Perceiving Graphical and Pictorial Information via Hearing and Touch

IEEE Signal Processing Society, Santa Clara

May 21, 2015

Thrasos Pappas
EECS, Northwestern University
On Sabbatical Leave at LLNL

Lawrence Livermore National Laboratory

LLNL-PRES-669870

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344. Lawrence Livermore National Security, LLC

People

- Pubudu Madhawa Silva, Northwestern Univ. – now at Intel
- Andrew Seward, Northwestern Univ.
- Dzung Nguyen, Northwestern Univ.

- Joshua Atkins, Johns Hopkins Univ. – now at Beats by Dr. Dre
- James E. West, Johns Hopkins Univ.

- William M. Hartmann, Michigan State Univ.

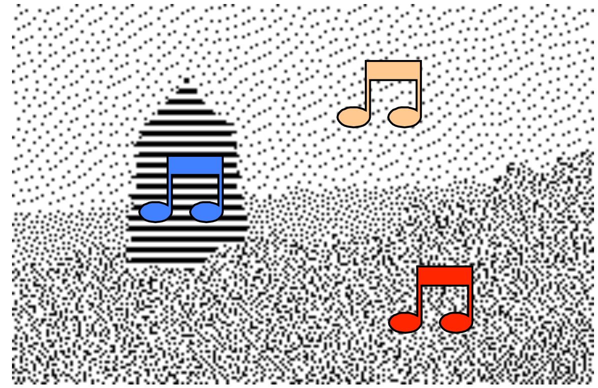
- James D. Johnston, independent consultant – formerly Bell Labs

- Karen Gourgey, Baruch College, CUNY
- Ilona Kretzschmar, CCNY, CUNY
- Vivien Tartter, CCNY, CUNY

Visual to Tactile-Acoustic Mapping


Original Image


Semantic Segmentation

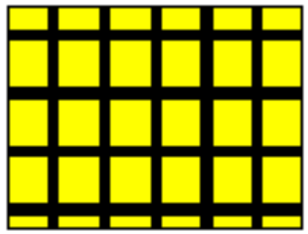

Acoustic-Tactile Layout

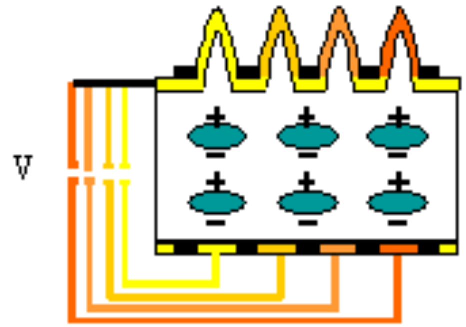
- Present a picture as a collection of segments with perceptually distinct tactile-acoustic textures
  - Tactile texture and sound (plus vibrations, variable friction, etc.)
- Active exploration with the finger or a pointer
  - Kinesthetic feedback
- Haptic space and scene perception
  - Advantages over line drawings

Other Approaches

- Invasive prosthesis
- Tongue display [Bach-y-Rita *et al.,* 2002]
- "vOICe" [Meijer, 1992]
  - 64x64 image of 16 graylevels mapped to tones
  - Vertical: frequency; horizontal: time and stereo panning; loudness ~ brightness
- "SoundView" [Doel *et al.*, 2003]
  - Tablet with pointer, colors mapped to sounds
- Raised line drawings [Wijntjes *et al.,* 2008]
- NOMAD [Parkes, 1988]
- Talking Tactile Maps [1994]
- Talking Tactile Tablet [Landau & Wells '03]
- Halftoning [Nayak & Barner '04]
- Dynamic variable friction displays
  - Tesla touch [Xu '11; Israr '12]
  - On Glass [Winfield '07; Chubb '09; Marchuk '10; …]
- Audio tactile maps [Jacobson '98; Parkes '88; Blenkhorn '94; Landau, 2003; Parente '03]

Motivation: Dynamic Tactile Tablet





- "Dynamic Tactile Interface for Visually Impaired and Blind People"
  - I. Kretzschmar, K. Gourney, V. Tartter, L. Abts, J. West, T. Pappas
- Three layers
  - Top: deformable electroactive polymer film
  - Middle: electrodes to address positions on the film
  - Bottom: touch sensitive screen
- Display dynamic tactile patterns (fast dynamic response)
- Fully addressable
- Detect finger position
- Audio feedback

Braille

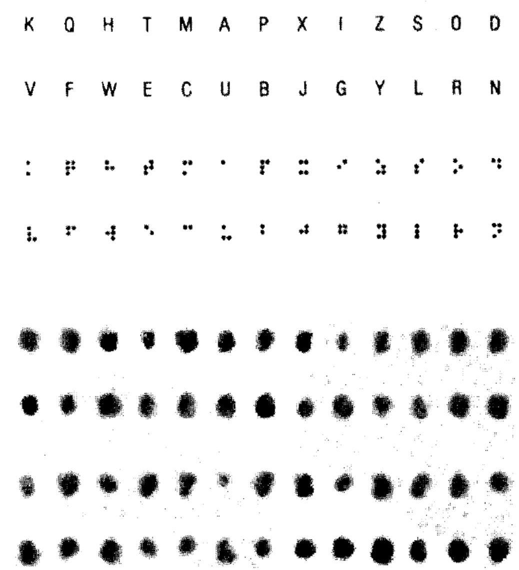
Dot diameter: 1.3 mm
Dot spacing: 2.5 mm
Dot Height: 0.5 mm [NLS 2005]

Letters:

a ⠁	b ⠃	c ⠉	d ⠙	e ⠑	f ⠋	g ⠛	h ⠓	i ⠊	j ⠚
k ⠅	l ⠇	m ⠍	n ⠝	o ⠕	p ⠏	q ⠟	r ⠗	s ⠎	t ⠞
u ⠥	v ⠧	w ⠺	x ⠭	y ⠽	z ⠵				

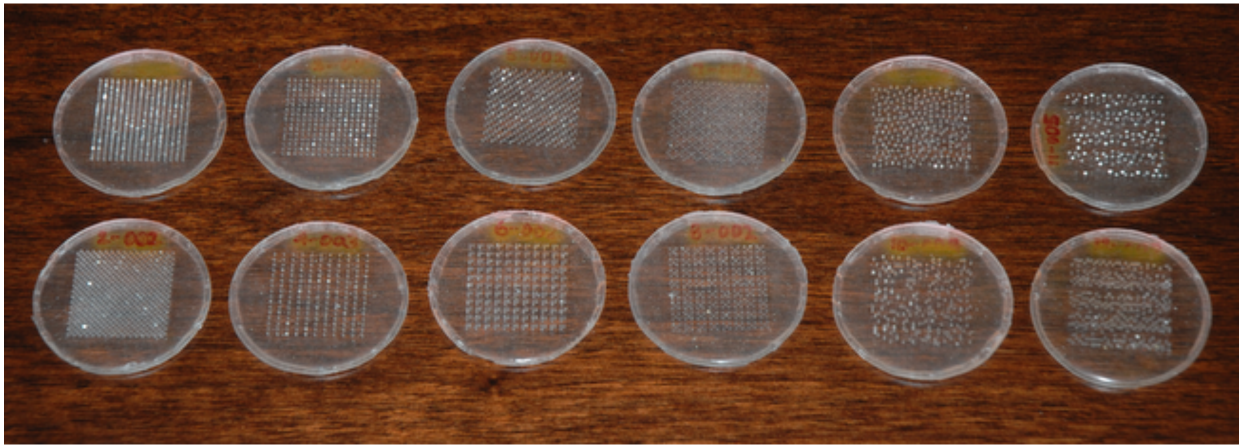
Numbers:

1 ⠼⠁	2 ⠼⠃	3 ⠼⠉	4 ⠼⠙	5 ⠼⠑	6 ⠼⠋	7 ⠼⠛	8 ⠼⠓	9 ⠼⠊	0 ⠼⠚



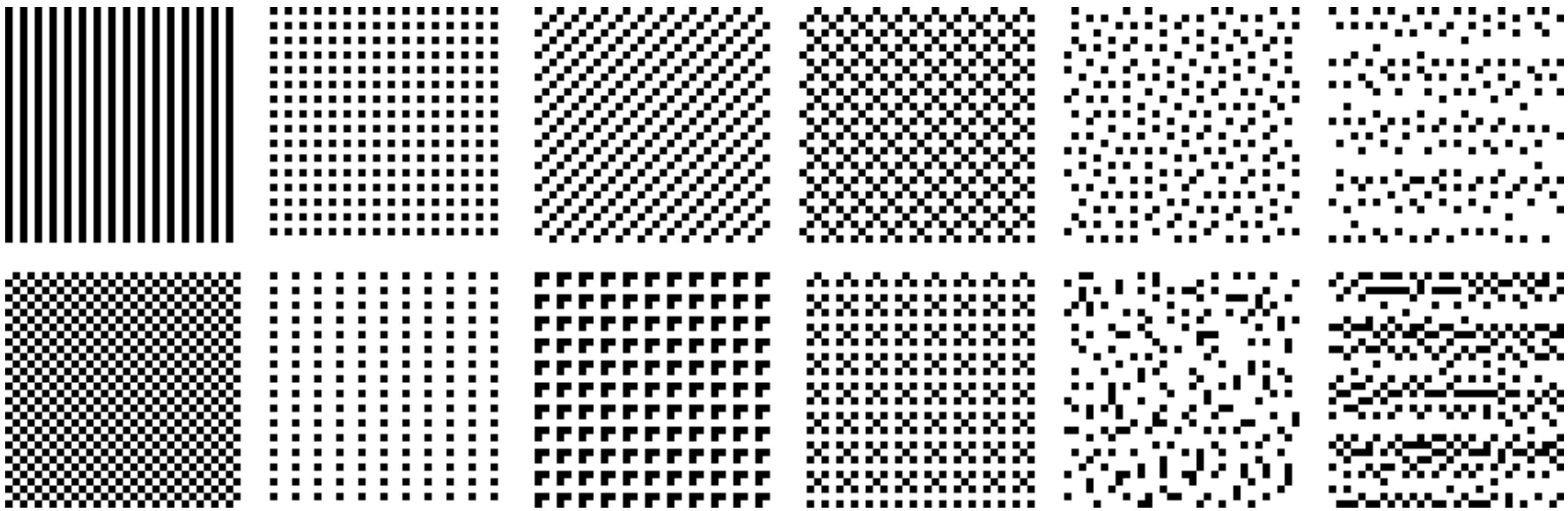
[Loomis 1981]

Device Model



- Simulate with polydimethylsiloxane (PDMS)
  - Milling machine to create molds in polypropylene or acrylic
- Assume
  - Fully addressable array, static patterns
  - Two states at each site (raised or flat)
  - Circularly symmetric, bell-shaped bumps
- Control spacing, diameter, height, and shape of dots
- Material properties: softness, friction
  - PDMS vs. embossed paper

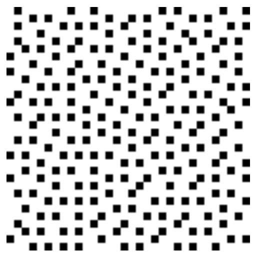
Tactile Pattern Generation



- Generate perceptually distinct tactile textures
- Leverage existing techniques: digital halftoning
- Visual patterns: Minimize visibility of halftone-induced textures
- Tactile patterns: Accentuate texture characteristics

