Signing Avatar Work at UEA

John Glauert, Richard Kennaway, Vince Jennings, Ralph Elliott

SLTAT Berlin 2011-01

Introduction

- VHG – Virtual Humans Group,
  School of Computing, UEA Norwich

  `http://vhg.cmp.uea.ac.uk/`

- Work on Signing Avatars
  - initially (late 1990s) based on Motion Capture
  - since 2000: focus primarily on *synthetic signing*
    - algorithmically generated animation parameters

Synthetic Signing System

- ViSiCAST project prototype system
- Spoken language text to SL performance by signing avatar
- Organised as two major stages
  - Natural Language Processing
  - Synthetic Signing
- Interface between stages: phonetic level SL sequence description — SiGML

Animgen

- *Animgen*: Central component of synthetic signing system
- Primary input: SiGML description of required SL sequence
- Output: animation parameters for performance by the given avatar of the given SL sequence
- Subsidiary inputs:
  - Data files describing physical characteristics, configuration parameters, of target avatar
- So input is avatar-independent, but output is avatar-specific

SiGML

- "Signing Gesture Markup Language" (XML application)
- Based on HamNoSys
  — i.e. based on HamNoSys model of SL articulation
- Various flavours:
  - HNS SiGML: HamNosys symbol names in <. . ./>
  - Gestural SiGML: "structured HamNoSys data"
    (somewhere between concrete and abstract syntax tree)
  - Also allow animation data
    (per-frame joint angles and morph factors)

SiGML — Manual and Nonmanual

- Manual HamNoSys 4 =
  Manual HamNoSys 2/3 with limited additions

- Nonmanual HamNoSys 4:
  newly introduced repertoire of nonmanual actions
  - Several tiers: shoulders, body, head, eye-gaze,
    facial expression, mouthing (with sub-tiers for the last two)
  - use SAMPA codes for speech mouthing
  - Limited structure — essentially a list of actions

Avatar Technology — Standard

- Skeleton — hierarchy of virtual bones
- Bound to textured surface mesh
- Set of morph targets — support facial nonmanuals

Avatar Technology — Non-standard

- SL Feature Points: a set of named locations (approx. 400) on the avatar's surface mesh that are needed for signing
- i.e. points to which HamNoSys locations can be mapped (directly or indirectly)
- Use our own *ARP Toolkit* to generate Feature Point definitions for each avatar
- Feature Point definitions are all that Animgen knows about the mesh

Animgen Implementation Strategy

- Manual component:
  determine required key postures for hands, use IK to determine frame-by-frame sequences of shoulder, arm, hand movements needed to reach those postures

- Facial nonmanuals:
  use avatar morphs, driven by configuration file mapping HamNoSys nonmanuals to morph timing and weighting characteristics

- