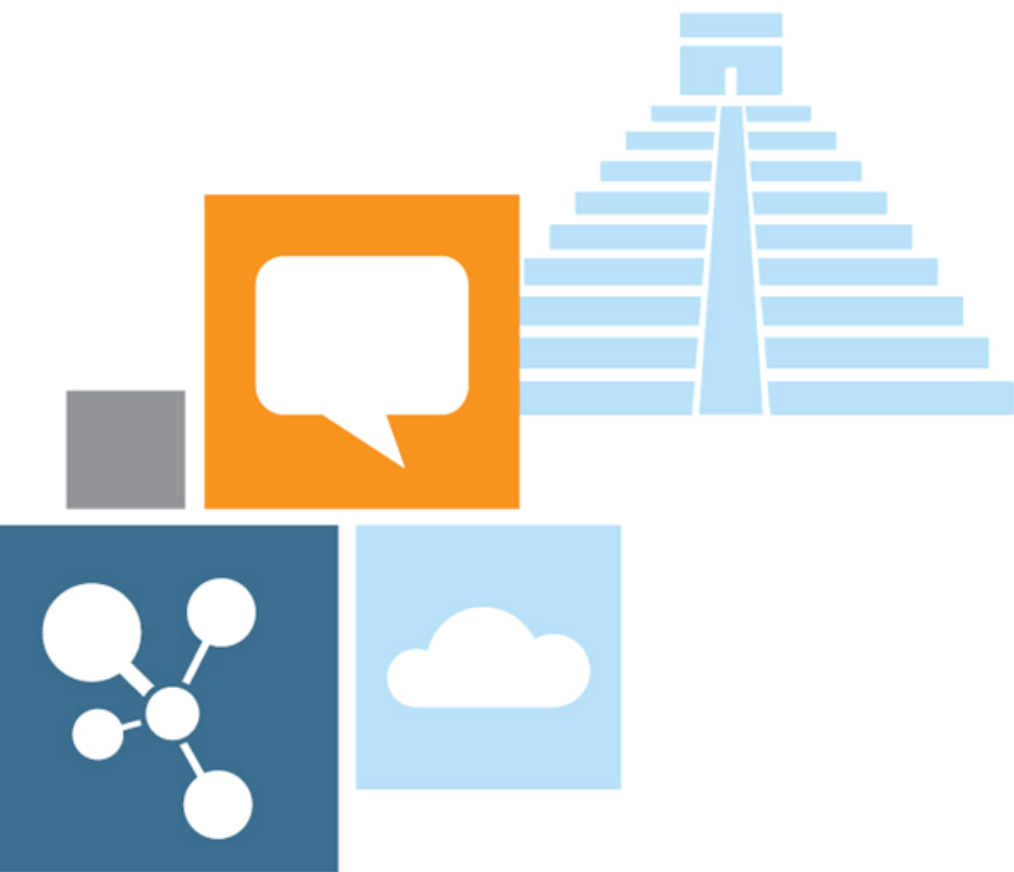


Microsoft



Microsoft® Research

Faculty Summit 2012

Riviera Maya, Mexico | May 23-25 | In partnership with CONACYT



Kinect for Windows

An Update for Researchers

Stewart Tansley, PhD

Director, Natural User Interfaces
Microsoft Research Connections

May 24, 2012



Kinect for Windows

An Update for Researchers

The Kinect
for Windows
Story

Key Research
Technologies

Example
Research
Applications

Kinect for
Windows
Today

Natural User
Interface

Your Story

The Kinect Effect video



KINECT™
for Windows®



Kinect for Windows

An Update for Researchers

The Kinect
for Windows
Story

Key Research
Technologies

Example
Research
Applications

Kinect for
Windows
Today

Natural User
Interface

Your Story



The Kinect for Windows Story – So Far

June 2009:	Project Natal
November 2010:	Kinect for Xbox 360
June 2011:	Kinect SDK from Microsoft Research
November 2011:	The Kinect Effect video
February 2012:	Kinect for Windows 1.0 product
May 2012:	Kinect for Windows 1.5 product



Natal → Kinect

- Decades of research in computer vision and audio
- Xbox prototype, they called MSR in September 2008
- First announced June 1, 2009 at E3 conference
- Launched in North America on November 4, 2010
(rapidly followed by EU, Japan, Korea, Australia...and many more)
- 10 million sold (at March 9, 2011)
Guinness world record: fastest selling consumer electronics device of all time





Origins: Project Natal



Source: Wikipedia

- Named after the Brazilian city, meaning “relating to birth”
- *“The birth of the next generation of home entertainment”*
(Alex Kipman)
- Not just the device. The sensor provides “eyes and ears”...
→ **but it needs a brain**
- Raw data from that sensor is just *“a whole bunch of noise that someone needs to take and turn into signal”*
→ **that is what our software does: find the signal**



Kinect for Windows

An Update for Researchers

The Kinect
for Windows
Story

Key Research
Technologies

Example
Research
Applications

Kinect for
Windows
Today

Natural User
Interface

Your Story



The Problem

- Find the people in the scene, ignore background
- Find their limbs and joints, which person is which
- Find and track their gestures
- Map the gestures to meaning and commands
- *Also: recognize faces*
- *Also: recognize voices and commands*

P.S. And play the game!



Software Magic!

- Machine Learning
- Effectively...
 - Evaluate trillions of the possible body configurations of 32 body (skeletal) segments
 - Every video frame
 - 30 times a second
 - On <10% of the CPU



Behind the Magic

- Decades of computer vision research between industry and academia, including our own at Microsoft Research and Xbox
- State of the art in human body tracking in 2007 had the ability to track a wide range of motion
→ **but with limited agility and not in real-time**
- Xbox's requirement:
All motions, all agilities, 10x real-time, for multiple bodies!
- But Xbox did have a low-cost 3D camera...



Vision Algorithm – Paper



- CVPR 2011 Best Paper:

Real-Time Human Pose Recognition in Parts from a Single Depth Image

Jamie Shotton, Andrew Fitzgibbon, Mat Cook, Toby Sharp,
Mark Finocchio, Richard Moore, Alex Kipman, Andrew Blake

<http://research.microsoft.com/apps/pubs/default.aspx?id=145347>



Paper



Supplementary

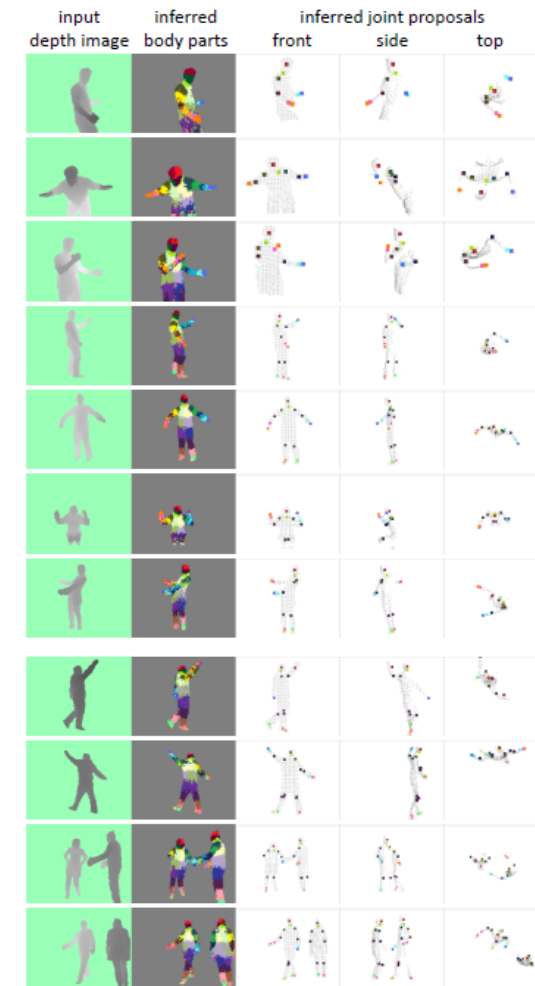


Video

<http://cvpr2011.org>

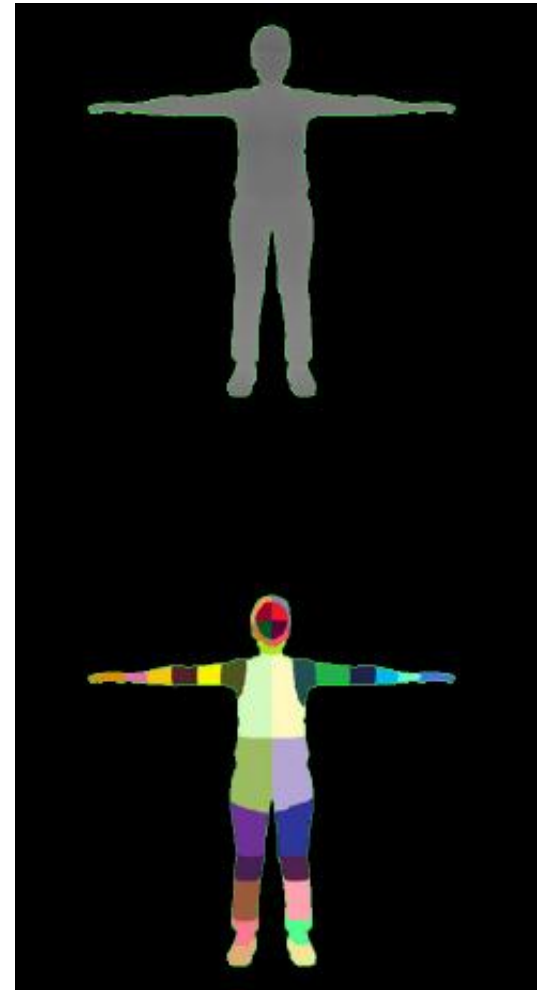
Vision Algorithm – Summary

- Quickly and accurately predict 3D positions of body joints
- From a single depth image, using no temporal information
- **Object recognition approach**
- Intermediate body parts representation that maps the difficult pose estimation problem into a simpler **per-pixel classification problem**
- Large and highly varied training dataset allows the classifier to estimate body parts invariant to pose, body shape, clothing, etc.
- Generate confidence-scored 3D proposals of several body joints by re-projecting the classification result and finding local modes
- System runs at 200 frames per second on consumer hardware
- Evaluation shows high accuracy on both synthetic and real test sets
- State of the art accuracy in comparison with related work and improved generalization over exact whole-skeleton nearest neighbor matching



Vision Algorithm – In Practice

- Collect training data – thousands of visits to global households, filming real users, the Hollywood motion capture studio generated billions of images
- Apply state-of-the-art object recognition research
- Apply state-of-the-art real-time semantic segmentation
- Build a training set – classify each pixel's probability of being in any of 32 body segments, determine probabilistic cluster of body configurations consistent with those, present the most probable
- Millions of training images → Millions of classifier parameters
- Hard to parallelize → New algorithm for distributed decision-tree training
- Major use of DryadLINQ (large-scale distributed cluster computing)

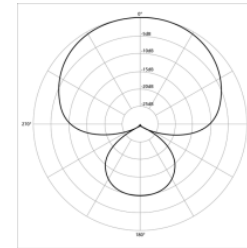


Don't Forget the Audio!

- 4 supercardioid microphone array in Kinect:



Source: Wikipedia



- See: 1-hour MIX conference presentation by Ivan Tashev:
<http://channel9.msdn.com/events/MIX/MIX11/RES01>
- "The talk will cover the overall architecture and algorithmic building blocks of the Kinect device, **especially the audio pipeline**. We will present the opportunities it opens for building better human-machine interfaces, new user experiences, and other potential applications. No specialized signal processing background is required."

