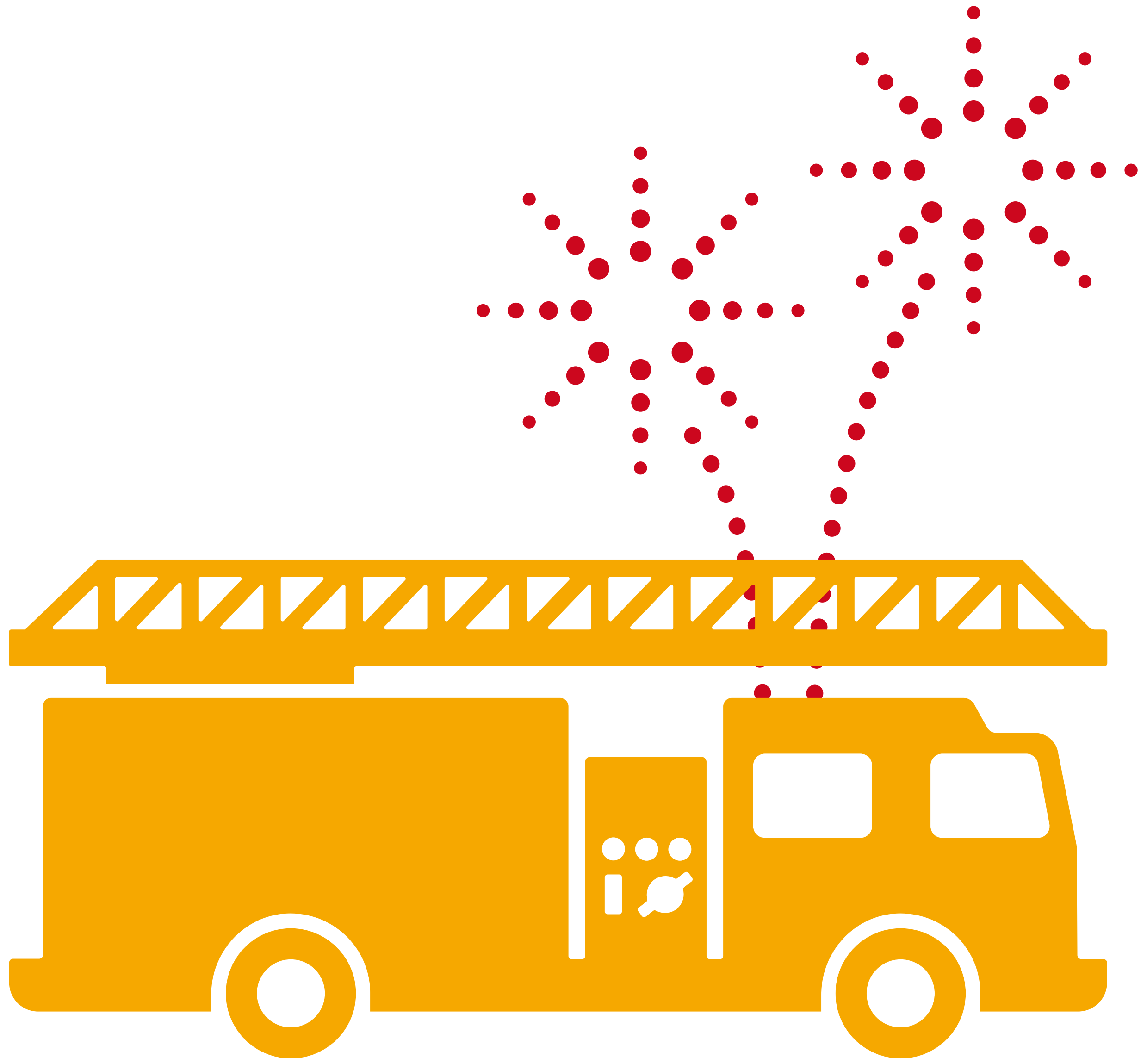
Noise	Volume, duration
Beep	Volume, pitch, duration
Melody	Volumes (emphasis), pitches, durations, sequence
Chords	Melody, plus harmony (minor/major, dissonance)
Speech	Textual contents

Two Ways of Listening

Musical Listening

Perceiving Notes

Two Ways of Listening



Musical Listening

Perceiving Notes

Everyday Listening

Perceiving Events

Audio Output: Noise

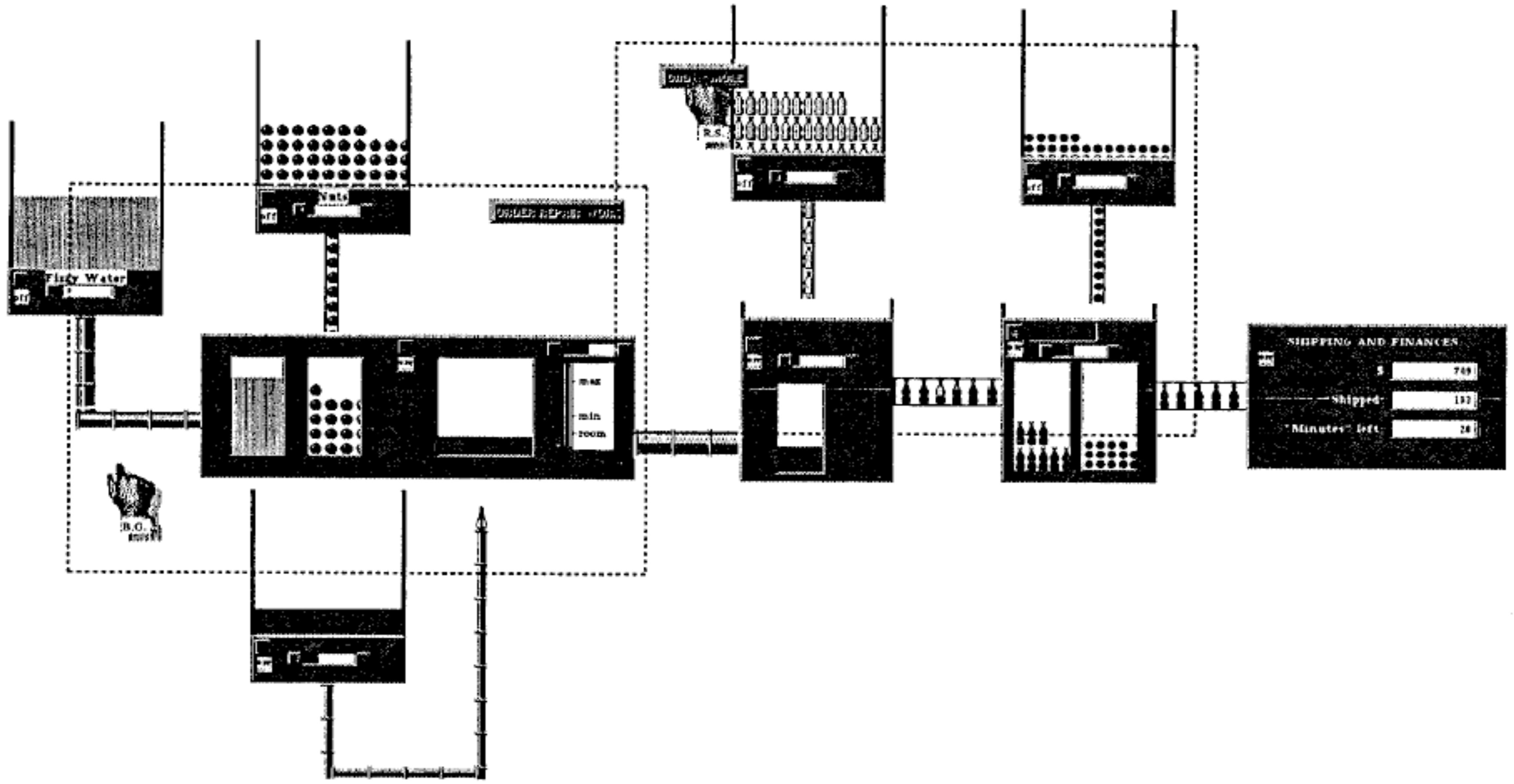
- Only audio output that may be "natural"
- Also used artificially in applications: **Auditory Icons**
  - Play everyday sounds along with actions of the system
  - Particularly useful in complex situations (too much to watch)
  - But: not everything maps to a sound