Dot diameter: 1.0 mm   Spacing: 1.0 mm   Height: 0.2 mm

Visual Vs. Tactile Pattern Perception



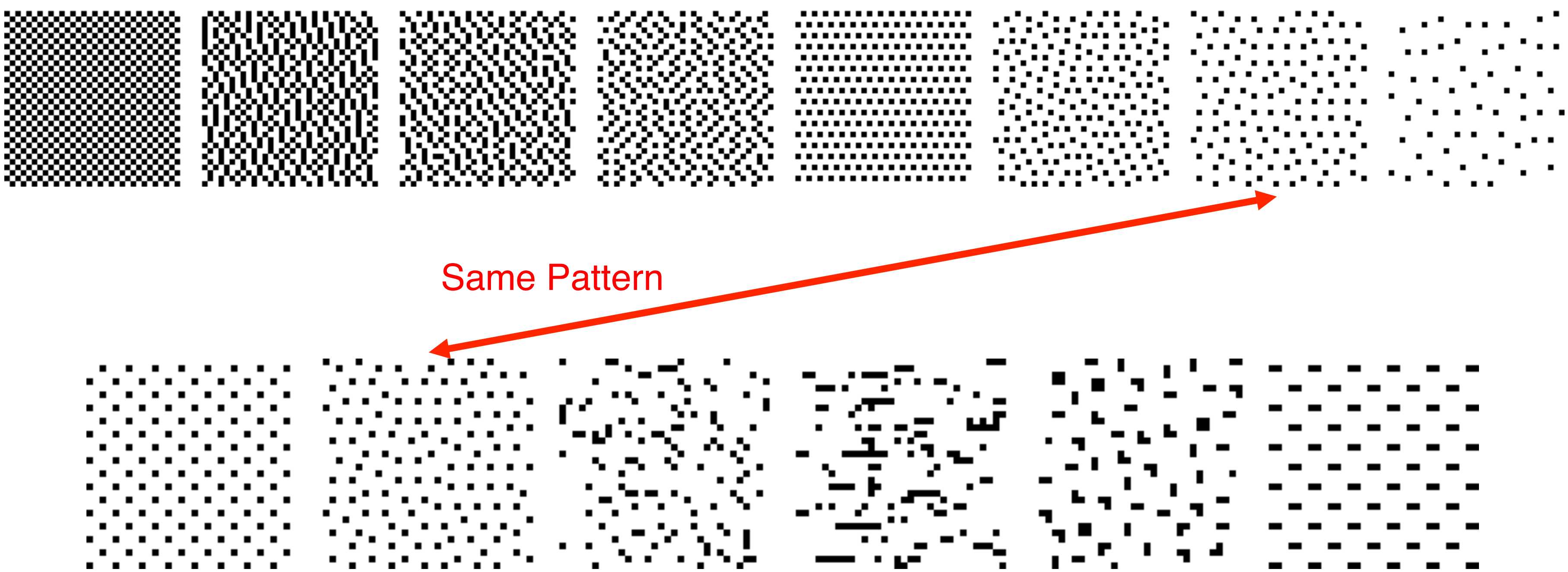
- Visually pleasing blue noise pattern
  - Floyd-Steinberg error diffusion
  - High frequency noise, less visible
- Tactile impression: smooth, boring

- Visually less pleasing
  - Error diffusion with weight perturbations
  - Contains more low frequencies
- Tactile impression: interesting, exciting

- Visually impaired and blind subjects
- Visually blocked subjects

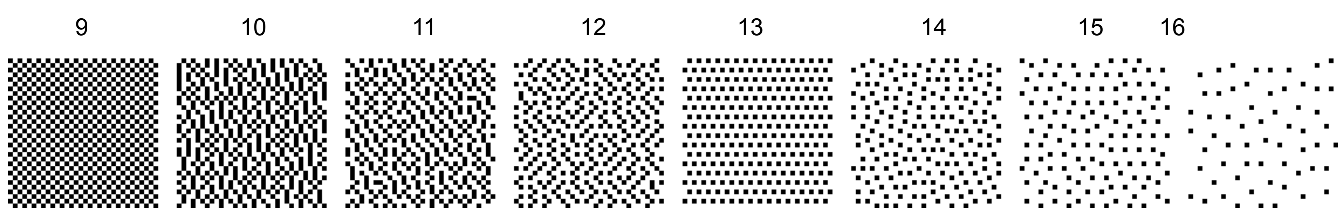
Tactile Patterns

Decreasing Density

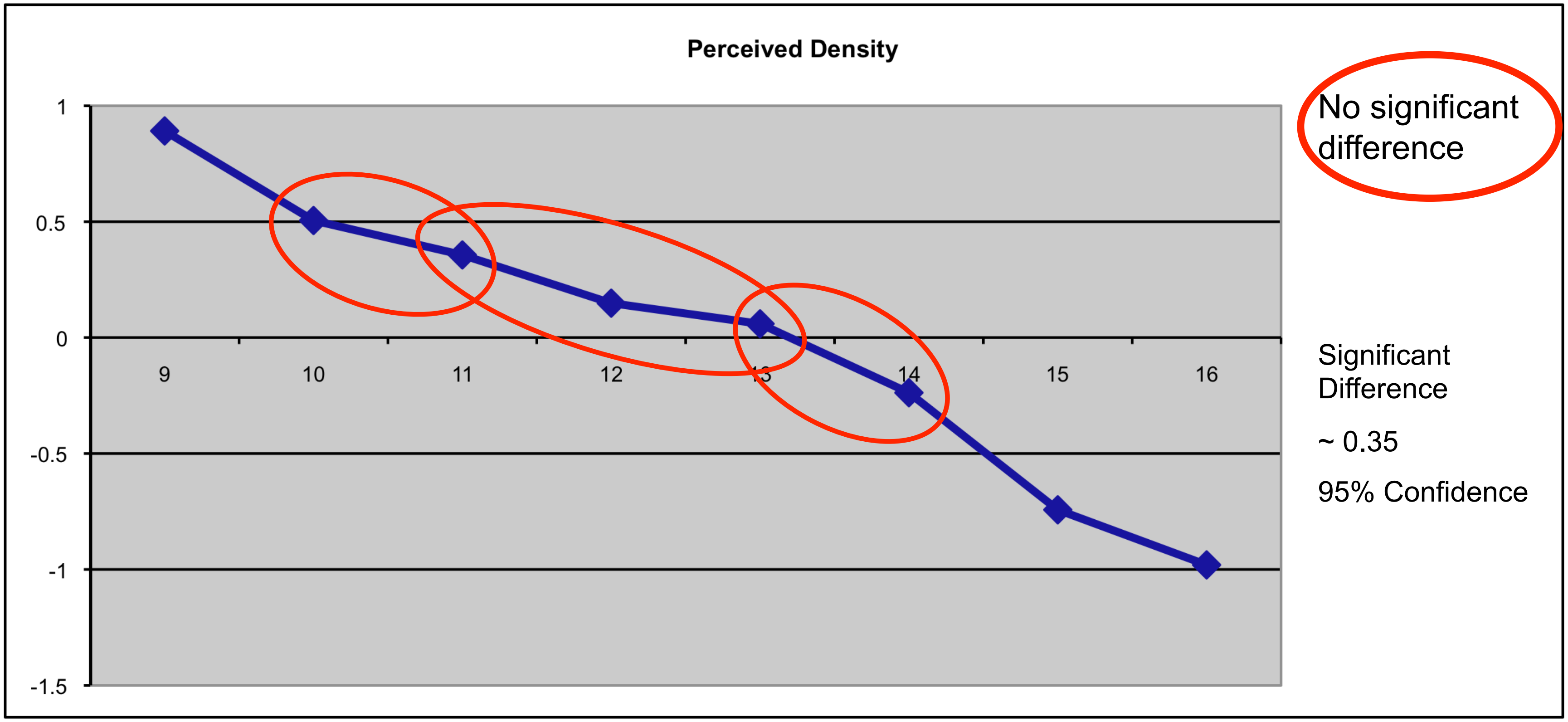


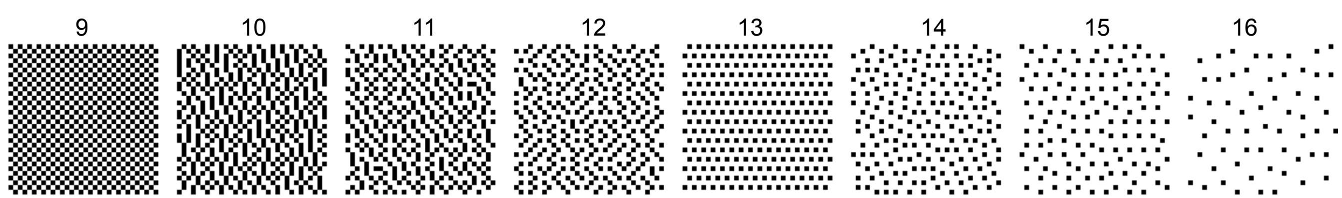
Equal Density – Different Pattern

Physical Density (Across Dot Density)



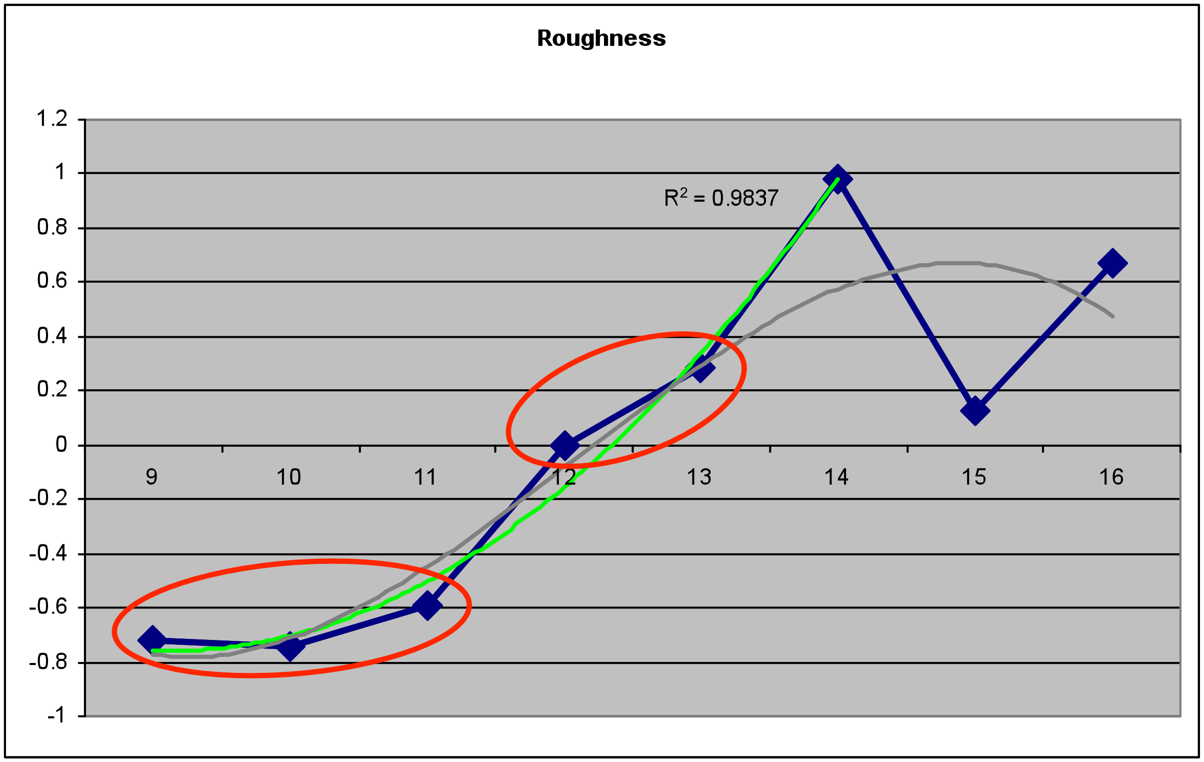
Perceived Density (Across Dot Density)





- Perceived closely matches physical density

Roughness (Across Dot Density)