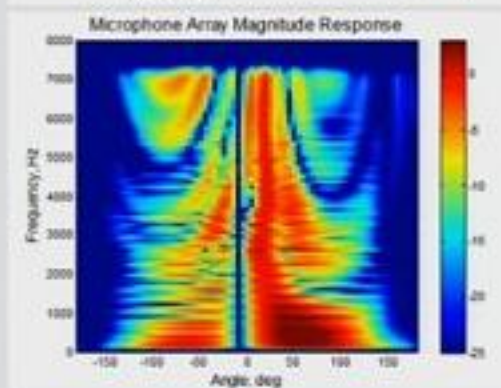
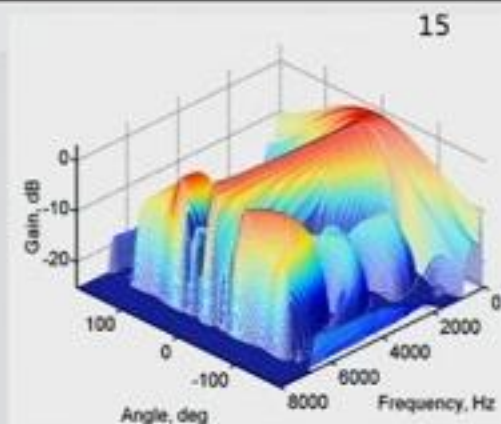
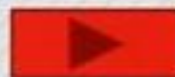
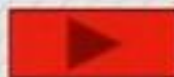
Ivan is the creator of most of the audio algorithms in the Kinect pipeline





Adaptive beamforming

- 1 On the fly computation of the weights
- 2 Higher CPU requirements
 - 3 Does null steering
- 3 MVDR beamformer
 - 4
$$\mathbf{W}_{MVDR}(f) = \frac{\mathbf{D}_e^H(f) \Phi_{NN}^{-1}(f)}{\mathbf{D}_e^H(f) \Phi_{NN}^{-1}(f) \mathbf{D}_e(f)}$$
- 5 Nulls can be enforced if known
- 6 Two microphone array demos





Kinect for Windows

An Update for Researchers

The Kinect
for Windows
Story

Key Research
Technologies

Example
Research
Applications

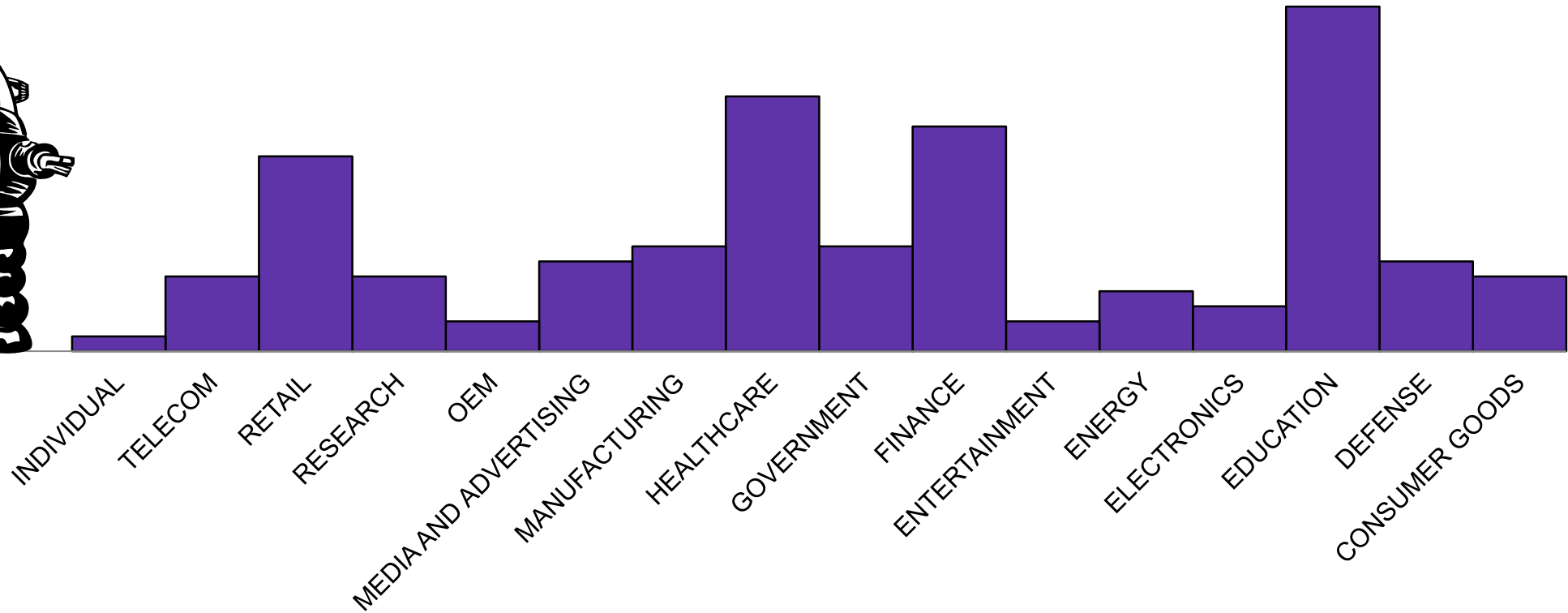
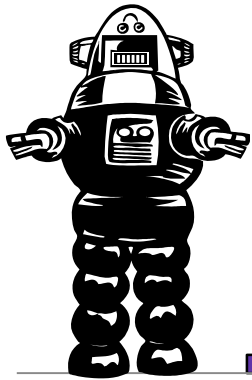
Kinect for
Windows
Today

Natural User
Interface

Your Story



Example Applications

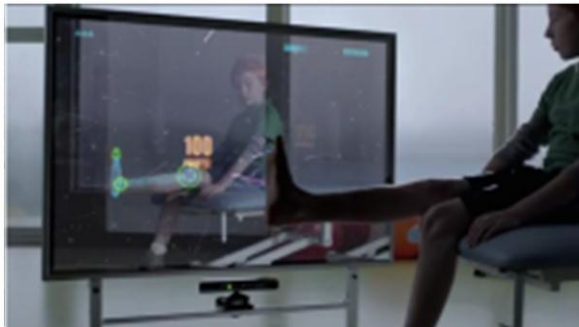


World Class Partners – Examples





Rising Key Scenarios



THERAPY



HEALTHCARE



TRAINING



RETAIL



Kinect for Windows

An Update for Researchers

The Kinect
for Windows
Story

Key Research
Technologies

Example
Research
Applications

Kinect for
Windows
Today

Natural User
Interface

Your Story



Where to get it

<http://kinectforwindows.com>

USE THE POWER OF KINECT TO CHANGE THE WORLD

Be at the forefront of innovation. Explore how Kinect for Windows transforms the way people interact with technology. Help unlock the possibilities.