Audio Output: Noise: The SonicFinder (1989)

- Prototype from 1989 that enhanced the Mac's Finder
- Auditory icons for desktop objects and operations
  - Files wooden, folders papery, applications metallic, bigger = deeper sound
  - Throw into wastebasket: smashing
  - Copying: a problem!
- Filling jug with water, rising pitch = progress
- Mostly satisfied users, but natural analogies are limited
- Also used in experimental Mosaic web browser and commercial products
  - Since Mac System 7 Finder, MS Office,…

Audio Output: Noise: ARKola

- Simulation of Coke bottling plant
- 2 remote users collaborating
- Noises reflect status (bottles clink when released, liquid splashes if wasted, bottles break if wasted)
- Result: improved collaboration

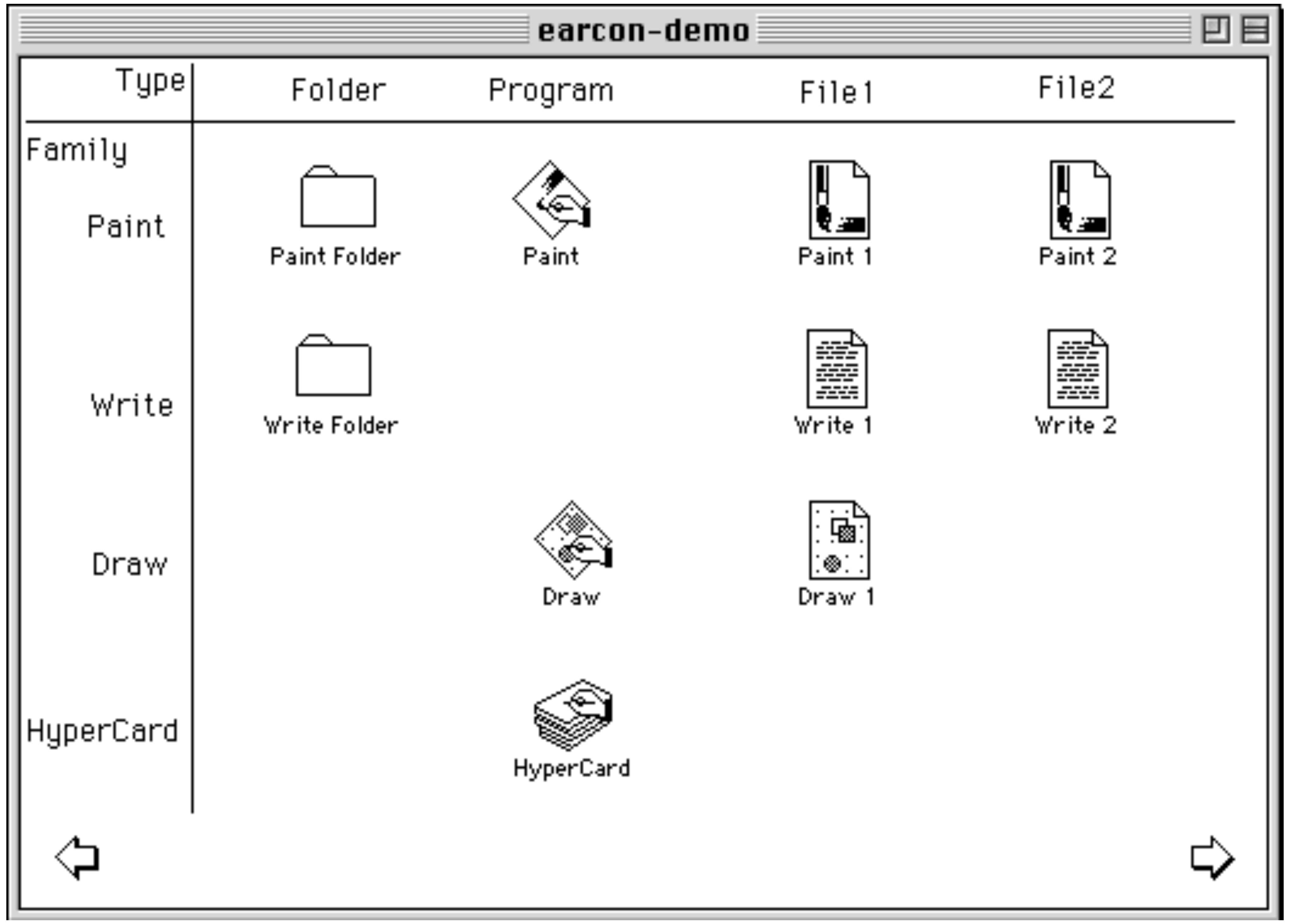


Audio Output: Melodies

- **Earcons** (Blattner 1989)
  - Abstract, synthetic tones, non-verbal
  - Building block: motif
    - Short, rhythmic pitch sequence with variable intensity, timbre and register
  - Combine into complex messages

Audio Output: Melodies: Earcons

- Use instrument timbres for distinction
- Pitch & register don't work well on their own, unless with large differences, use pitches 150Hz–5kHz
- Vary rhythms greatly, do not use notes shorter than 8ths, vary number of notes. Can parallelize.
- Intensity: Bad for distinction. Keep in small range (10–20db over background noise), user may adjust volume. Use localization.
- Join motifs with .1s gaps