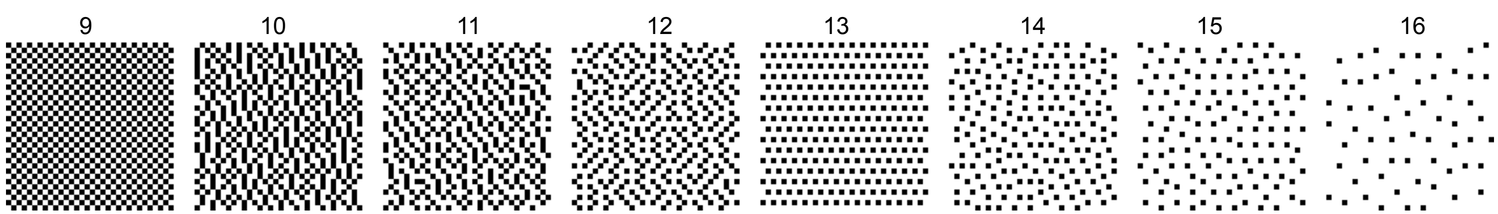


No significant difference

Significant Difference

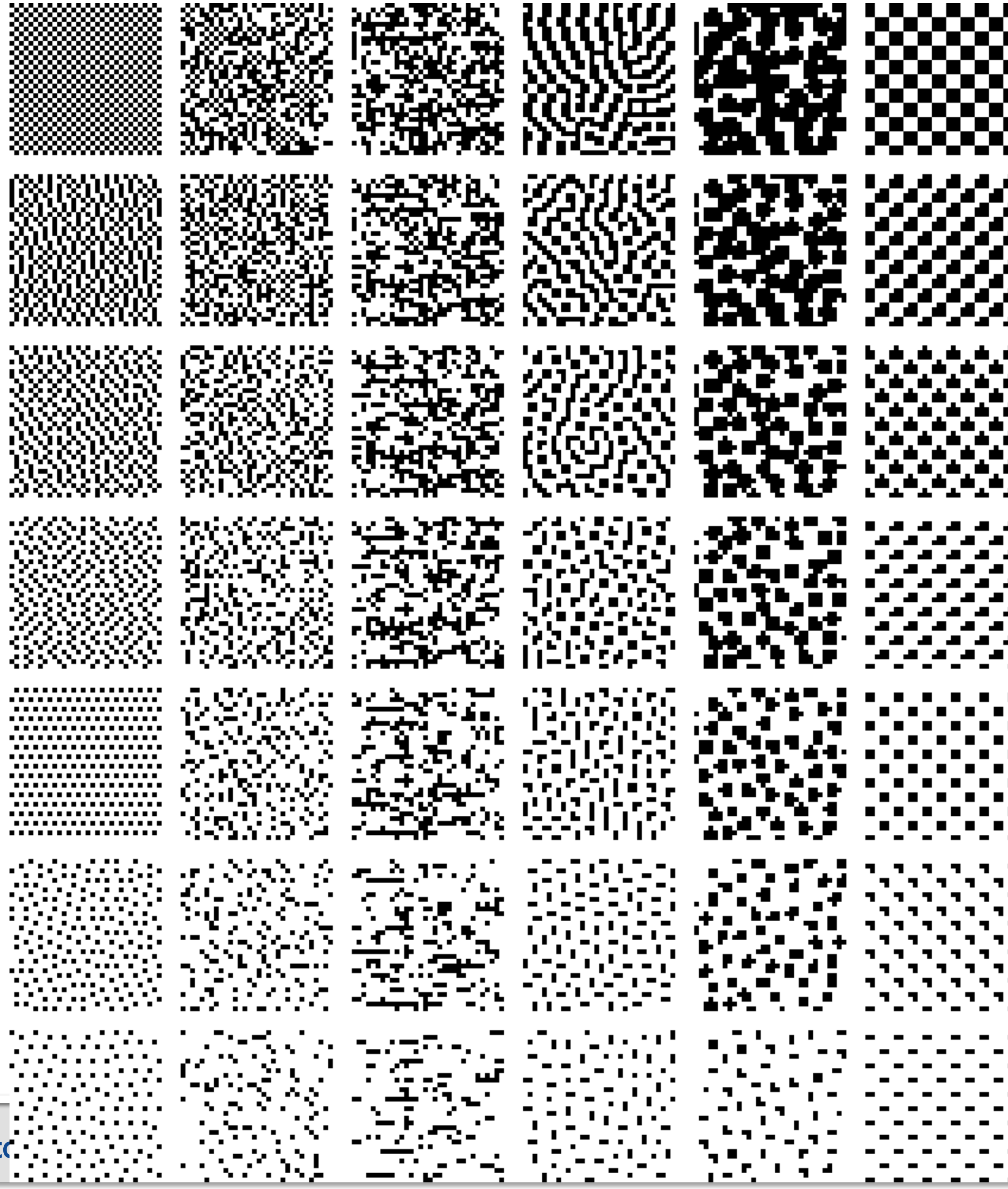
~ 0.3

95% Confidence



- “When the elements get too sparse, on the order of 3-4 mm apart, people do not perceive the surface as textured.” [Klatzky, Lederman ’02]

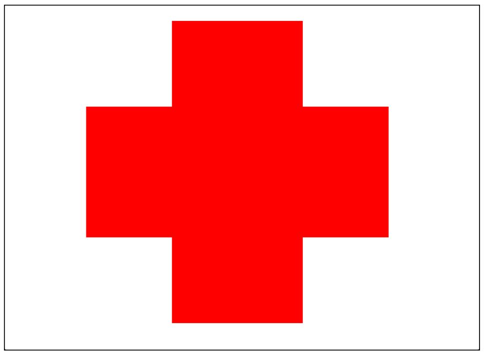
Patterns



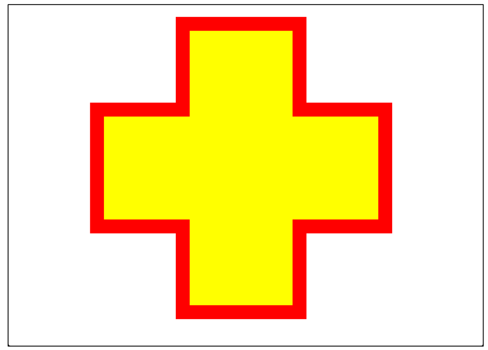
Devices

- Tactile Sensing
  - iPhone, iPad, other touch screen interfaces
- Tactile Display?
  - Variable friction
  - Vibration
  - Mechanical pin arrays
  - Dynamic electrically activated dot patterns?
    - — Dynamic Tactile Tablet
- Acoustic Display
  - Use finger to actively explore 2-D layout on touch screen
  - Touch used as pointing device
  - Provides kinesthetic feedback
  - Static tactile overlay

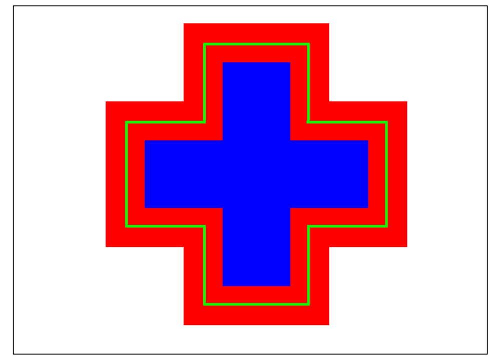
Conveying Shape



C1



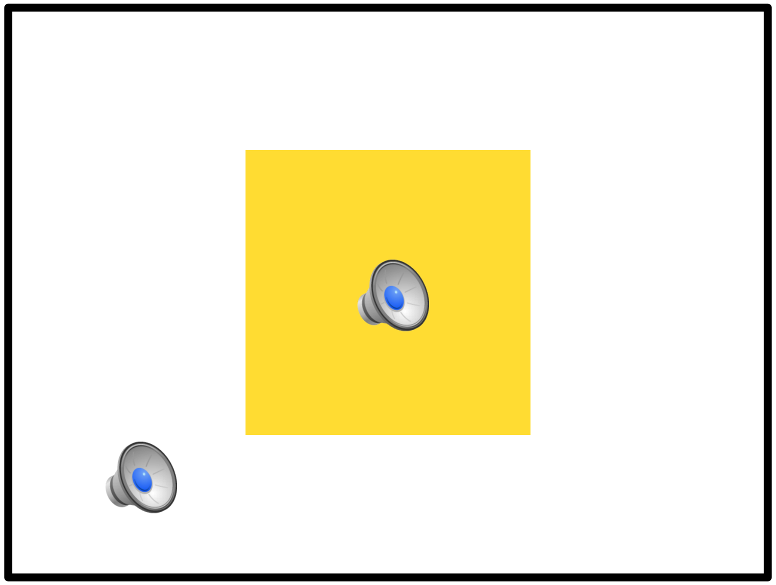
C2, C4, C5



C3

C1	2 Constant sounds
C2	3 Constant sounds
C3	2 Tremolo sounds with varying border rate
C4	3 Sounds with varying border intensity
C5	3 Sounds with HRTF (directionality and proximity) in background and border – original: KEMAR, modified: human

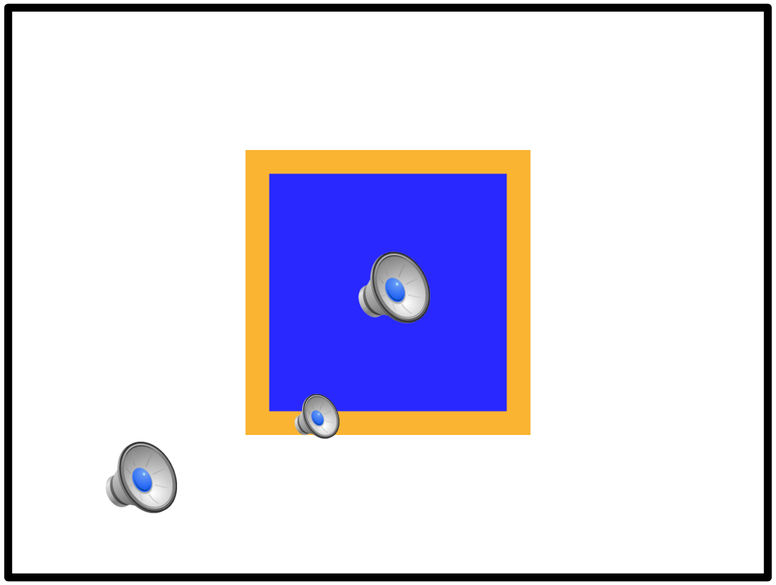
Shape C1: Two Constant Sounds



- Two regions with distinct constant sounds

Silva, et al., ICASSP 2011

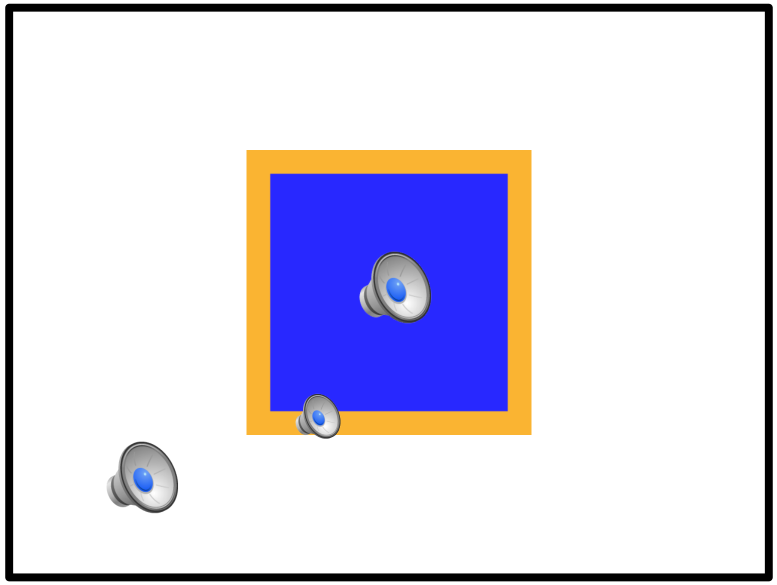
Shape C2: Three Constant Sounds



- Three regions with distinct constant sounds

Silva, et al., ICASSP 2011

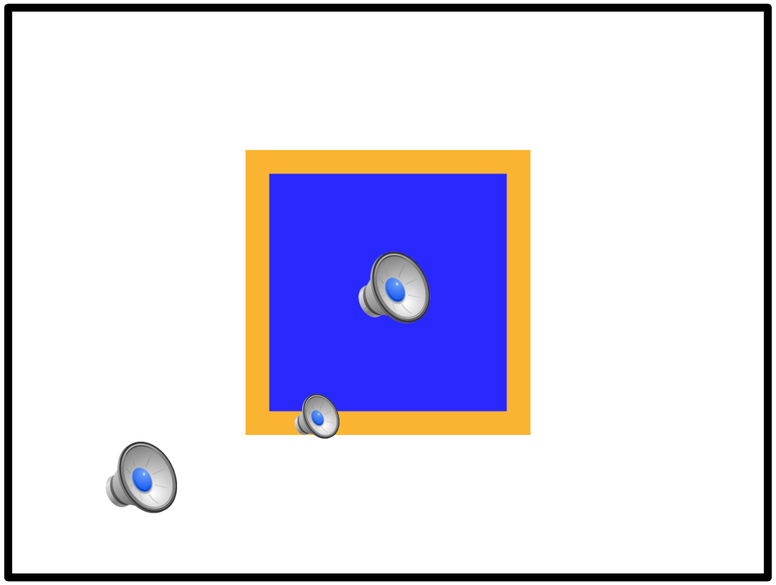
Shape C4: Three Sounds, Variable Intensity Border



- Background and object: constant sounds
- Border: variable intensity

Silva, et al., submitted

Shape C5: Three Sounds, Directional Sounds



- Use directional acoustic feedback (background and border)
- Head Related Transfer Function (HRTF)
- Playback via stereo headphones

Silva, et al., ICASSP 2011

Shape - Experiments

- Two sets of experiments with basic shapes
- First set with 21 subjects – touch screen users (except one)
  - C1-2cons, C2-3cons, C3-2trem, C5-3hrtf-ke
- Second set with different subjects
  - Unaltered C2-3cons – for comparison
  - Added C4-3int
  - Modified C5-3hrtf-hu (better sounds, human HRTFs)
  - 6 subjects – touch screen users (experienced)
  - 5 subjects – little experience with touch screen devices
- Apple iPad touch screen
- SENNHEISER HD595 headphones

Experimental Procedure

- Subjects had no prior information about the objects they were going to be tested on

- Training example with the same task but different object (or scene) at the beginning of each experiment
- Initially, the subject was able to see the pattern/shape and the scanning finger on the touch screen
- Then, the subject repeated the trial without seeing the pattern/shape or the finger

Shape - Results

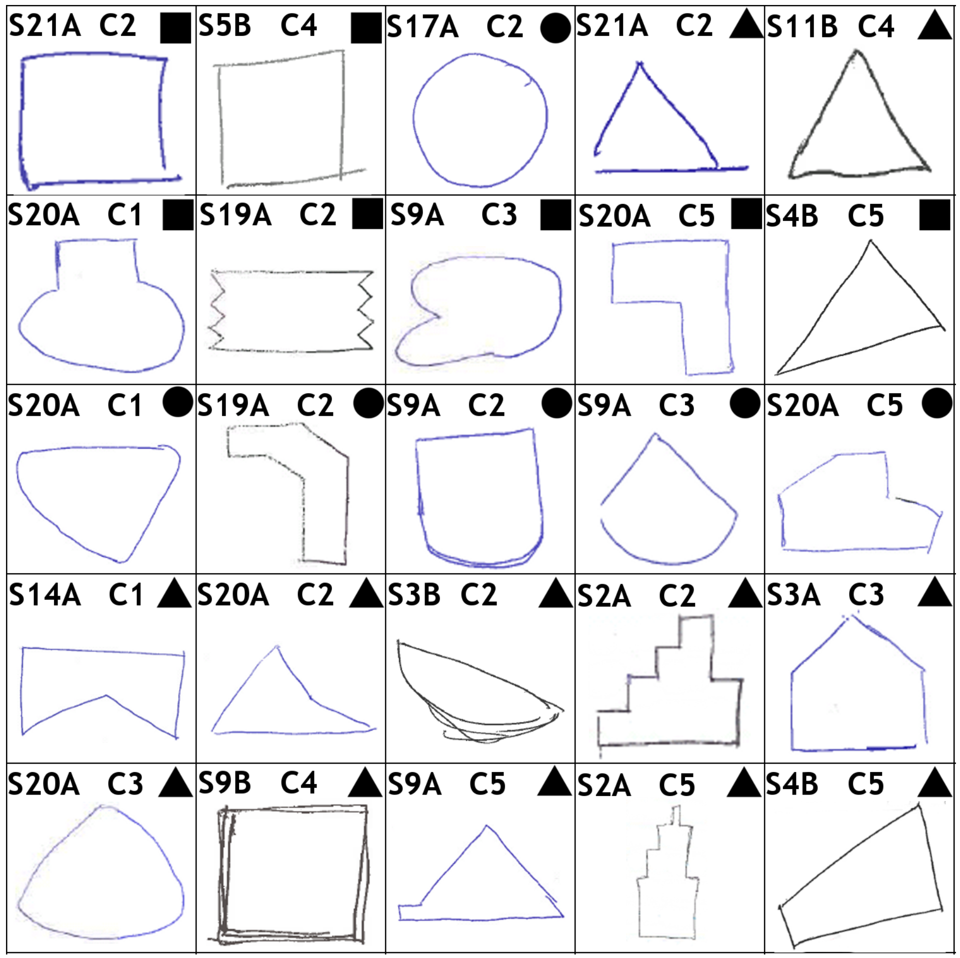
First set of experiments

	C1-2cons	C2-3cons	C3-2trem	C5-3hrtf-ke
Accuracy	67%	81%	72%	80%
Time	236 s	228 s	181 s	182 s

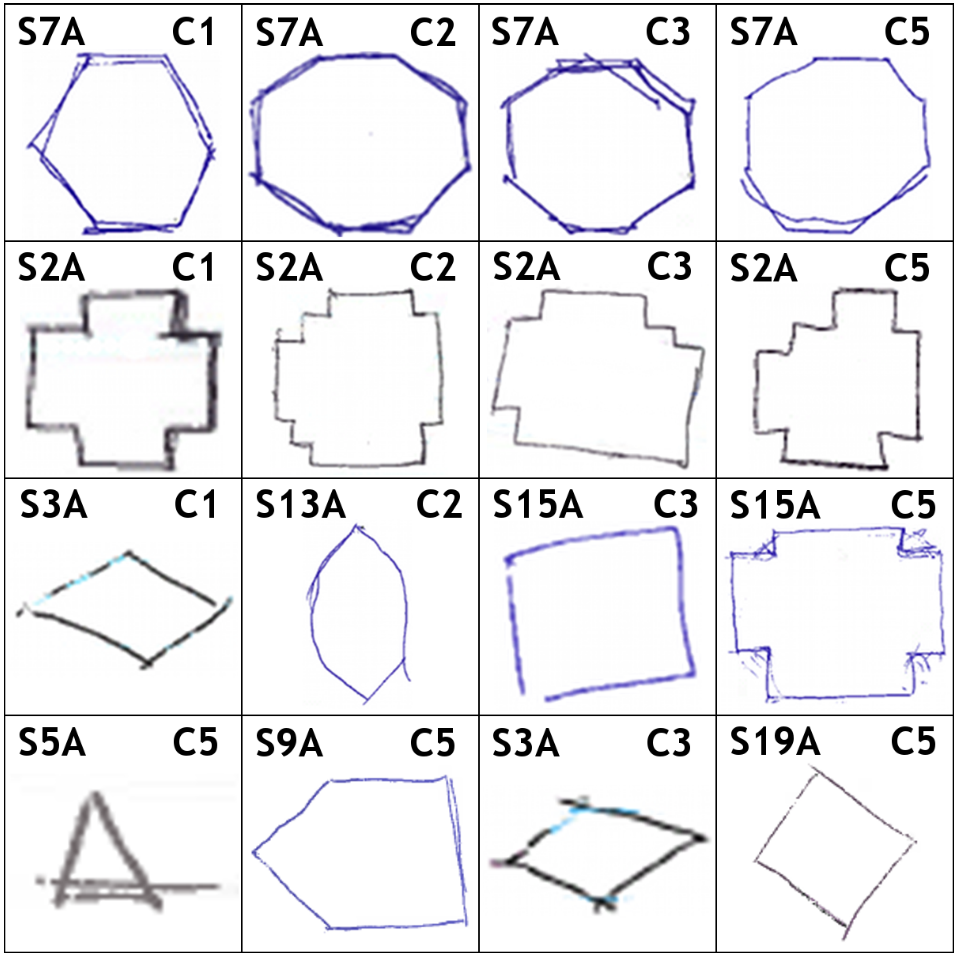
Second set of experiments

	C2-3cons	C4-3int	mC5-3hrtf-hu
Accuracy	70%	82%	73%
Accuracy (6 subjects)	78%	89%	89%
Time	331 s	259 s	189 s
Time (6 subjects)	243 s	212 s	103 s

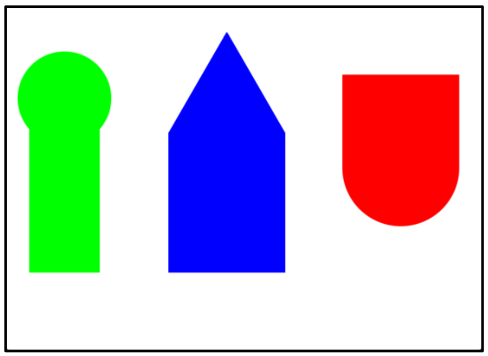
Shape Approximations: C1 – C5

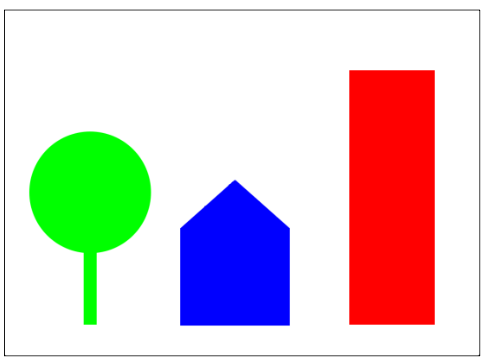


Circle Approximations: C1 – C5



Shape – Simple Layout





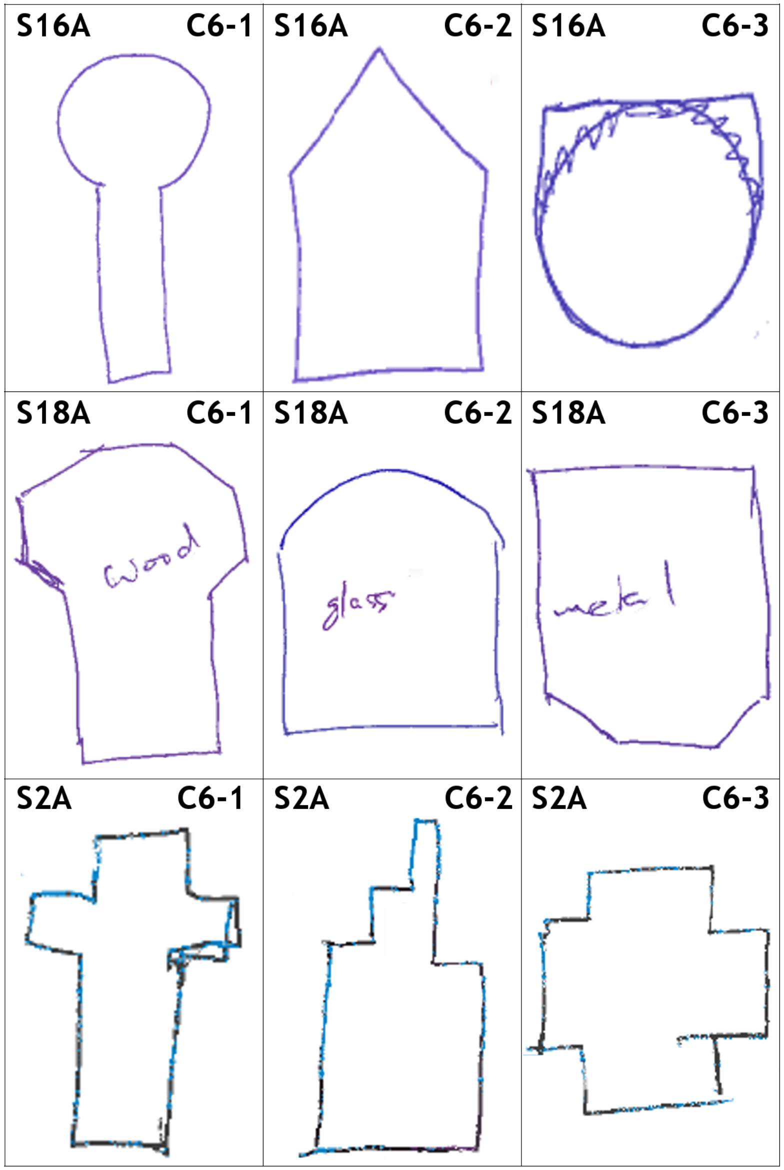
C6	Virtual cane – acoustic display with zoomed-in mode
C7	Virtual cane – acoustic display with tactile overlay

- Tapping sounds for objects; silent background
- Zoomed-in mode: one shape presented with C5

	C6-cane-ac-zm	C7-cane-ac-ta
Number of Subjects	21	5
Number of Objects	100%	100%
Shape Accuracy	23%	100%
Material Identification	80%	73%
Time	745 s	240 s

Drawings for C6

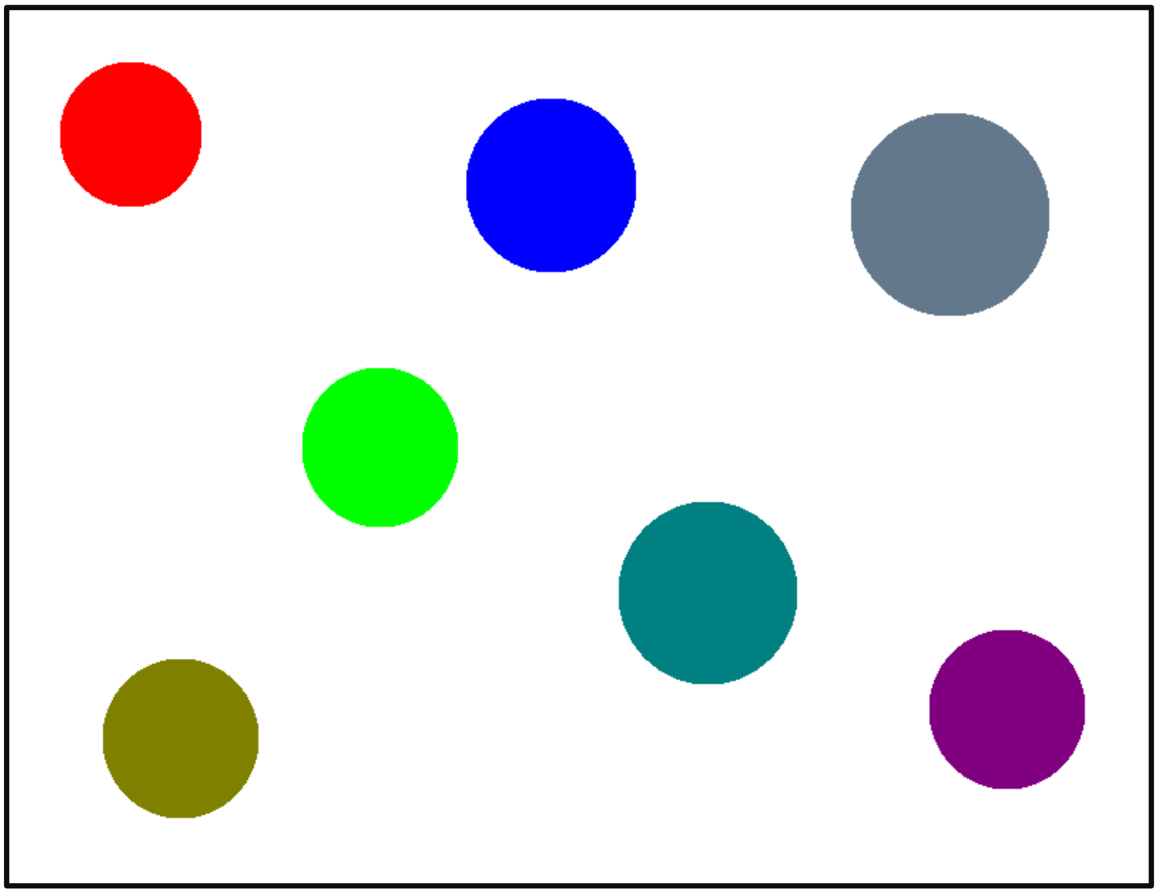
- Subjects were asked to draw objects in scene indicating their relative positions, shapes, and the material they are made off



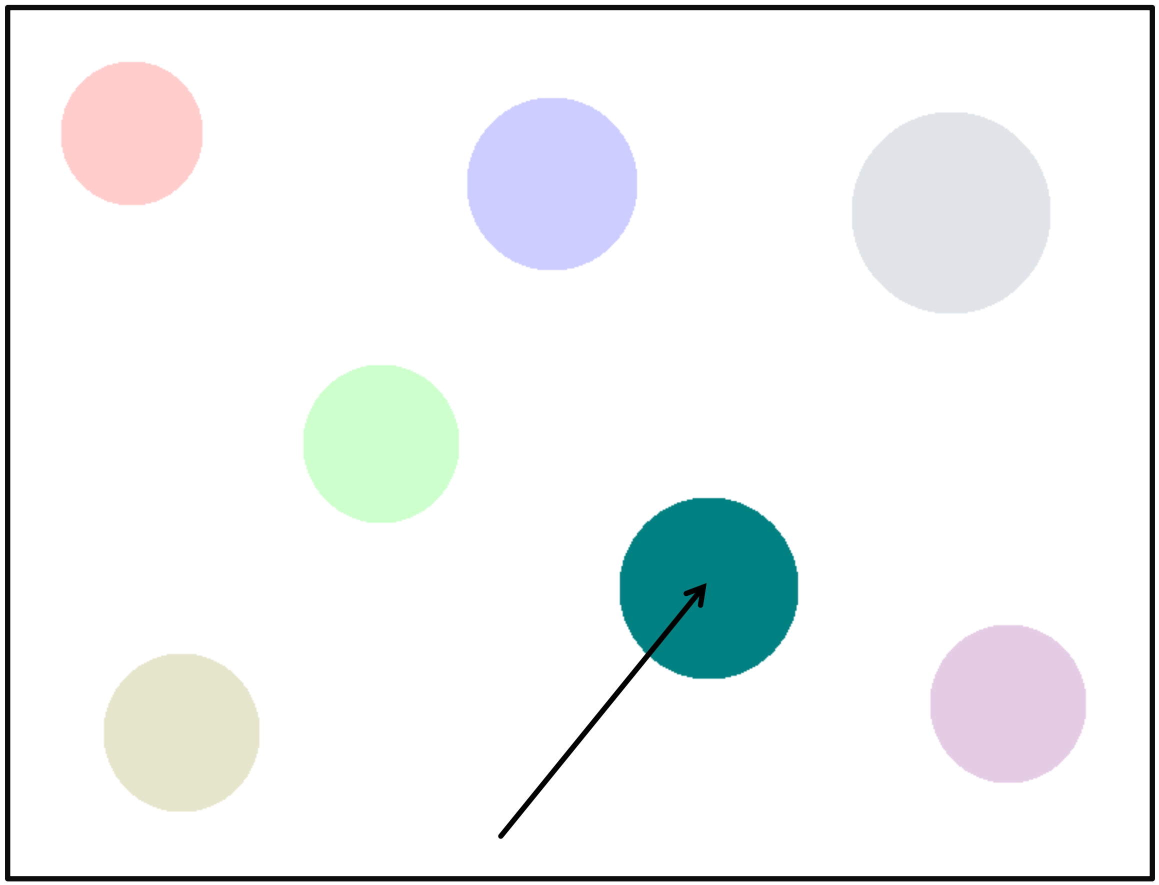
Shape - Conclusions

- Outperformed all existing techniques
  - Acoustic displays: SoundView, vOICe
  - Dynamic tactile displays: Tesla Touch
- Considered advantages of proximity & directionality cues
- Virtual Cane
  - Acoustic-tactile (C7) significantly better than acoustic with zoom (C6) in both accuracy and time
  - **Raised dot patterns best for shape rendition**
- Considerable learning curve
  - Significant difference in accuracy (with comparable or better time) between experienced and inexperienced subjects
  - Performance could be improved by systematic **training**