PRODUCT FEATURES



Purchase

Learn where to purchase a Kinect for Windows sensor, and start developing today.

[BUY ONLINE >](#)

Discover

What's possible with Kinect for Windows? See how Kinect is being applied to fields beyond gaming.

[EXPLORE GALLERY >](#)

Develop

Download the SDK and Toolkit, along with access resources to help develop Kinect for Windows applications.

[DOWNLOAD SDK >](#)



International Availability

Now:

Australia, Canada, France, Germany, Hong Kong, Italy, Ireland, Japan, Korea, Mexico, New Zealand, Singapore, Spain, Taiwan and the United Kingdom.

June:

Austria, Belgium, Brazil, Denmark, Finland, India, the Netherlands, Norway, Portugal, Russia, Saudi Arabia, South Africa, Sweden, Switzerland and the United Arab Emirates.

Kinect for Windows – Today

SDK

Raw sensor streams

Near mode (40 cm)

Skeletal tracking for 1-to-2 users

Advanced speech and audio capabilities

Familiar development environment

Supports 4 Kinect sensors on a single PC

Sensor





What's New in v1.5 (May 2012)

Seated mode skeletal tracking

10 joint

Improved skeletal tracking

40cm near

Face tracking capabilities

3D mesh

Developer toolkit

Lots of good things!

Kinect Studio
debug

Record, playback,

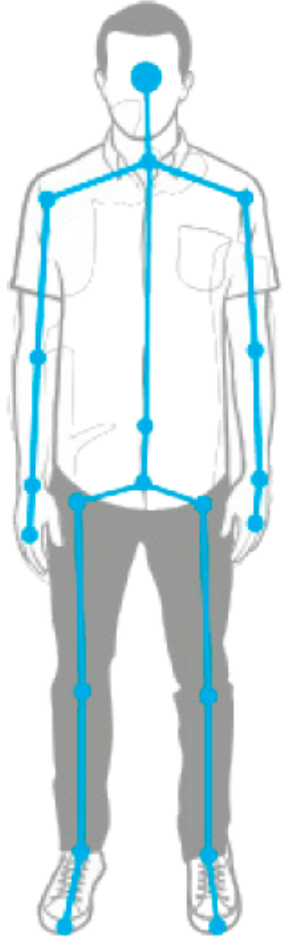
Speech recognition options

4+11 language regions



New Documentation

<http://go.microsoft.com/fwlink/?LinkID=247735>



HUMAN INTERFACE GUIDELINES

Kinect for Windows v1.5.0



HIG Extracts (70 pages total)

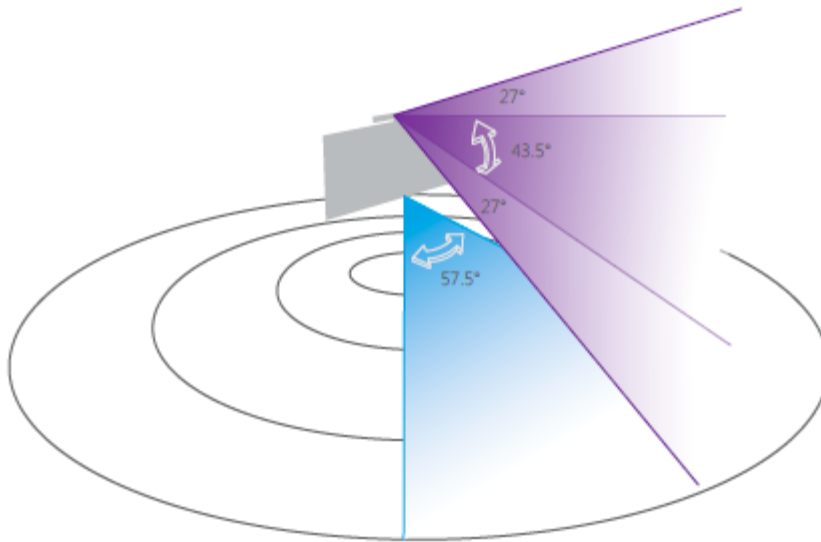
Physical Capabilities

Angles of Kinect vision (Depth and RGB)

Horizontal: 57.5 degrees

Vertical: 43.5 degrees with

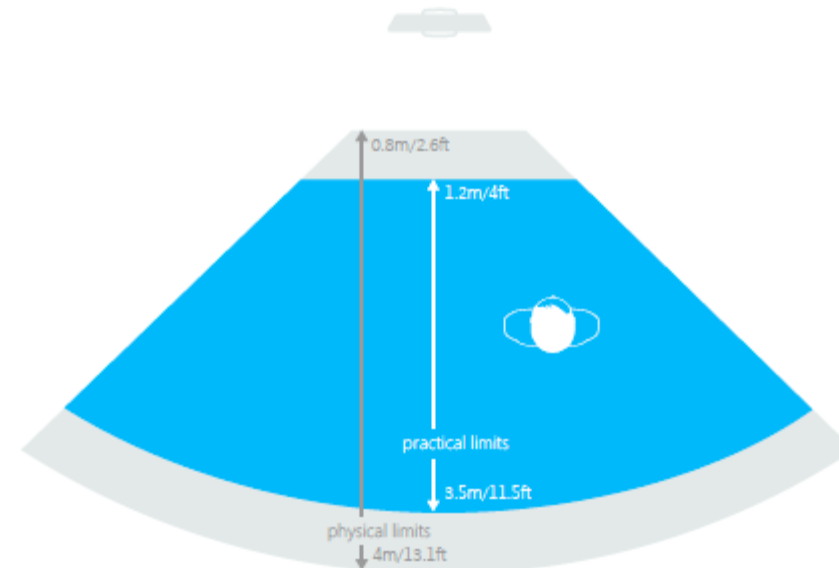
-27 to +27 degree tilt range up and down



Distance ranges for Depth (default mode)

Physical limits: 0.8 to 4m

Practical limits: 1.2 to 3.5m

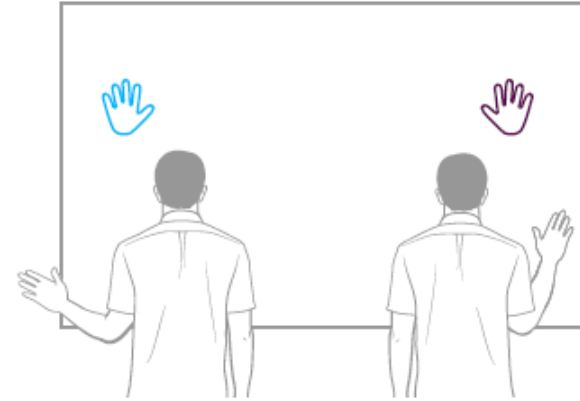




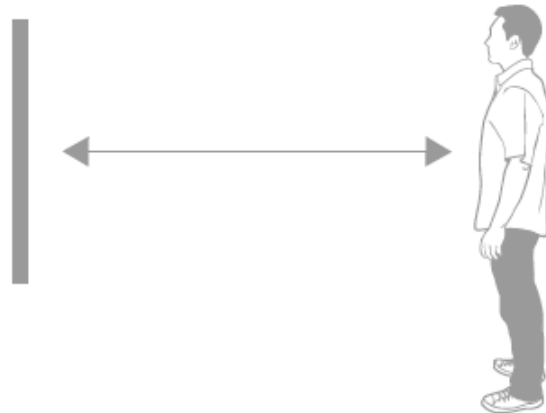
HIG Extracts



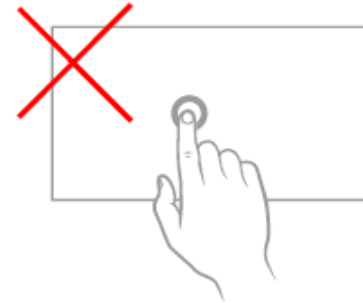
Keep in mind that gesture use should be purposeful



Multiple users will be interacting with your application at once

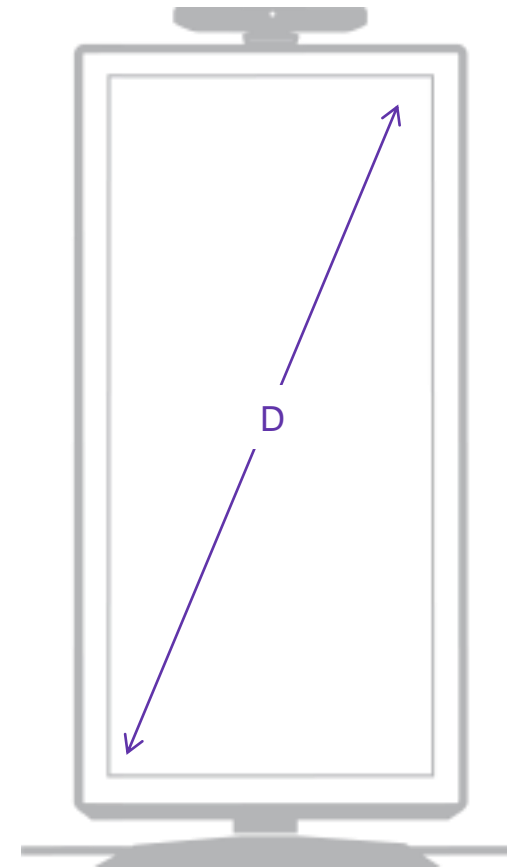
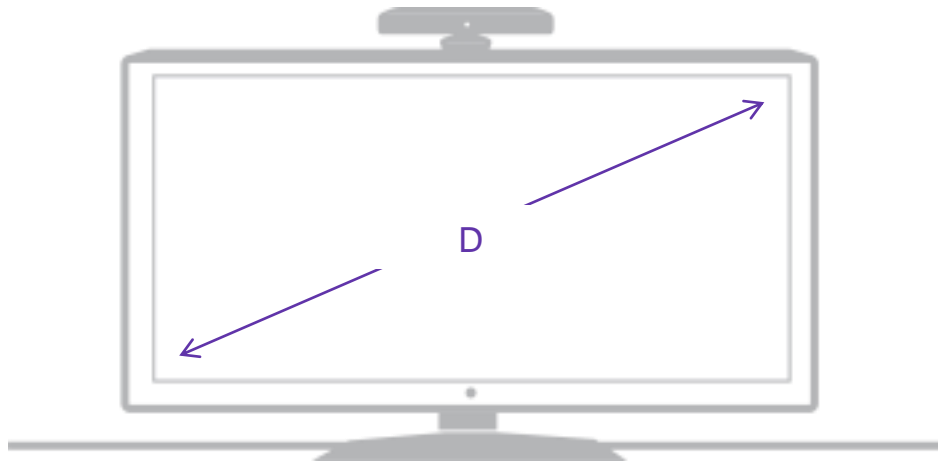