Audio Input: Noise

- Use raw audio data to trigger events or control system
- Examples
  - Clap to turn on an interactive room
  - Monitor noise level to switch between remote sites displayed in multi-party video conferences
  - Localize current speaker to pan video conferencing camera
  - Monitor noise level to adjust phone ringer volume
- Fairly simple, used in commercial applications (and devices) today

Audio Input: Melodies

- **Data**: Musical Input
  - Record music
  - Synthetic (MIDI) or real (audio) data
- **Search**: Query By Humming (QbH)
  - Given a database of music, let users hum a melody to find the corresponding piece in the database
  - Difficult algorithmic problem

CHAPTER 34

Speech

Speech Output

- Also known as Text-to-speech (TTS) systems
- Using recorded chunks of different size
- 3 Levels of quality:
  Phonemes (flexible, small database) < words < sentences (more realistic)

Speech Output

- **Advantages**
  - Natural, familiar, emotional
  - For the visually impaired
  - Eyes-free
- **Disadvantages**
  - Slower than visual (bandwidth)
  - Transient/ephemeral
  - Hard to browse/search
  - Synthetic missing "prosody" (melody,…)
  - Unlike noise, cannot fade into background (hard to ignore)

Speech in Humans

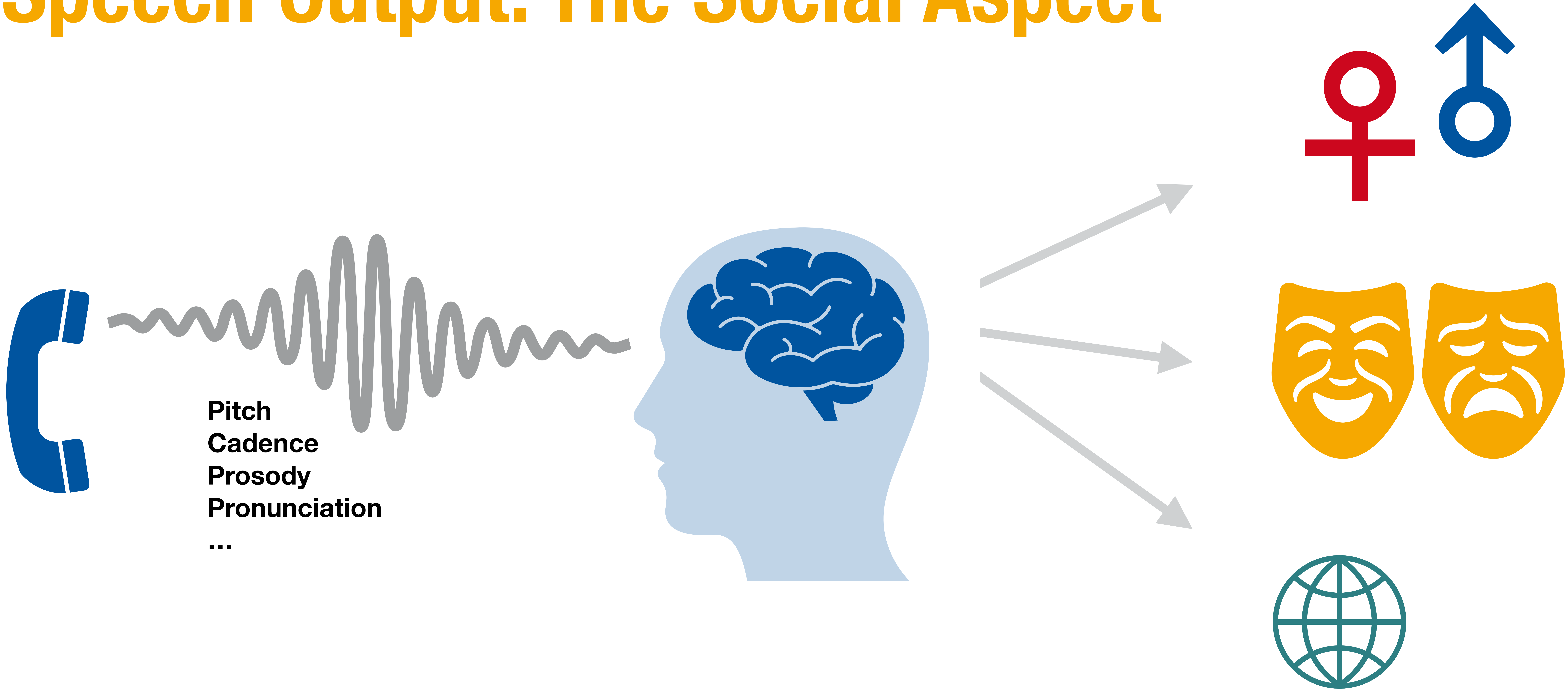
The primary way of human communication

IQ score as low as 50 is sufficient to speak

Speech implicates more parts of the brain then other functions

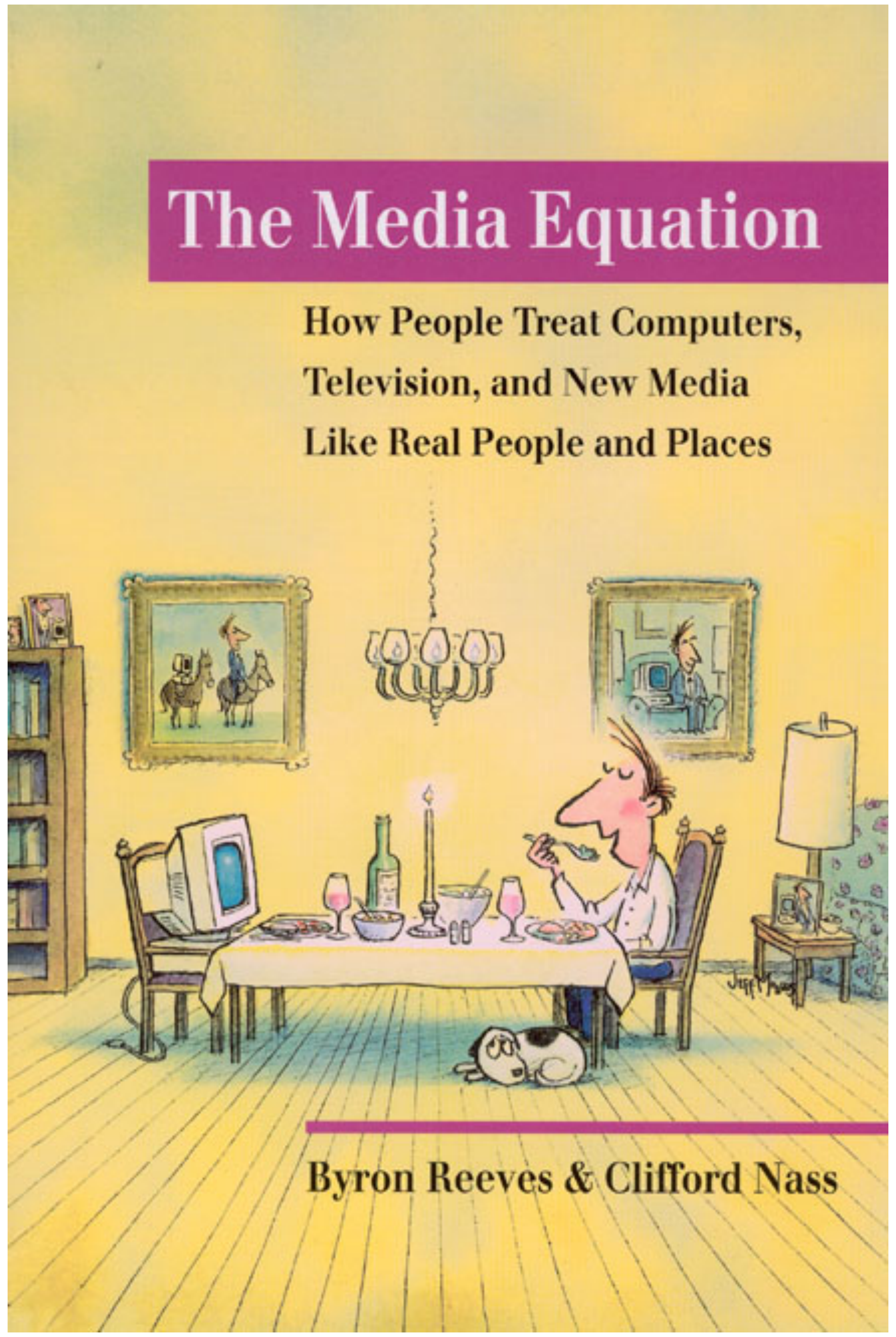
Left brain hemisphere is specialized to process speech

Speech Output: The Social Aspect



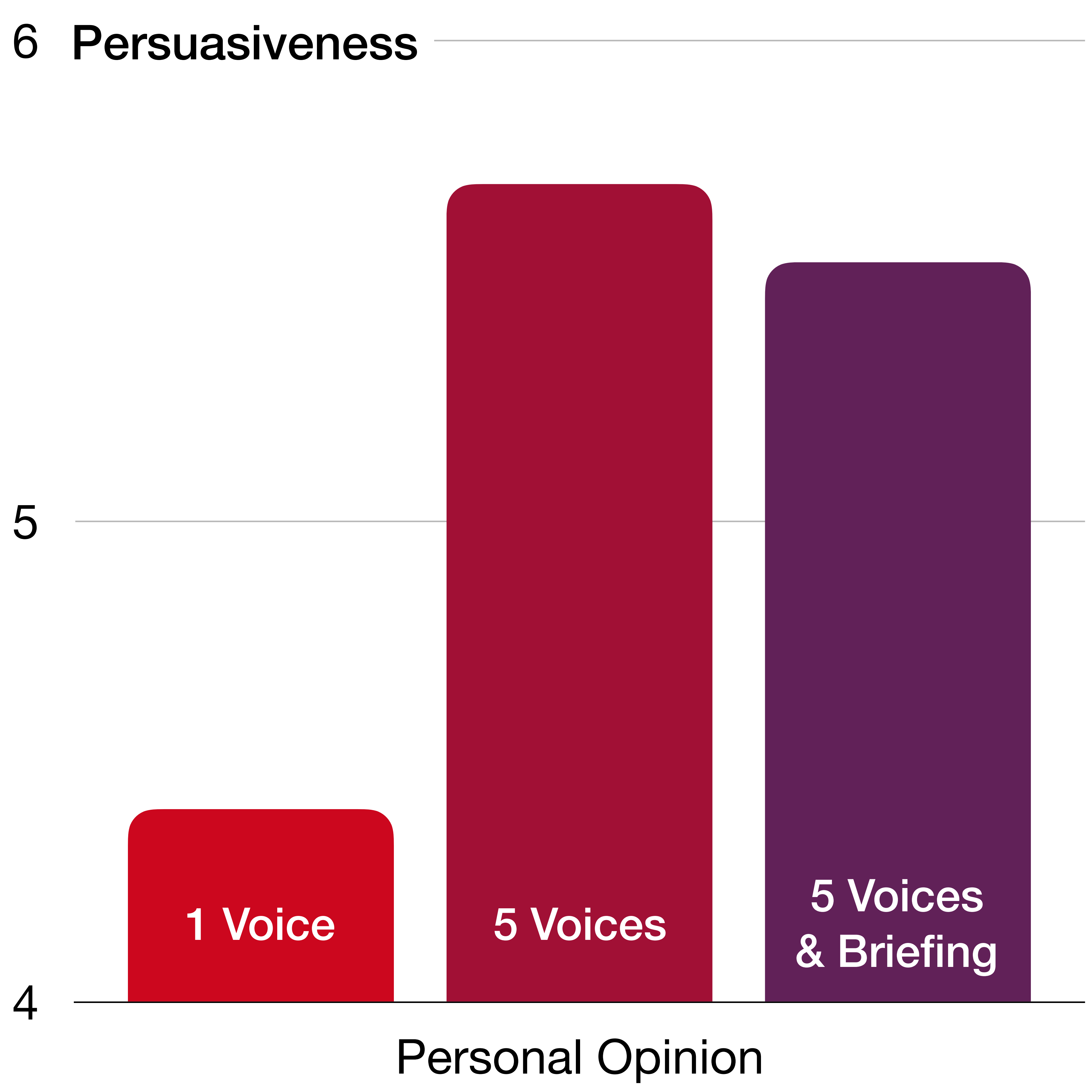
Speech: New Media, Same Old Rules

- Regardless whether a voice originated from a human speech or TTS system, the same areas in the brain are activated
- The same rules are applied to make conclusions about the speaker
- Opportunity for HCI to create voice interfaces that increase liking, trust, efficiency, …
- In the following examples form “The Media Equation” by Clifford Nass showing that computers are social actors



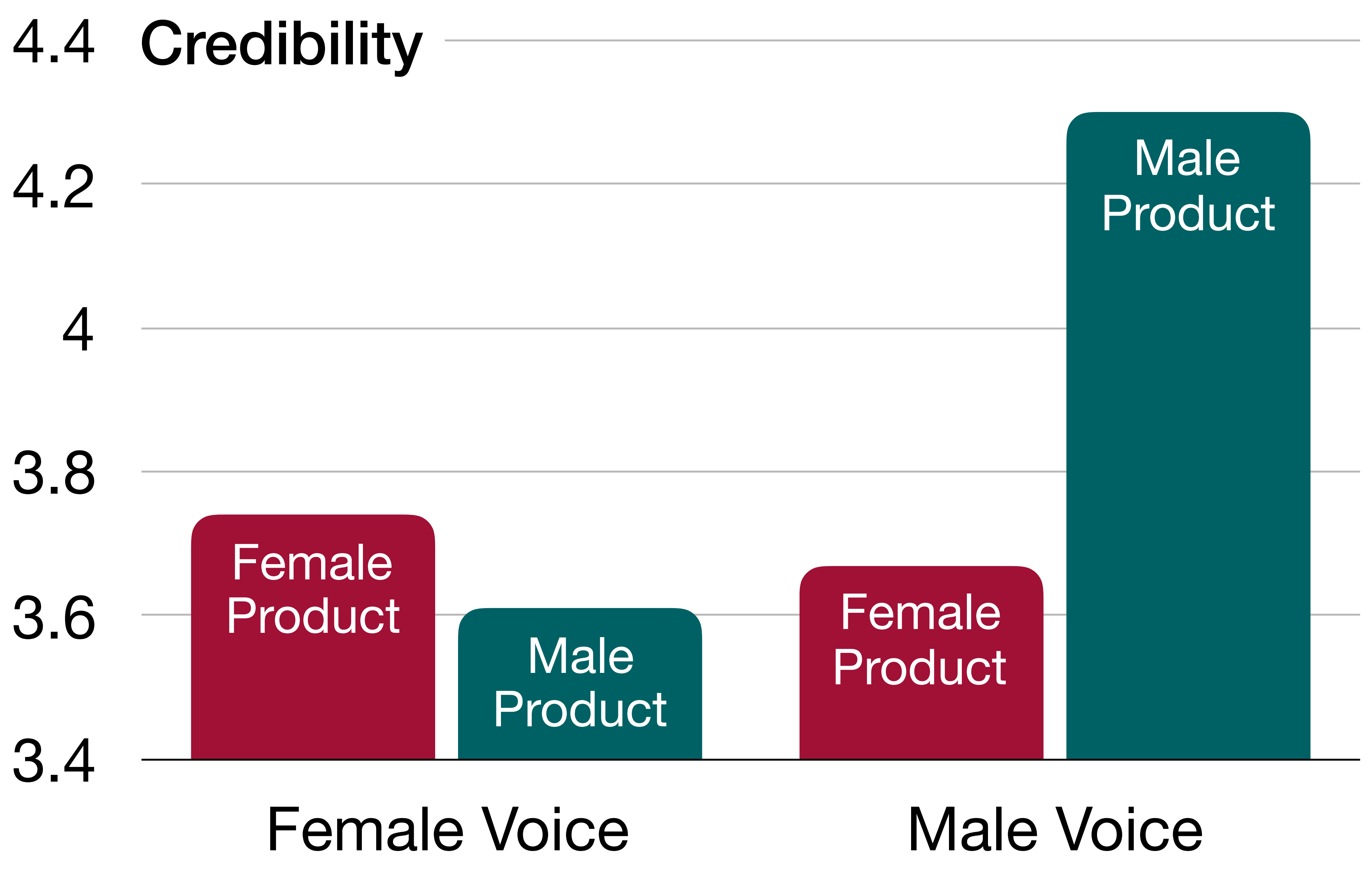
The Media Equation: Persuasiveness

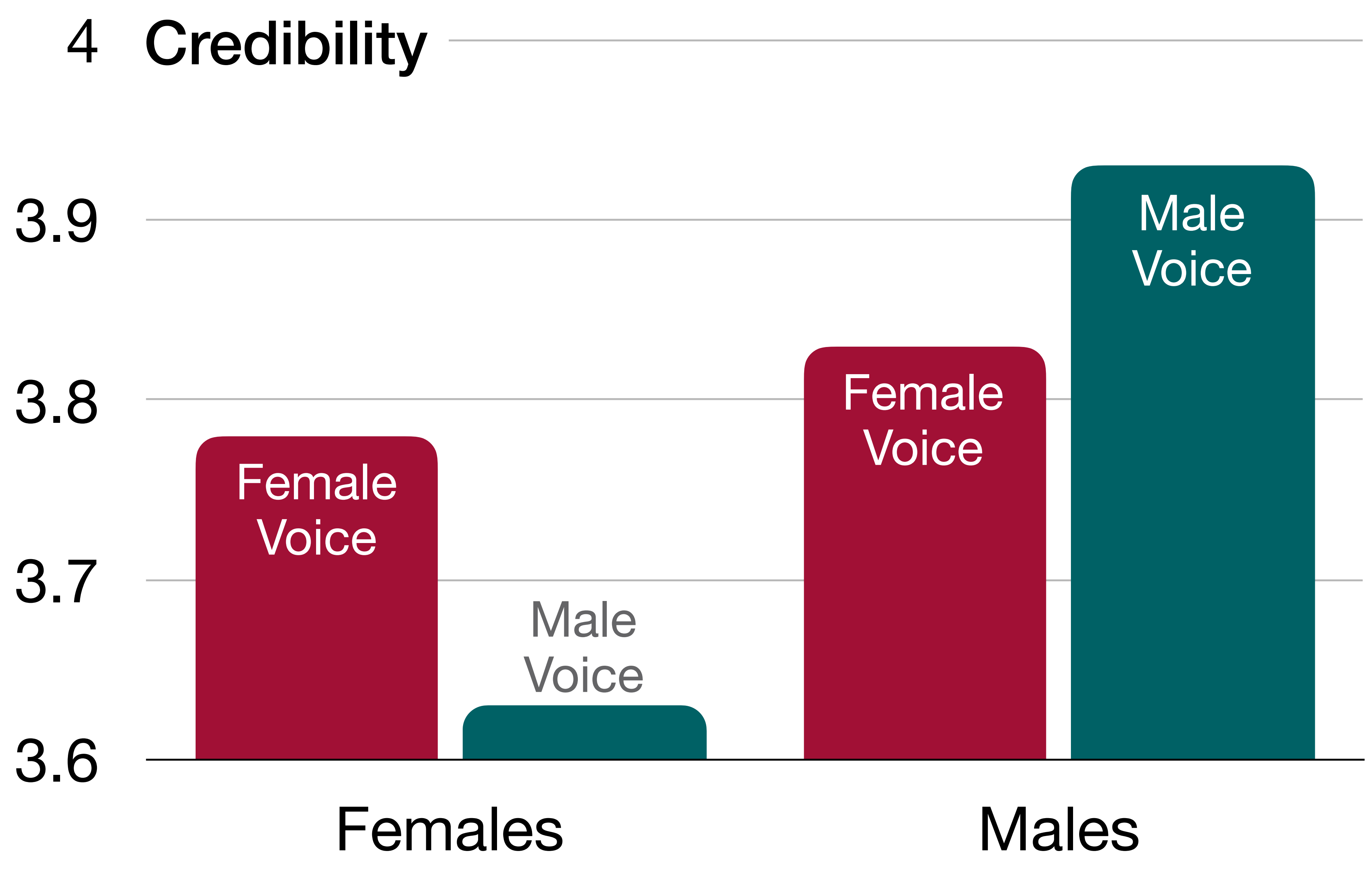
- Sample Experiment:
  Five positive Amazon customer recommendations read
  by one vs. several voices
- Multiple voices more persuasive
- Does not work with multiple fonts



The Media Equation: Credibility

- Product reviews are most credible, when gender of product, voice, and user match





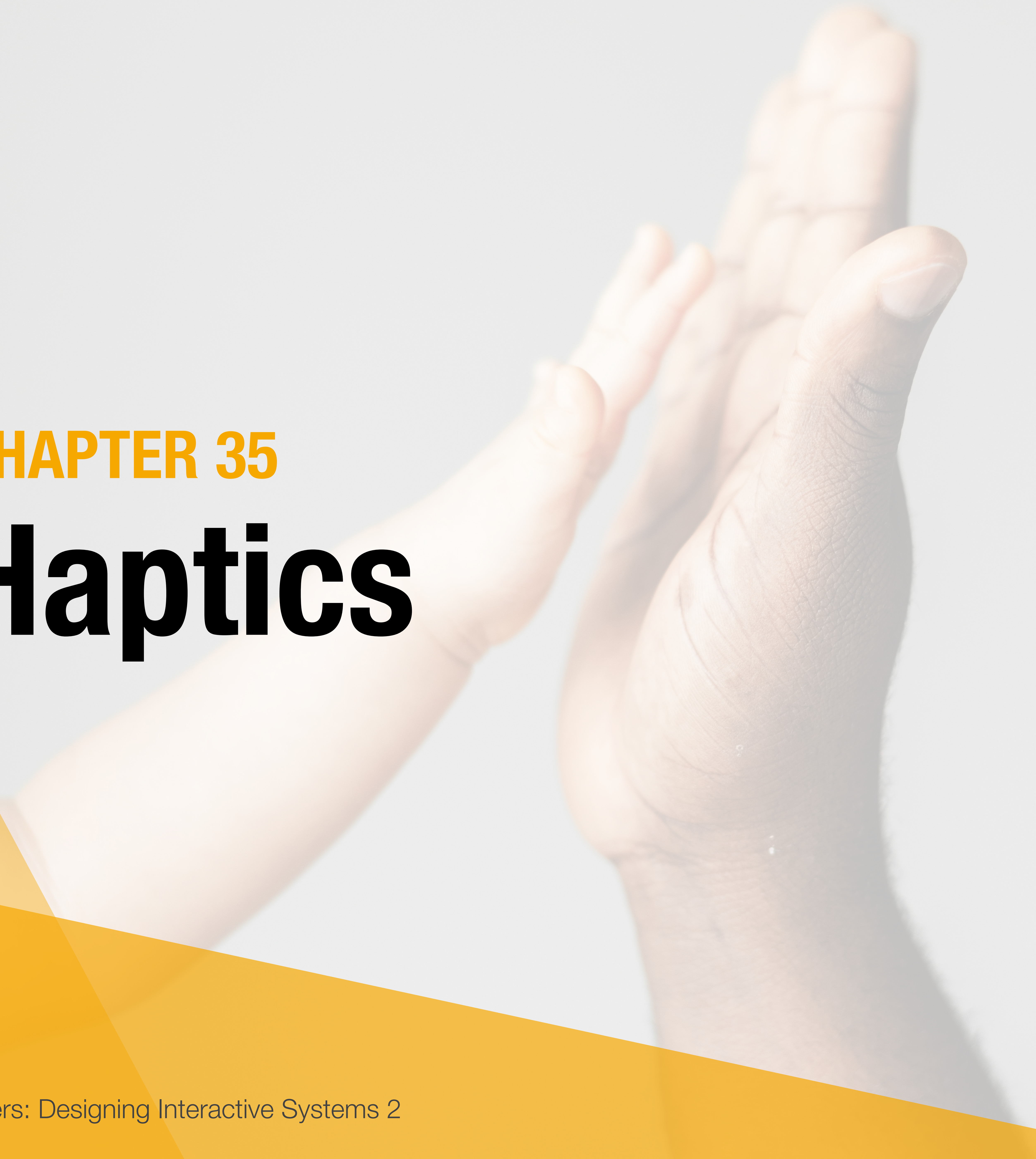
Speech Input: Problems and Challenges

- Speaker dependency
(accent, intonation, stress, volume,…)
- Vocabulary dependency
- Background noise very critical
- Detection precision is high, but getting the semantics out of a sentence still an issue
- Many syntax combinations for same semantics

- Continuous speech
    - Humans disambiguate blurred word boundaries and multiple semantics
    - "How to wreck a nice beach"
- Higher cognitive load
    - Hand-eye coordination can happen in parallel to planning and problem solving
    - Speaking while thinking is more difficult

CHAPTER 35

Haptics




JACK CRV
15
CHROME VANADIUM
18

What makes these feel different?

- Hardness
- Height maps
- Temperature
- Damping
- Friction
- Flexibility
- Spatial / temporal features

Touch is Special

- Bidirectional
- Socially intentional-committing, invasive
- Gestural-expressive (functional and emotional signals)
- Many parameters: force, texture, temperature, moisture,…
- Poor absolute but high relative resolution
- Touch to do, probe, poke, fidget, communicate, verify, enjoy, connect,…
- Inhibitions: dirty, painful, forbidden, too intimate,…

Main Types of Haptic Interfaces

- **Cutaneous stimuli**
  - On the skin, i.e. tactile
  - E.g., heat, pressure, vibration, slip, pain
- **Kinesthetic stimuli**
  - Bodily movements
  - Detected in muscles, tendons, and joints
  - E.g., limb position/motion/force

Cutaneous Feedback

Kinesthetic Feedback

RWTHAACHEN UNIVERSITY

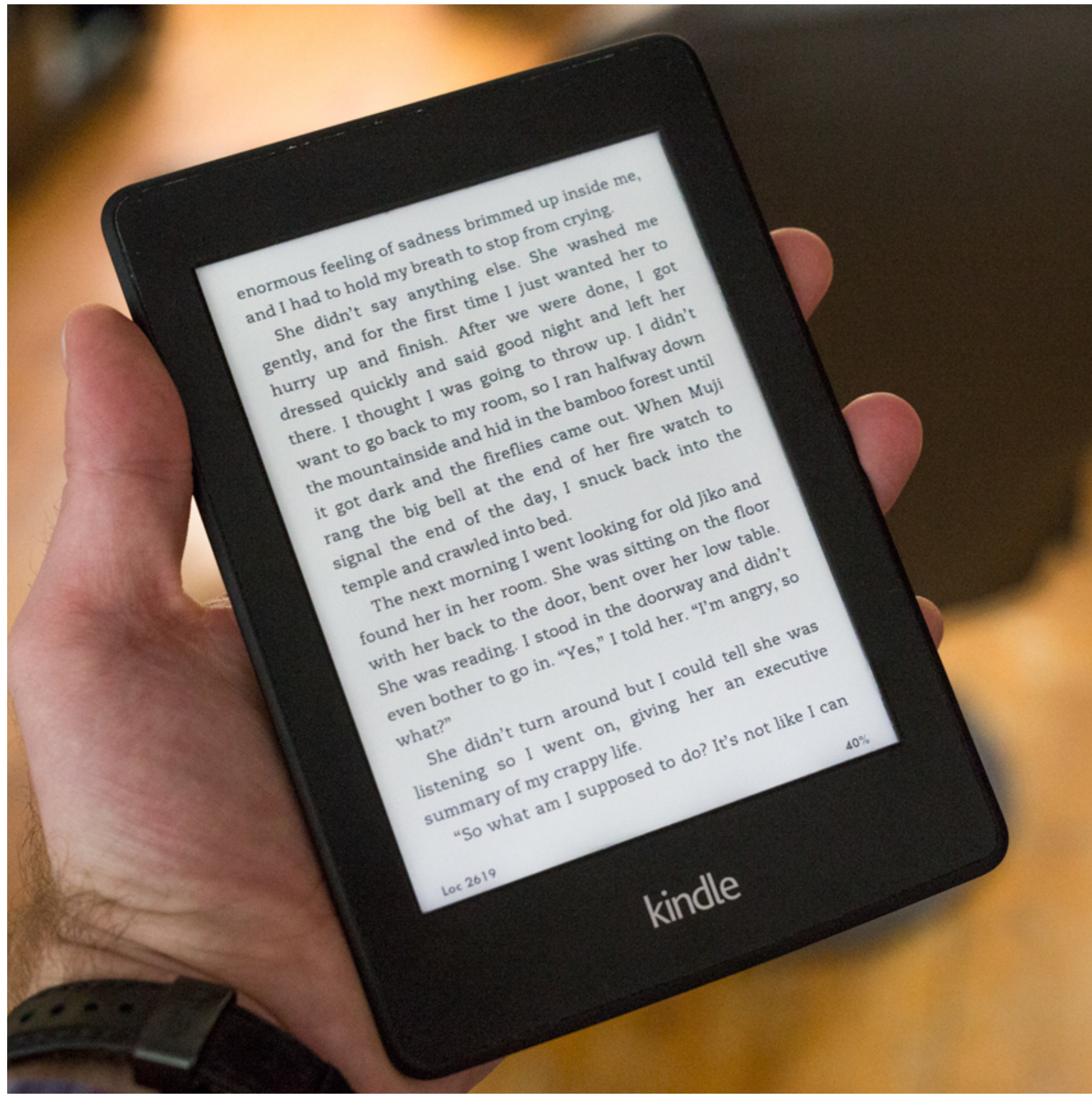
Haptic Output

- **Advantages**
  - Realistic
  - Intimate
  - Eyes-free
  - Needs no screen space
- **Disadvantages**
  - Limited resolution
  - Intimate
  - Unexpected

CHAPTER 36

Vision

Visual Interfaces



Visual Input

- What can the computer see?
- Brightness
- Color
- Edges
- Shapes
- Objects

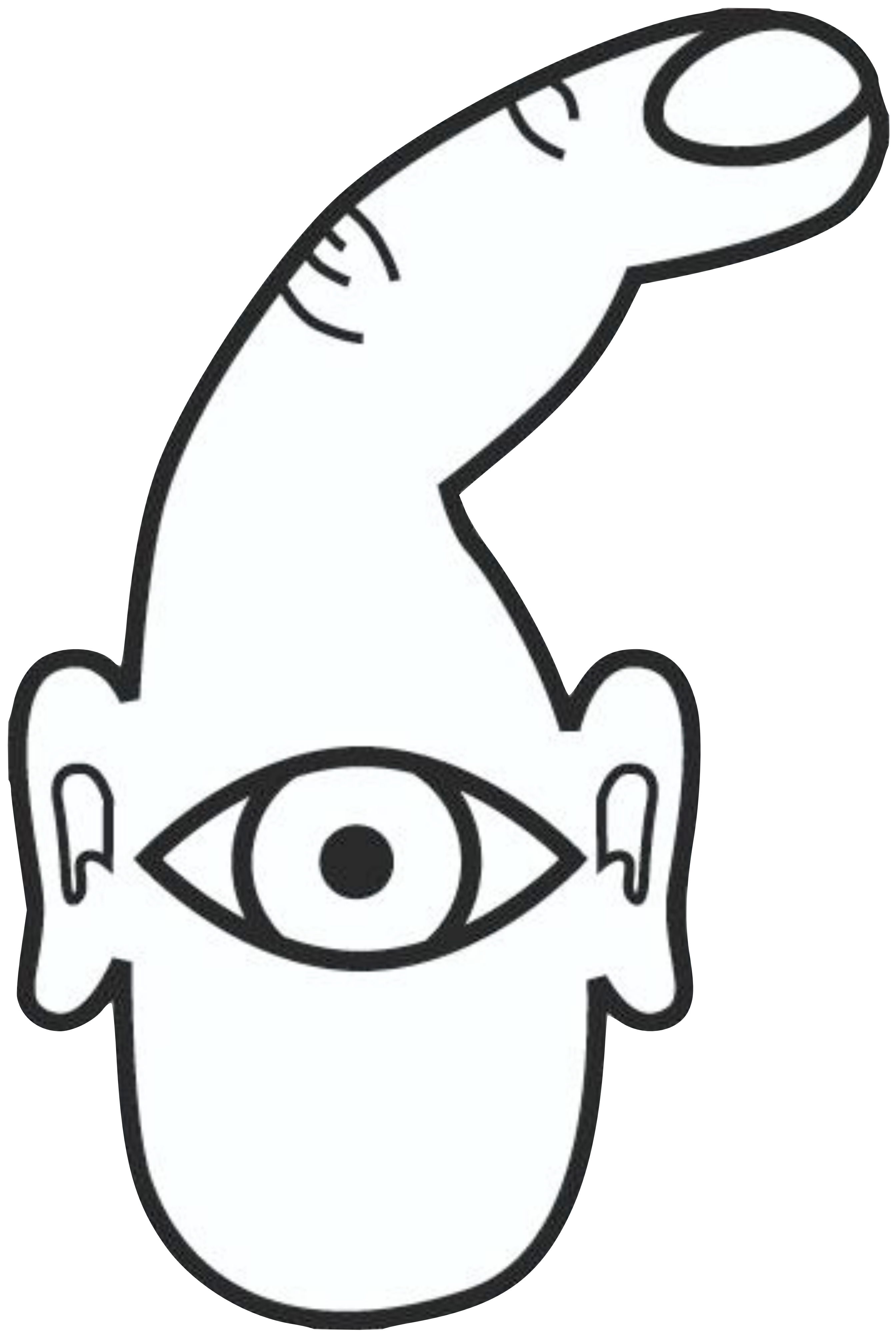
Visual Input: Processing Pipeline

- Capture frame
- Apply basic filtering to enhance image
- Look for regions / objects in scene
- Classify, match with world model
- React as result of observed state change in model world

CHAPTER 37

Creating Multimodal Interaction

How does the computer see us?