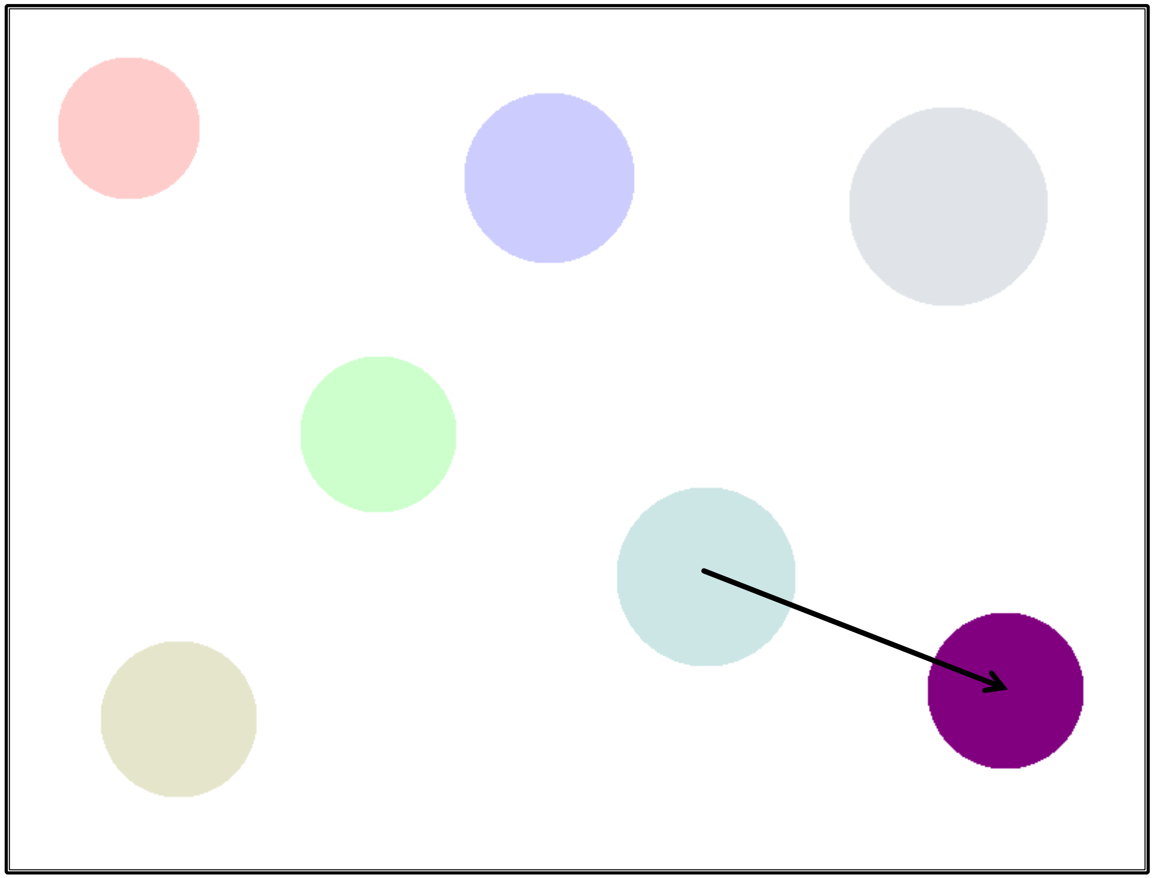
2-D Object Layout



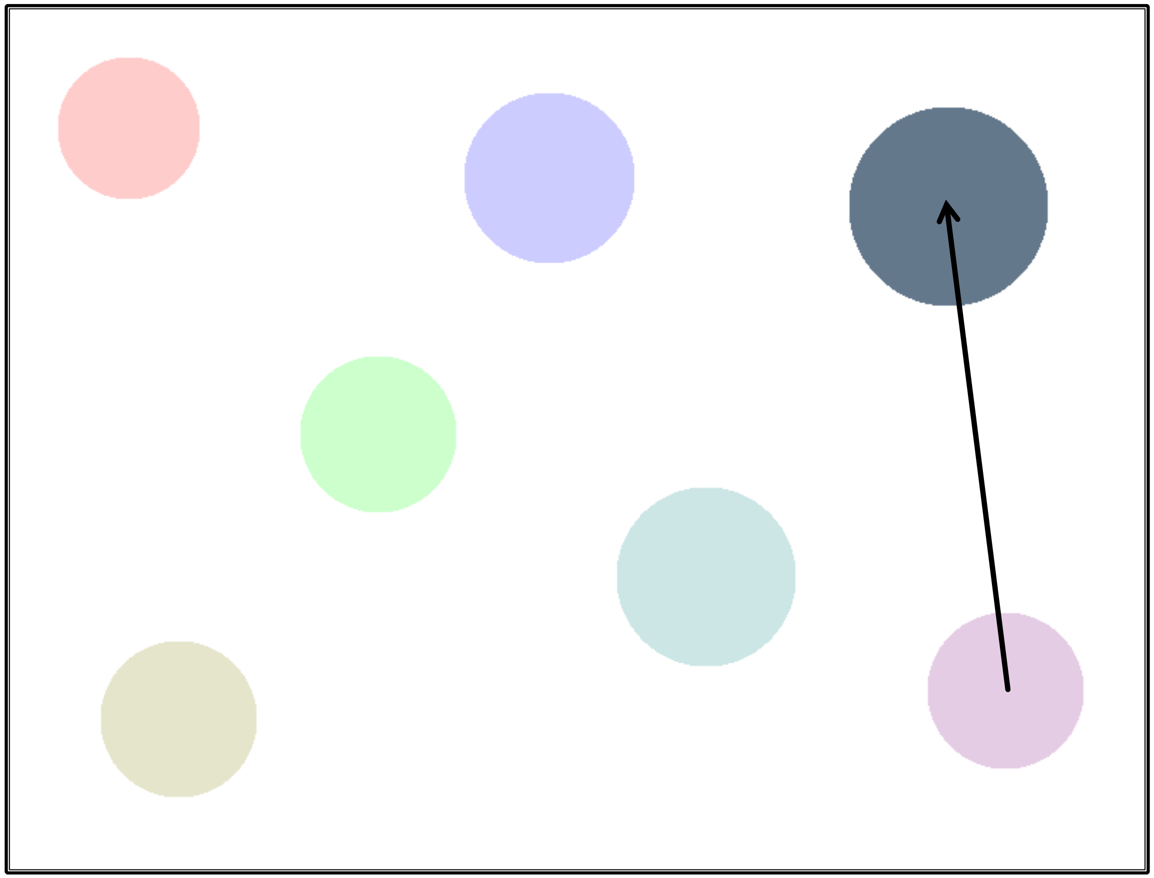
2-D Object Layout



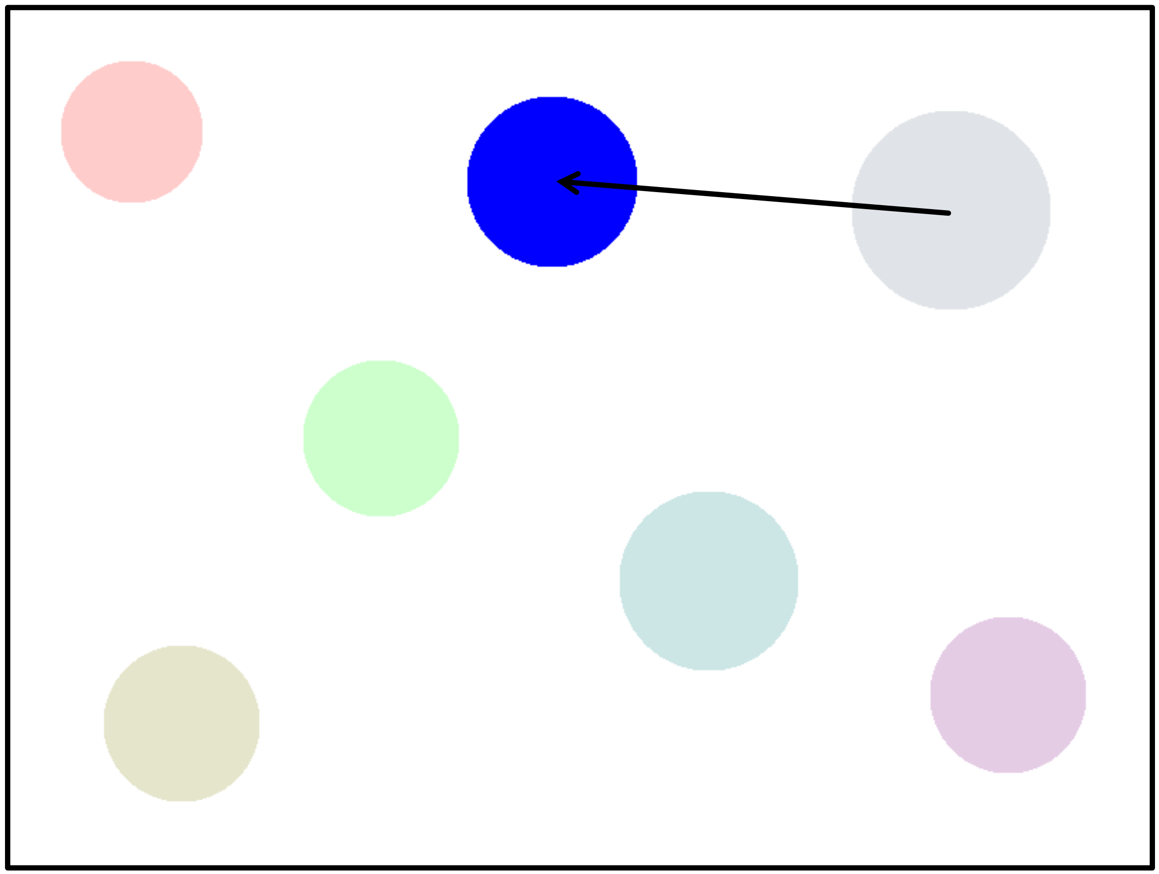
2-D Object Layout



2-D Object Layout



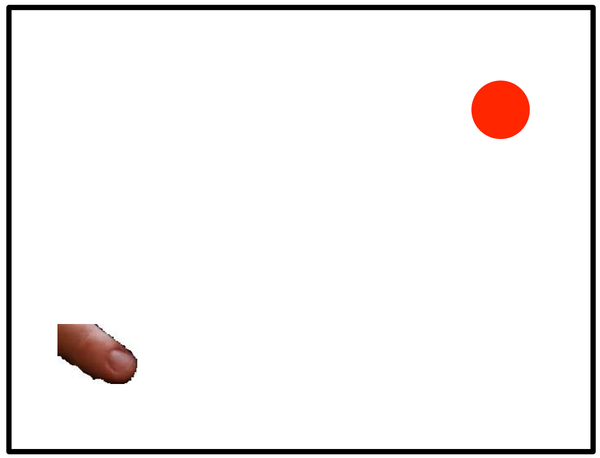
2-D Object Layout



Locating Object: Sound Rendering

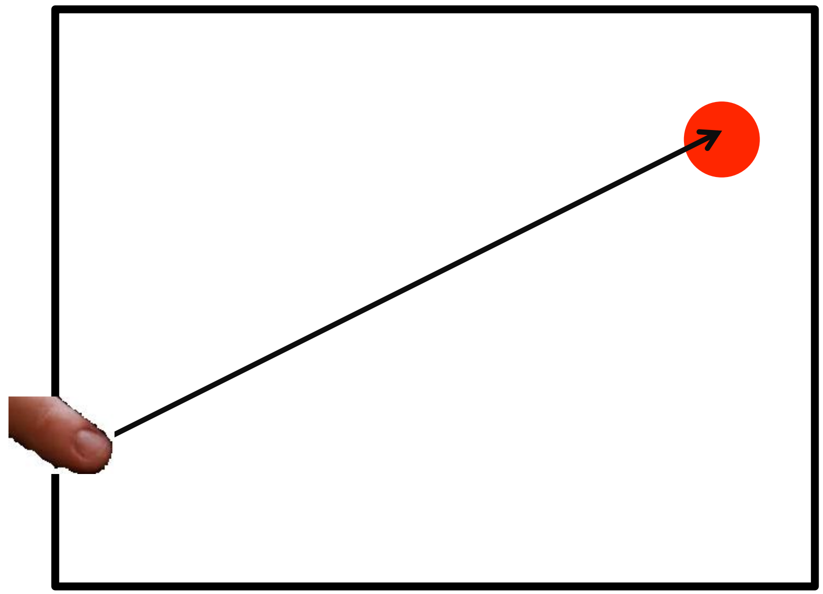
- **Directionality**
  - Head Related Transfer Function – HRTF
  - Natural cue: Models acoustic signals that enter ear from source at given location
- **Proximity**
  - Natural cues: Intensity, direct-to-reverberant energy ratio, spectrum distortions, binaural differences
  - Humans are consistently inaccurate in acoustic proximity judgments [Zahorik ICAD’02, JASA’02]
  - Use intuitive cues (not realistic)
    - Relative intensity of object tone and background noise
    - Tempo variations

Nav: Navigation Experiment



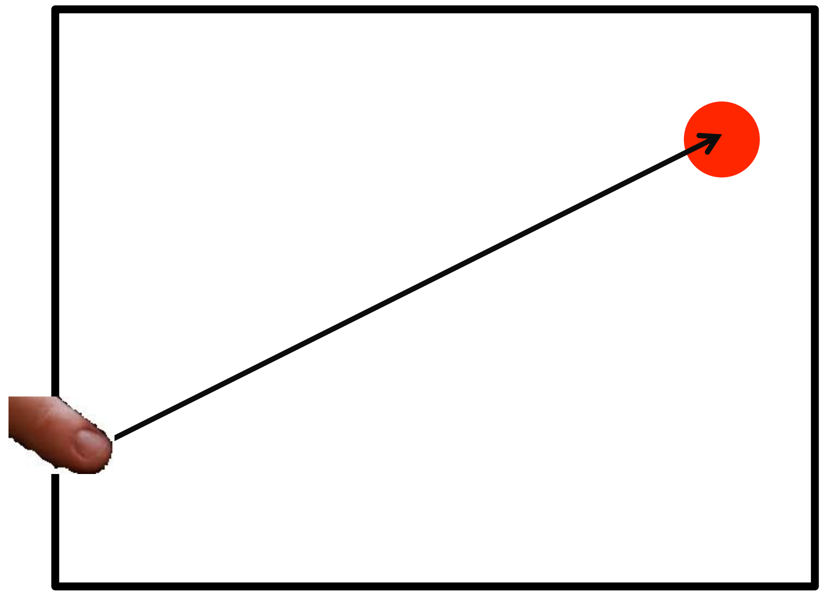
- Task: locate a single dot as fast as possible – and notify
- Multiple trials in a fixed time window (10 minutes)
- Random object placement in each trial
- Measure time per trial
- **Object:** tone; **background:** tone + white noise
- Proximity via **intensity** or **tempo** variations

Locating Object: Intensity



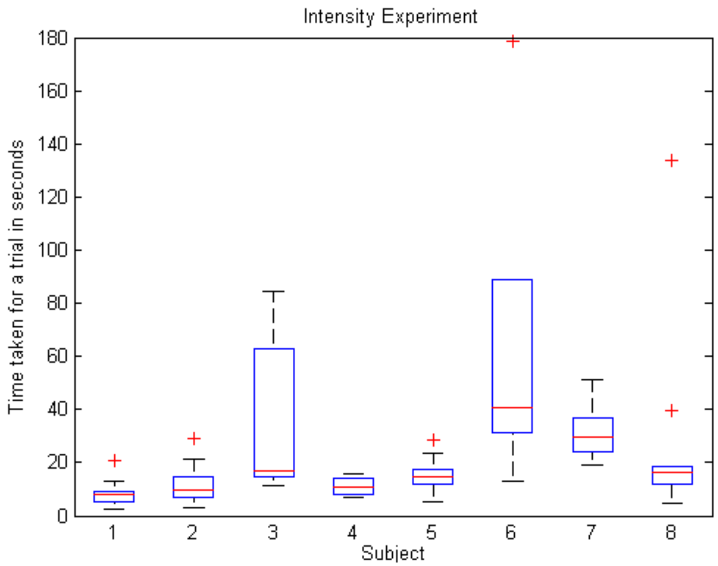
- Directionality via HRTF
- Proximity via relative intensity
  - Tone-to-noise intensity ratio

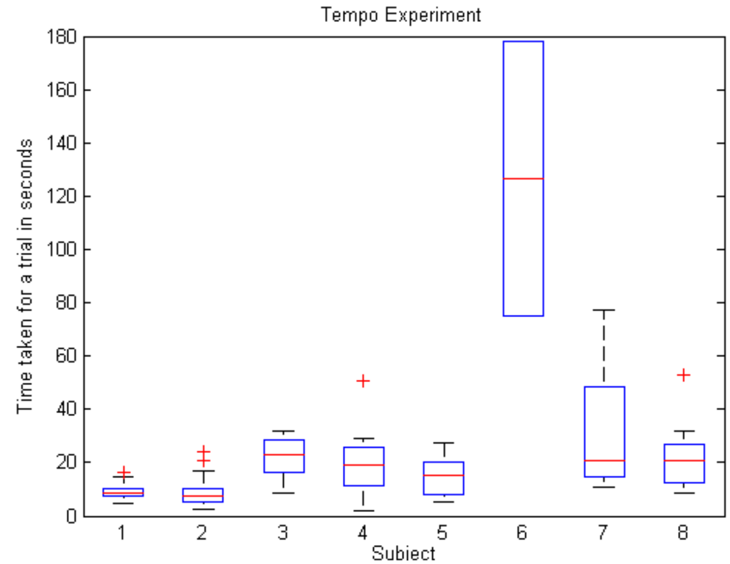
Locating Object: Tempo



- Directionality via HRTF
- Proximity via tempo viariations
  - Quantized to 3 levels (1 – 3 Hz)

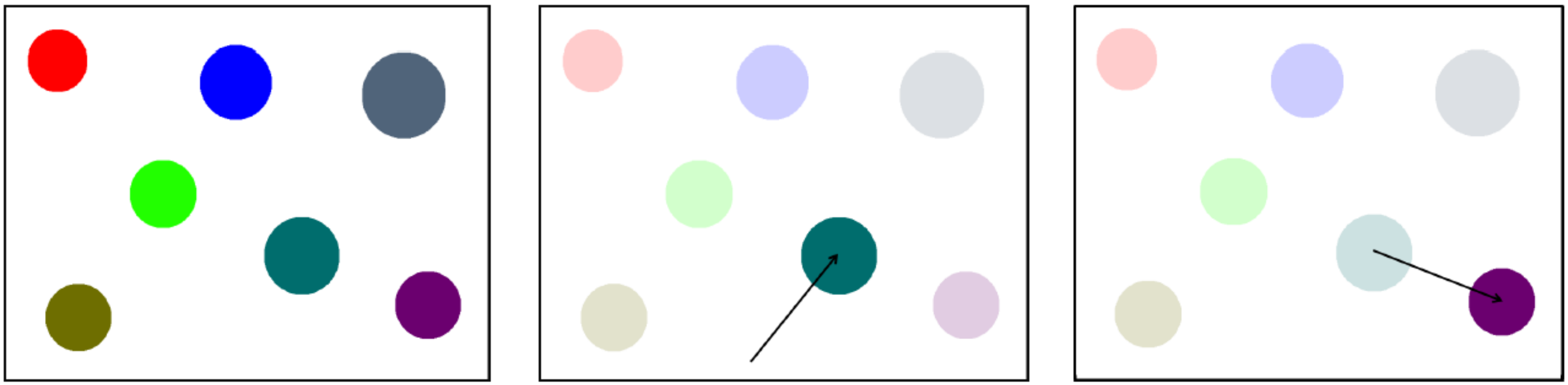
NAV: Intensity *vs.* Tempo





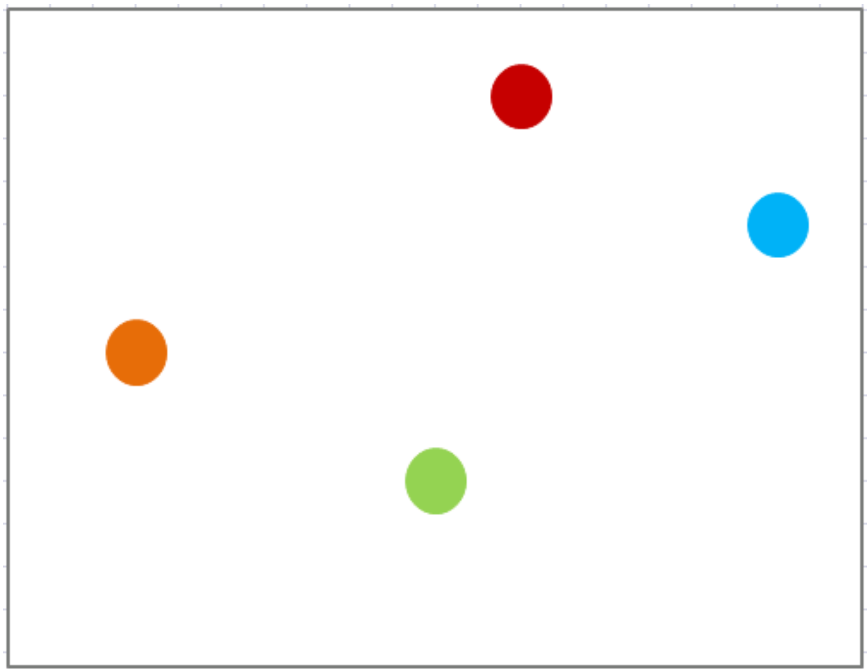
- 8 subjects: 4 male, 4 female
- Intensity: mean = 25.6 s, median = 15.6 s
- Tempo: mean = 32 s, median = 19.8 s
- No significant differences (t-test: $p = 0.31$)
- Intensity: rendition is instantaneous; continuum of values (vs. a few levels of tempo)

Serial Layout – L1



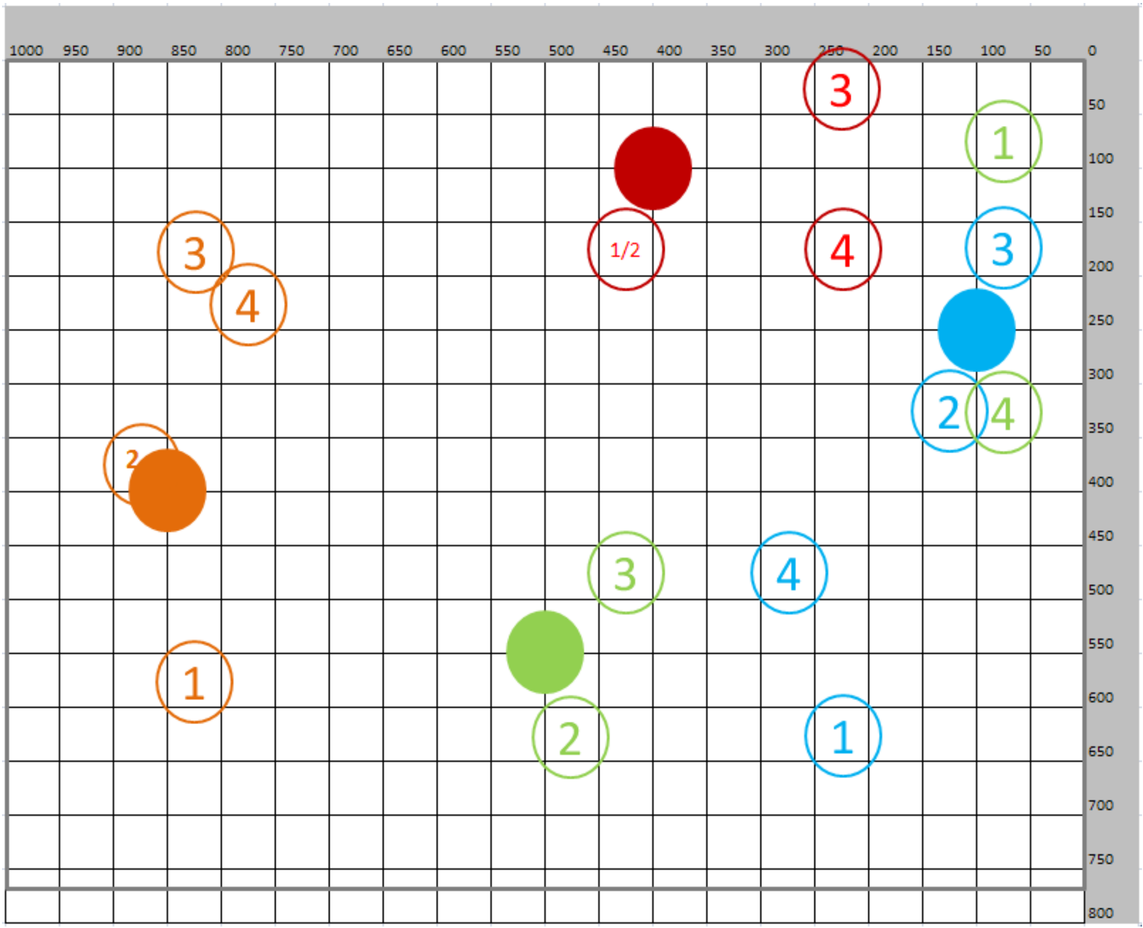
- Objects presented serially, one at a time
- Starts with object closest to the finger
- Double taps to get the next
- Presented in cycles and visited marked ‘inactive’

Serial Layout – L1: HRTF + Intensity



- HRTF for directionality; intensity (tone-to-noise ratio) for proximity
- Pitch for object identification (452, 652, 852 &1052 Hz)
- Subjects asked to draw object corresponding to each sound in graph sheet
- No time limitations; not allowed to draw during exploration
- Implemented on iPad 1
- 4 Subjects, 2 male, 2 female

Serial Layout – L1: Results



- Average time: 7 mins

Serial Layout – L1: Problems

- Proximity rendering via intensity (tone-to-noise ratio)
  - Insensitivity of intensity for small finger movements
- Object confusions
  - Hard to discriminable and memorize sounds
- “Manhattan scanning”
- Serial exploration

Layout: Enhancements

- Directionality rendering
  - 54 quantization levels: 5^0 steps for [-45^0, 45^0]; 10^0 steps otherwise
  - Boosted high frequencies of sounds
- Calibrated proximity vs. volume curve
  - Measured the relationship between tablet volume and intensity
  - SPL at headphones measured for 50 uniform volumes of 1KHz sine
  - Curve designed such that SPL is uniformly varied with proximity
- Proximity via direct-to-reverberant ratio
  - Natural proximity cue
- Musical instrument sounds
- Listener orientation
- Non-serial scanning

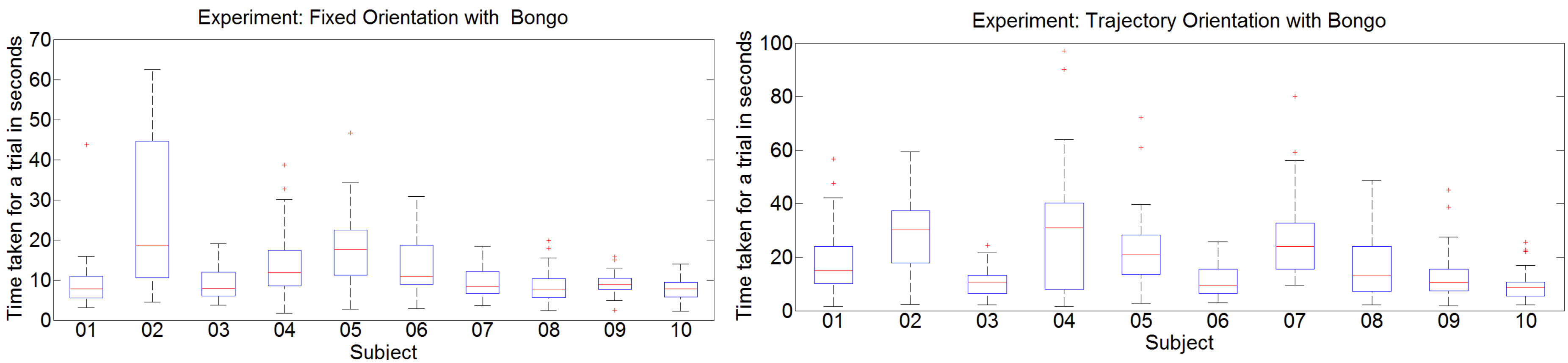
Sound Selection: Percussion vs. Wind

- Navigation experiment
  - 4 male subjects
  - Object: Bongo roll vs. trumpet
  - Background: object sound + reverb + directionality + proximity
  - Different tempo for object and background
- Time per trial
  - Bongo roll: mean = 14 s, median = 10.1 s
  - Trumpet: mean = 13.4 s, median = 9 s
  - No significant difference (t-test: $p=0.47$)
- Conclusion: Enables use of diverse set of sounds

Virtual Listener Orientation

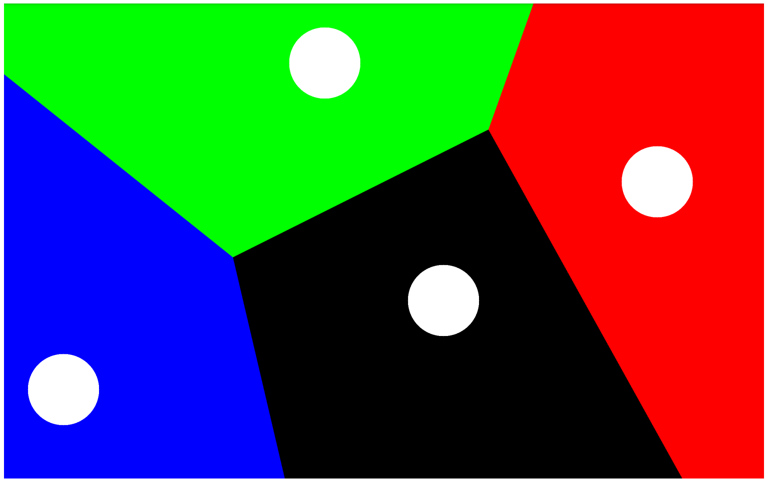
- Not the subject's physical head orientation
- Fixed orientation (FO): north
  - May be a reason for Manhattan scanning – ears most sensitive to head-on directionality changes
- Use direction of scanning pointer (virtual listener) movement
  - Based on the scanning trajectory (TO: trajectory orientation)
  - Analogous to natural human behavior
    - Face object as you move toward the object
    - Will this eliminate Manhattan scanning?
    - Will this add confusion?