Scenario requires users to interact from a distance



Your scenario requires that users not touch anything directly

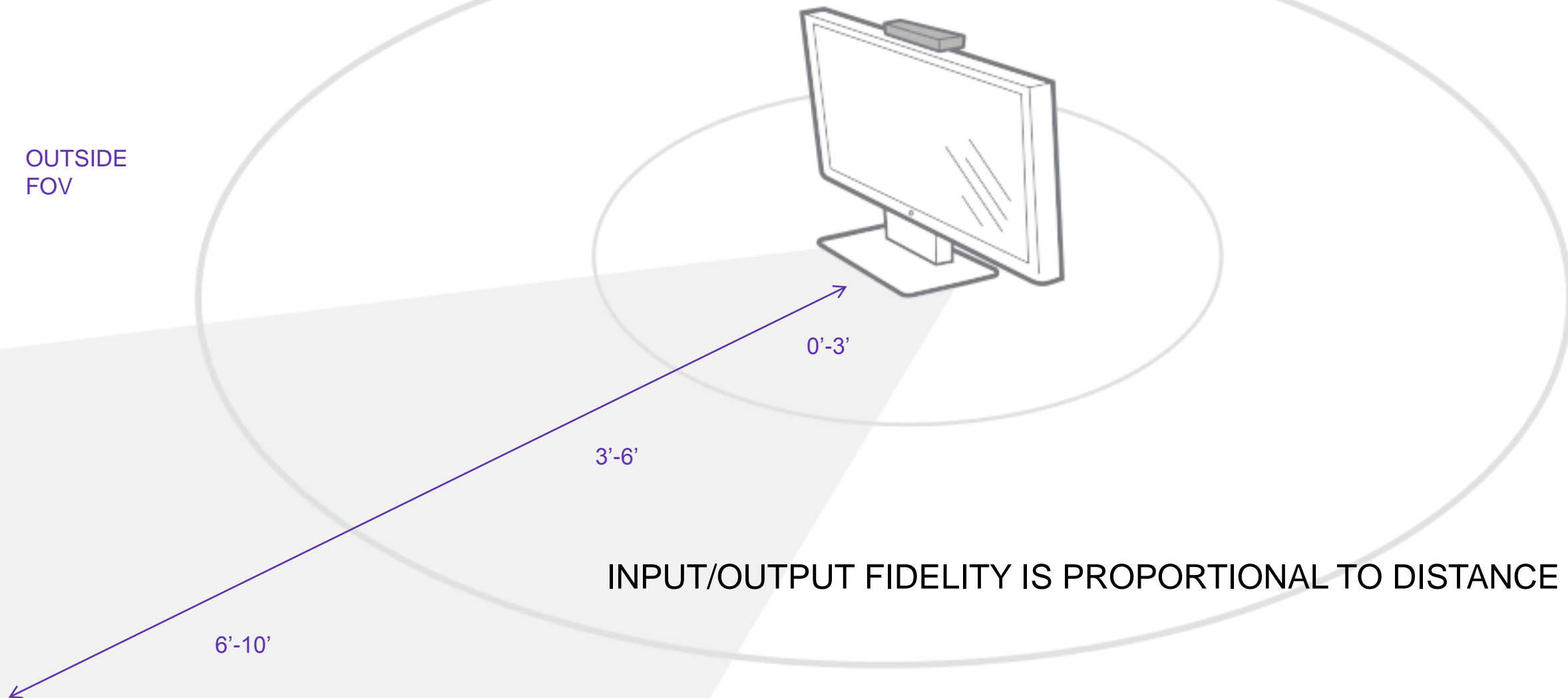
DISPLAY SPECTRUM



MAP INPUT AND OUTPUTS TO SCREEN SIZE AND ORIENTATION

PROXIMITY SPECTRUM

OUTSIDE
FOV



0'-3'

3'-6'

6'-10'