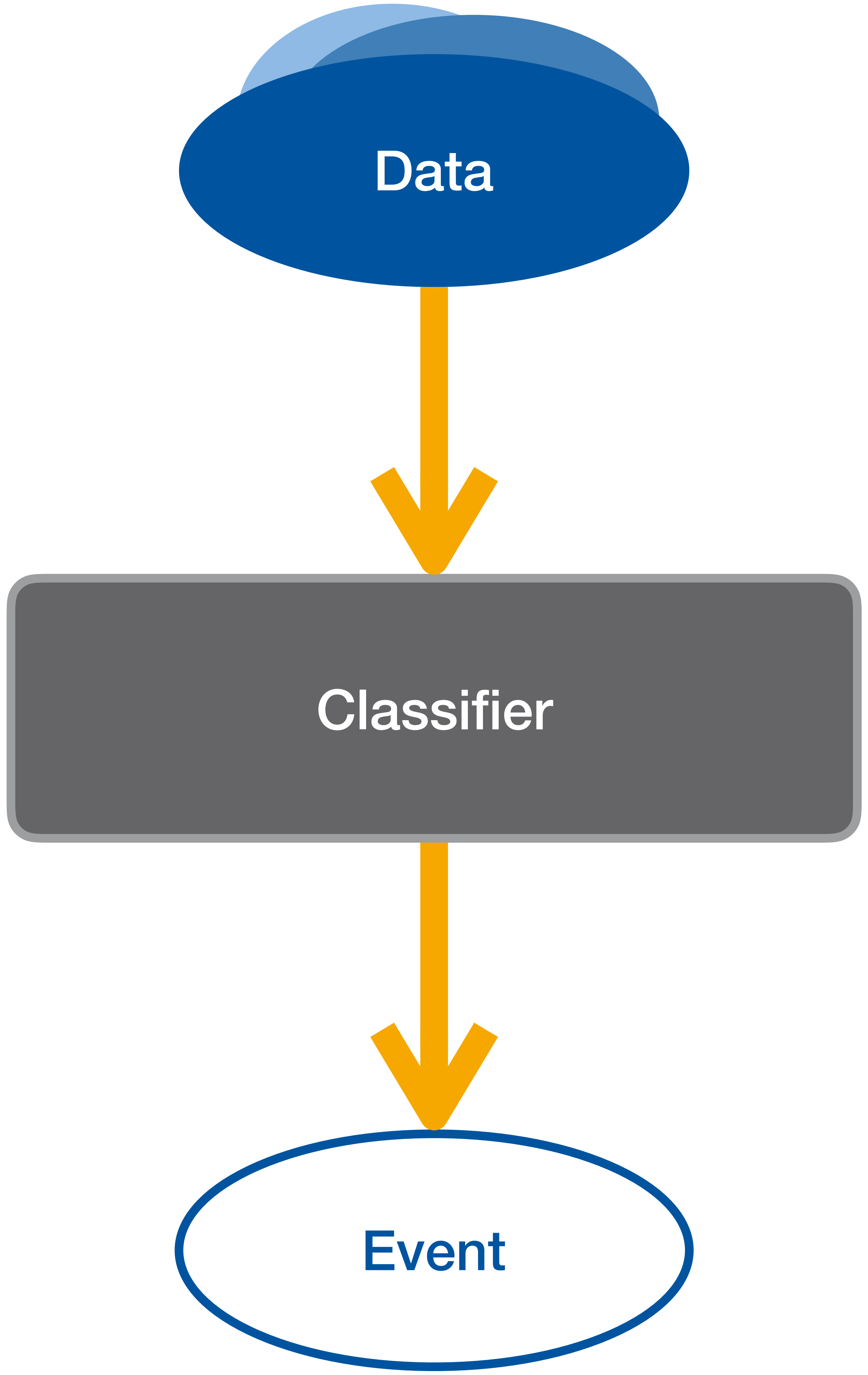
Design Space of Multimodality

	Visual		Auditory		Haptic	
	Input	Output	Input	Output	Input	Output
Control						
Data						

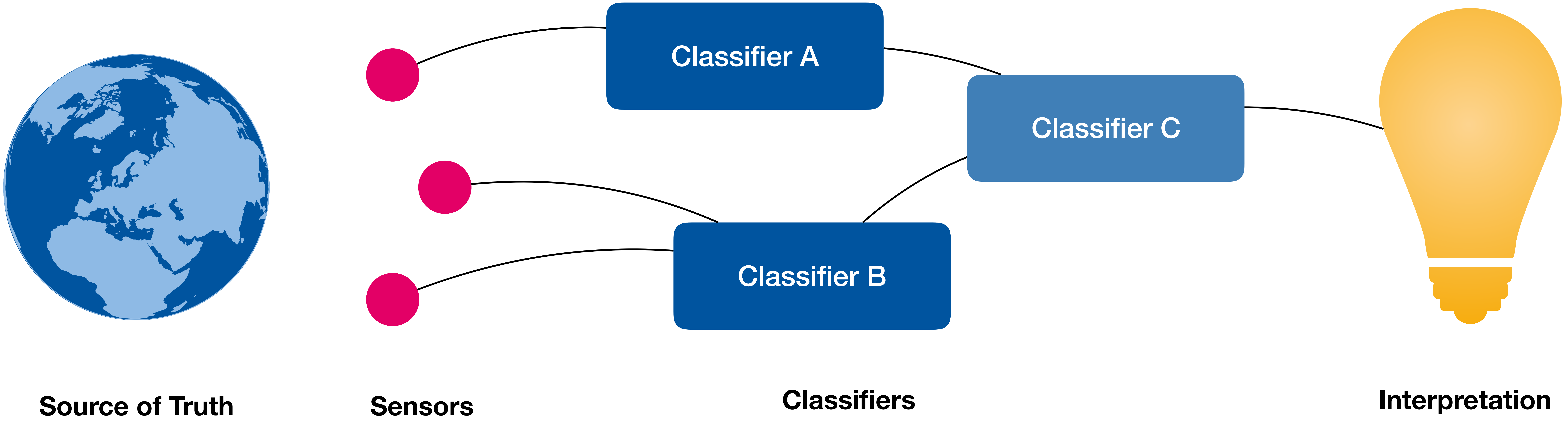
How to **process** inputs from multiple modalities?

Issues During Interpretation

- For a single sensor
  - Noise (sensor, channel, modality-specific)
  - Non-universality
- For multiple sensors
  - Ambiguity due to contradicting information
  - Different formats and sampling rates



Fusion



Classifier A
Classifier C
Classifier B
Source of Truth
Sensors
Classifiers
Interpretation

Example: Support Vector Machines