Fixed vs. Trajectory Orientation



- Navigation Experiment to determine the best method
  - 10 subjects, bongo roll for object sound
- Time per trial
  - FO: mean = 12.1 s (median = 9.5 s) TO: mean = 17.9 s (median = 13.2 s)
  - Significant difference (t-test: p=0.01)
- Subject ratings
  - Difficulty: FO: mean = 3.3 TO: mean = 5.8
  - Cognitive load: FO: mean = 2.95 TO: mean = 5.7
  - Significant difference in Difficulty (p=0.02) and Cognitive load (p=0.01)

Voronoi Layout

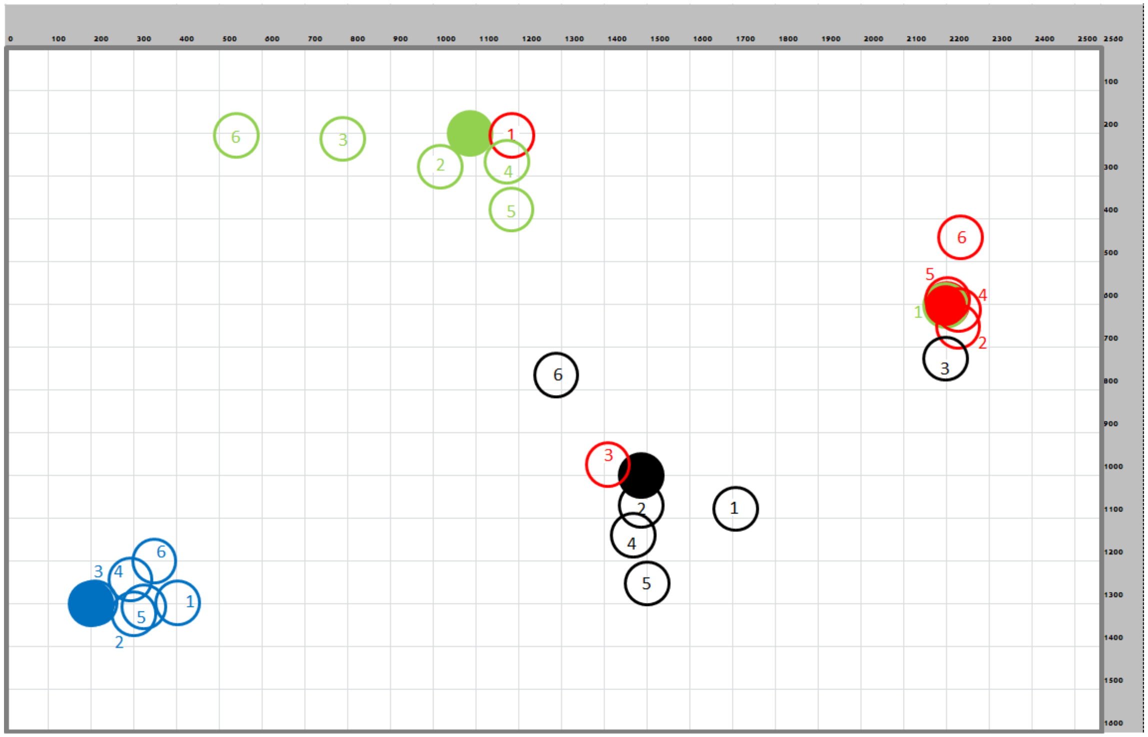


- All objects are available on the screen at once
- Subject hears only sound of object closest to finger location
  - Sound guides to object in the region
- Screen is partitioned to Voronoi regions of object centers
- Each object's background is limited to its Voronoi region

L2-vor: Voronoi Layout

- Initial mode: Serial introduction of objects
- Main mode: Voronoi layout
- 6 subjects, FO
- Sounds
  - G3 (note G of 3rd octave) of bass clarinet
  - B3 of oboe
  - D5 of trumpet with no vibrato
  - Bongo roll

L2-vor Results

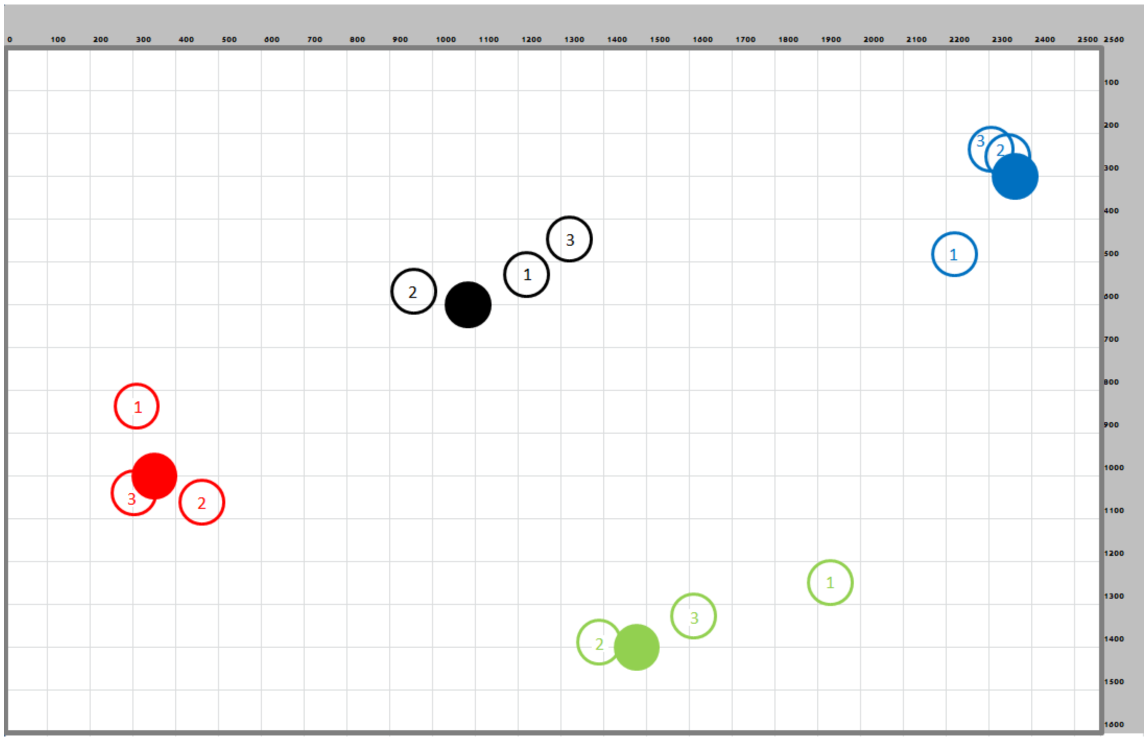


- Average time: 8 minutes

L3-ung: Unguided Layout

- Each object identified by characteristic sound
- Background is silent
  - Arbitrary scanning
- Used as benchmark to analyze the effectiveness of acoustic guidance (directionality and proximity)
- Experiment
  - 3 subjects (out of 6 subjects of L2-vor)
  - Same 4 sounds as L2-vor experiment
  - Layout was the transpose of L2-vor

L3-ung Results



- No confusions
- Average time 14 minutes

Comparison: L1, L2 & L3

- Error of reproduction (EOR)
  - Displacement between object location and subject placement
    - Measured in pixels
    - Averaged across objects and subjects
  - Normalized for resolution
    - Expressed as percentage of maximum distance for given resolution

Layout	EOR	Time
L1-ser	16.6%	7 minutes
L2-vor	9.9%	8 minutes
L2-vor after correcting for confusions	4.9%	
L3-ung	4.1%	14 minutes

Does Acoustic Guidance Help?

- Objects were represented by dots of 0.2 inch radius
- 4 dots took 4.5% of the total screen area

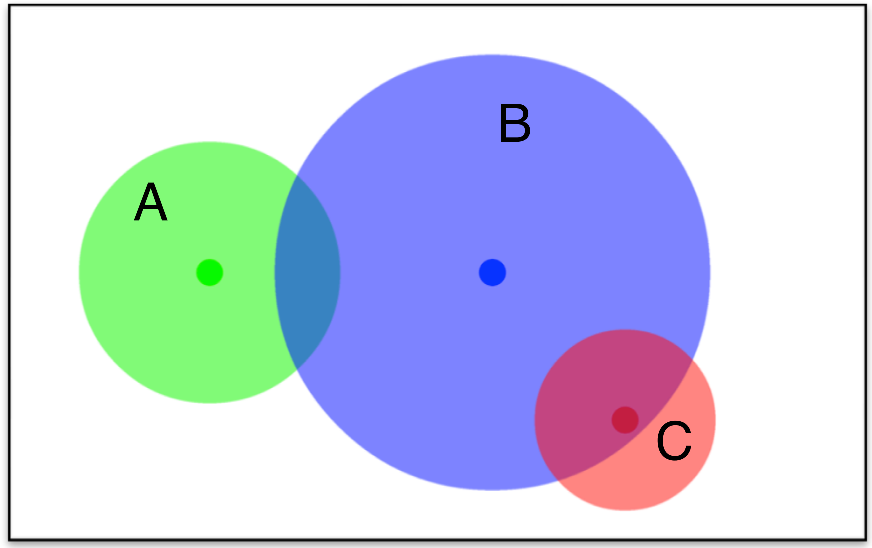
- How does the performance scale with object size?
- Conducted 3 navigation experiments
  - Silent (unguided) background (UG)
  - Guided background (GG)
    - With directionality and proximity as in L2-vor
  - All with bongo roll assigned to object

NAV: Dot Size Effect on Localization Guided vs. Unguided

- Two dot sizes of
  - Large size: 0.20 inch radius – original size
  - Small size: 0.05 inch radius – 1/16 of original area
- Stylus scanning only
- 3 subjects – will add more subjects

Experiment	Mean of trial times	Median of trial times
GG - Large dot	12.1 s	9.5 s
GG - Small dot	11.7 s	11.0 s
UG - Large dot	93 s	80 s
UG - Small dot	382 s	331 s

Venn Diagram



- Background is silent
- Similar to the Voronoi layout representation
  - Circular area analogous to the Voronoi region
    - With the exception of possible overlaps between circles
  - Circle center represented by a small dot (radius 0.07 inches) analogous to the object
- Multiple sounds played in overlaps

Venn Diagram

- Proximity rendition similar to L2
  - But, intensity change depends on radius of each circle (maximum at center, minimum at the edge)
  - Thus, intensity gradient provides cue for circle size
- Goal is to convey
  - Relative position and size of each circle
  - Amount of overlap – as a percentage of the smaller circle area
- Experiment
  - Subjects had to select among different choices
  - Subjects were then asked to draw and label the diagram

Venn Diagram: Results

	Overlaps			Relative Locations			Relative Sizes		
	A & B	B & C	C & A	A / B	B / C	C / A	A	B	C
Accuracy (%)	100	50	100	83.3	100	100	83.3	100	100

- 6 Subjects
- Available choices
  - Overlap: none, 10 – 40%, 40 – 60%, 60 – 90%, full
  - Location: N, NE, E, SE, S, SW, W, NW, N
  - Size: small, medium, large
- Actual overlaps
  - A & B : 16%
  - B & C : 59%
- All accuracies are better than the chance values

Conclusions: Shape

(Red indicates statistically significant results)

- Use of spatial sound (directionality and proximity cues)
  - Offers faster shape rendition for comparable or better accuracy
  - Performance (accuracy) can be improved significantly with training
- Raised-dot patterns
  - provide best shape rendition (in terms of time and accuracy)
  - but current technology does not allow dynamic display
- Friction display
  - Inferior to both for shape rendition

Conclusions: Localization

- Use of spatial sound (directionality and proximity cues)
  - Fixed head orientation superior to trajectory orientation in terms of time, difficulty, and cognitive load
  - Advantages depend of scene resolution
  - Large dot size: spatial sound outperforms unguided localization in terms of time to dot
  - Small dot size: Performance remains the same for spatial sound; goes down significantly for unguided layout

General Conclusions

- Dynamic (and static) acoustic-tactile representation of visual signals
  - Designed and implemented several configurations
  - Conducted pilot subjective tests that offer some statistically significant results, but also, many valuable insights for the design of further systematic tests with visually impaired and visually blocked subjects
- Acoustic display
  - Dynamic
  - Good for object identification
  - Can be used for shape rendition and object localization
- Raised dot display
  - Good for shape rendition
- Simple and intuitive concepts yield better results than natural analogies (fixed vs. trajectory orientation)
- Applications: Virtual cane, Venn diagrams