INPUT/OUTPUT FIDELITY IS PROPORTIONAL TO DISTANCE

POSTURE SPECTRUM

SEATED



STANDING



ACTIVE



MAP INPUTS/OUTPUTS TO ACTIVITY LEVEL

ORIENTATION SPECTRUM

AUDIOVISUAL



AUDIOVISUAL

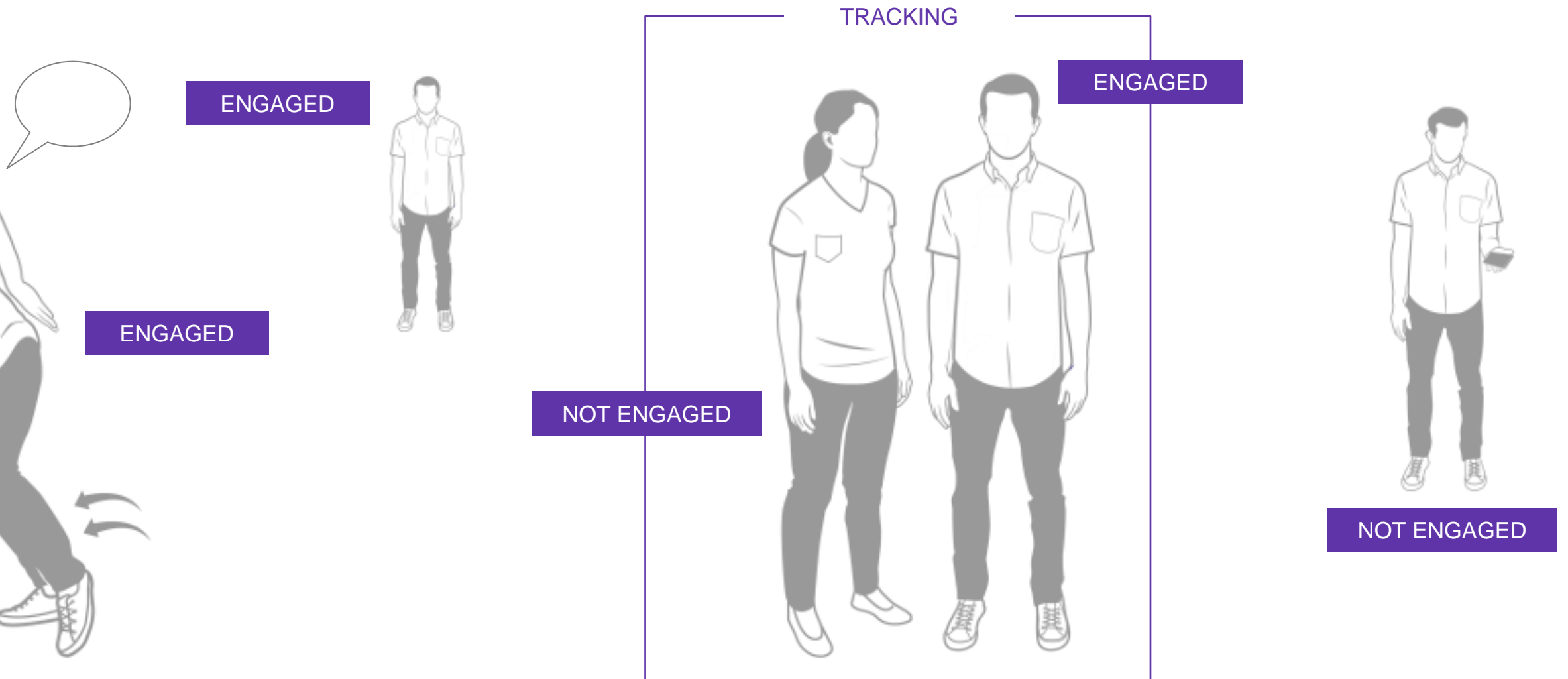


AUDIO



MAP OUTPUTS TO ORIENTATION

ENGAGEMENT SPECTRUM





Kinect for Windows

An Update for Researchers

The Kinect
for Windows
Story

Key Research
Technologies

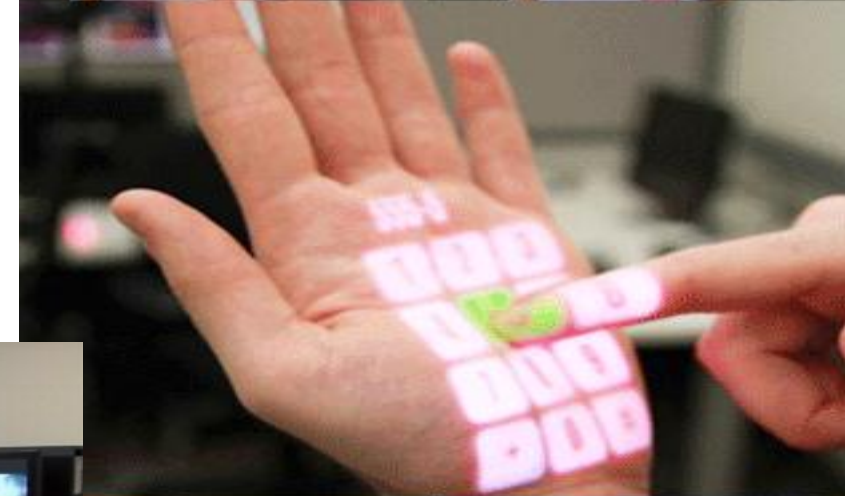
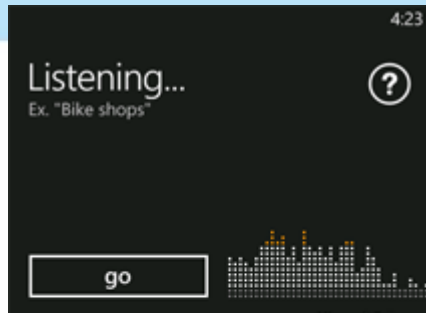
Example
Research
Applications

Kinect for
Windows
Today

Natural User
Interface

Your Story

Some directions in Natural User Interfaces



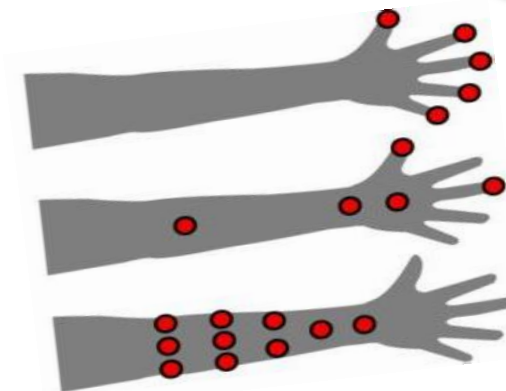
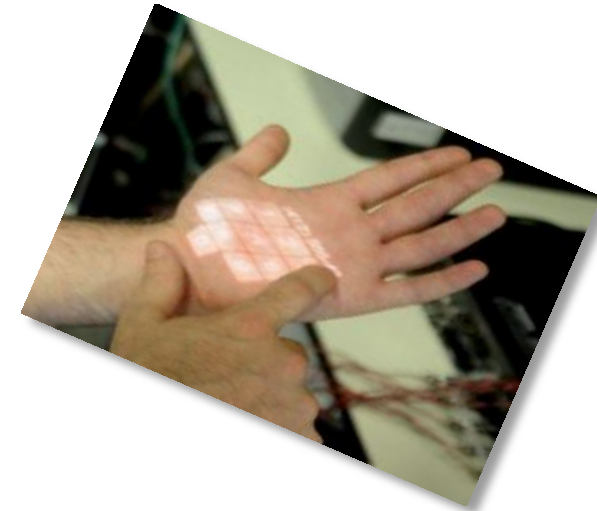
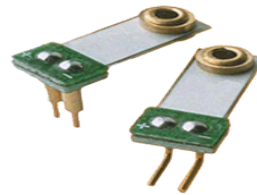
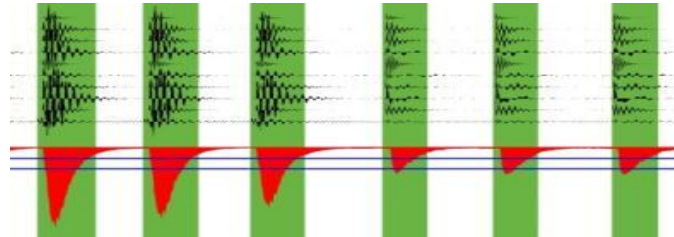
*How do we make
computing always,
instantly, intuitively
available?*





