Multimodal Interaction in Virtual Reality

Ipke Wachsmuth
Faculty of Technology
University of Bielefeld

Anthropomorphic Interfaces
- Interfaces following human form
- natural human-machine communication
- highly popular in interactive multimedia
- gestures, speech, coverbal gesture

Communicating in the situation
Central aspects in our work:
- Multimedia
- Multimodality
- Multiple representations
- Multi-agent Systems

Drawbacks in VR
- Tedious usage of synthetic models
- Deficits in interactive design support
- Manipulation not subject to physical laws
- Lack of user-friendliness in interaction

Our approach: indirect management techniques, based on gesture and speech
Three things are important in our work toward incorporating gestures as a useful tool in virtual reality:

• measuring gestures as articulated hand and body movements in the context of speech

• interpreting them by way of classifying features and transducing them to an application command via a symbolic notation inherited from sign language

• timing gestures in the context of speech in order to establish correspondence between accented behaviors in both speech and gesture channels
Gesture Recognition

- **Approach 1:** Basic interactions, direct feedback
  - Pointing
  - Grasp
  - GraspRelease
  - Rotation
  - Translation

- **Approach 2:** Formbased description
  - explicit (knowledge-based) approach
  - atomic form elements of gesture are composed to 'gesture words'

Symbols for Body Parts

Gesture notation with HamNoSys

For form description of gestures we work with a scenario-specific subset of HamNoSys (HNS').
From gesture to application

- **Actuators**
  - abstract placeholders of significant discrete features in body reference system; normalized (e.g., world coordinates)

- **Motion modifiers**
  - bind temporarily to actuators and filter motion data to object transformations (via manipulators)

- **Manipulators**
  - receive transformation commands and put them into effect in the 3D scene

**Multimodal Interface:**

An interface which allows the (sequential or concurrent) usage of multiple modalities

**Open input:**

Begin and end of interaction not known
Gestures: Kinesic Structure

Consistent Arm Use and Body Posture

Consistent Head Movement

Gesture-Unit

Gesture-Phrase

Preparation

Stroke

Retraction

Hold (pre-stroke)

Hold (post-stroke)

After McNeill, Levy & Pedelty (1990)

Timing of gestures and speech

- The gesture stroke is often marked by an abrupt stop which is correlated with accented words or syllables
- The stroke does not occur after an accented word but simultaneously or shortly before

⇒ hypotheses for establishing correspondence between accented behaviors in speech and gesture channels

Stroke Cues

- pre/post-stroke hold (kinesic structure)
- strong acceleration of hands, stopps, rapid changes in movement direction
- strong hand tension
- symmetries in two-hand gestures

Multimodal Integration

Two "logistic" problems to be solved (Srihari, 1995):

- **Segmentation Problem**
  How can a system be made to cope with "open input"? How can units be determined to be processed in one system cycle?

- **Correspondence Problem**
  How to determine cross-references between multiple modalities (speech/gesture)?
Earlier work: VIENA (1996)

Virtual Construction
Work with Bernhard Jung and Martin Hoffhenke

• Innovative approach to prototype the design of mechanical devices and explore their assembly in early stages of industrial development, based on CAD models

The things we have learned from investigating these issues help us to advance natural interaction with 3D stereographic scenes in a scenario of virtual construction.

Multimodal Interaction in VR
Work with Marc Latoschik, SGIM Project

Speech and Gesture Interfaces for Multimedia
In the first place we have dealt with pointing and turning, etc., commonly classified as deictic and mimetic gestures.

Where we would like to go next:

- Virtual Workspace (two-sided 3D projection)
- Manipulative gesture (grasp space) and
- Communicative gesture (distant space)

DEIKON (2000++)

- Systematic study of referential acts by coverbal gesture
- Aim: To investigate how complex signals originate from speech and gesture and how they are used in reference
- Contribution of gestural deixis for making salient or selecting objects and regions
- Making an artificial communicator to understand and produce (coverbal) deictic gestures in construction dialogs
Deictics & Iconics

Work with Timo Sowa and Marc Latoschik

In the DEIKON project, we have now started to research into more sophisticated forms of deictics in construction dialogues that include features indicating shape or orientation, which leads us into iconic gesture.

Lifelike Gesture Synthesis

Work with Stefan Kopp, Articulated Communicator

Pull-1
\( (\text{PARALLEL} \ (\text{Start} \ 1, \ 0) \ (\text{End} \ 2, \ 0)) \)
\( (\text{SEQUENCE} \ (\text{Start} \ 1, \ 0) \ (\text{End} \ 2, \ 0)) \)
\( (\text{PARALLEL} \ (\text{Start} \ 1, \ 0) \ (\text{End} \ 17, \ 0)) \)
\( (\text{STATIC} \ (\text{Start} \ 11, \ 0) \ (\text{End} \ 17, \ 0) \ (\text{HandLoc})) \)
\( (\text{STATIC} \ (\text{Start} \ 1, \ 0) \ (\text{End} \ 17, \ 0) \ (\text{HandShap})) \)
\( (\text{PARALLEL} \ (\text{Start} \ 1, \ 0) \ (\text{End} \ 2, \ 0)) \)
\( (\text{static} \ (\text{Start} \ 1, \ 0) \ (\text{End} \ 2, \ 0) \ (\text{HandLoc})) \)
\( (\text{STATIC} \ (\text{Start} \ 1, \ 0) \ (\text{End} \ 2, \ 0) \ (\text{HandShapeBSlist})) \)
\( (\text{STATIC} \ (\text{Start} \ 1, \ 0) \ (\text{End} \ 2, \ 0) \ (\text{PalmOrientation}) \) \)

Another issue in our work is the synthesis of lifelike gesture from symbolic descriptions for an articulated virtual figure where natural motion and timing are central aspects.

Find Lab Showcase & papers...

Technische Fakultät
Labor für Künstliche Intelligenz & Virtuelle Realität

www.techfak.uni-bielefeld.de/techfak/ags/wbski/
www.techfak.uni-bielefeld.de/~ipke/

WBS Team

Bernhard Jung
Stefan Kopp
Timo Sowa
Ian Voß
Marc Latoschik

Martin...

...Hoffhenke
...Fröhlich