Skinput Project

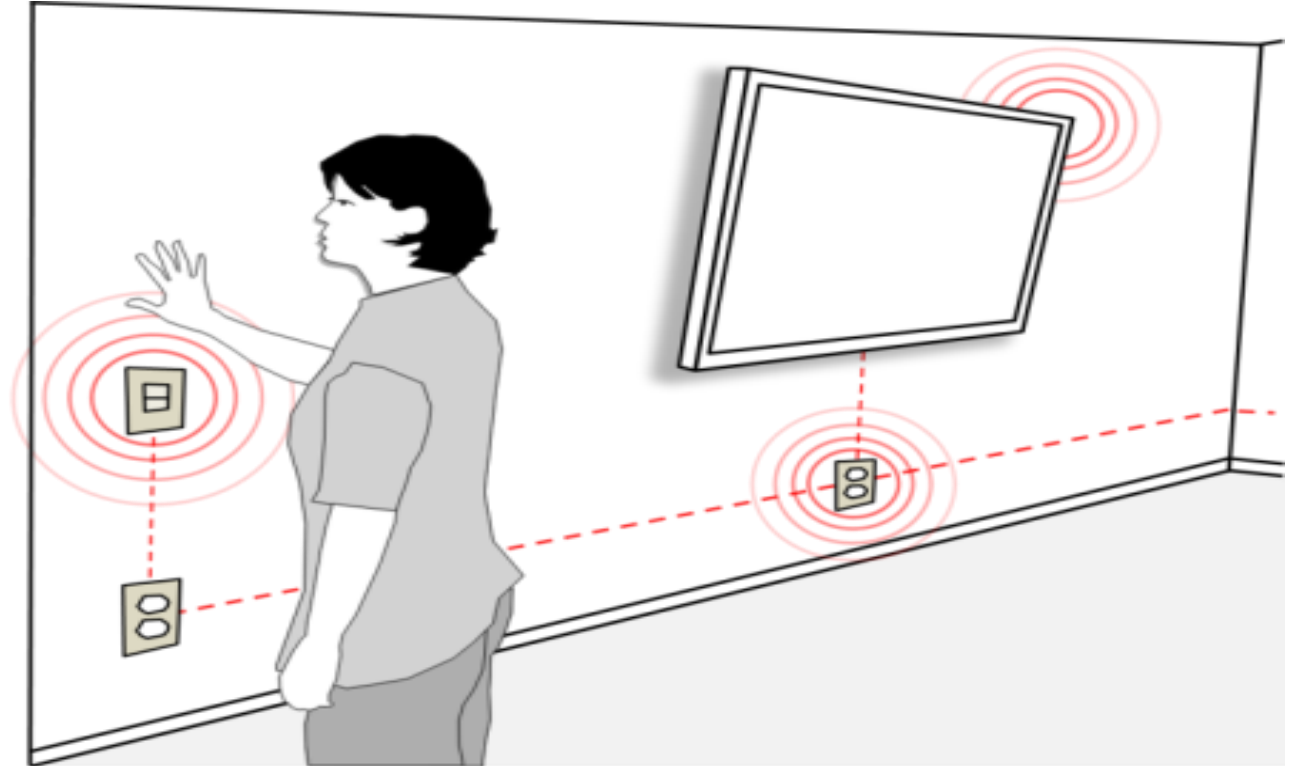
<http://research.microsoft.com/en-us/um/redmond/groups/cue/skinput/>





Humantenna Project

<http://research.microsoft.com/en-us/um/redmond/groups/cue/humantenna/>





Sensors & Devices Group

<http://research.microsoft.com/en-us/groups/sendev/>





Natural Interaction Group

<http://research.microsoft.com/en-us/groups/natural/>





Computer Vision Group

<http://research.microsoft.com/en-us/groups/vision/>



Medical image analysis



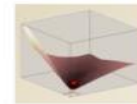
i2i: 3D visual communication



Image and video editing



Object class recognition



Discrete optimization in vision



C-Slate for remote collaboration



Geometric modelling from images



Visual tracking





And Many More...

<http://research.microsoft.com>

<http://research.microsoft.com/NUI>



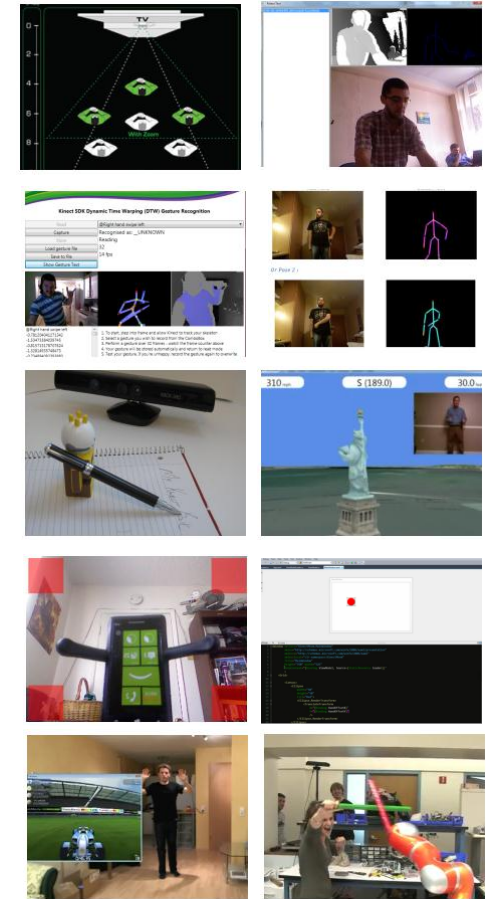
More Ideas – Community

Channel 9 Kinect Community

<http://channel9.msdn.com/coding4fun/kinect/>
Over 250 posts!

Imagine Cup

<http://www.imaginecup.com/>





Kinect for Windows

An Update for Researchers

The Kinect
for Windows
Story

Key Research
Technologies

Example
Research
Applications

Kinect for
Windows
Today

Natural User
Interface

Your Story



Your Story

What we've covered

The Kinect for Windows story – just so far...

Some of the key research technologies behind Kinect

Example applications and opportunities

Kinect for Windows product v1.5

Some directions in Natural User Interfaces

→ What is ***your*** research story with Kinect for Windows?



Kinect for Windows

An Update for Researchers

The Kinect
for Windows
Story

Key Research
Technologies

Example
Research
Applications

Kinect for
Windows
Today

Natural User
Interface

Your Story



Resources

<http://kinectforwindows.com>

<http://channel9.msdn.com/coding4fun/kinect/>

<http://microsoft.com/education/facultyconnection>

<http://research.microsoft.com>

<http://research.microsoft.com/NUI>

stansley@microsoft.com

<http://research.microsoft.com/~stansley>

@dswtan

#KinectWindows

The Microsoft logo is displayed in a large, bold, white, italicized sans-serif font. The letters are closely spaced, and the 'M' and 'S' are particularly prominent. A registered trademark symbol (®) is located at the top right of the 't'.

© 2012 Microsoft Corporation. All rights reserved. Microsoft, Windows, and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.

The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.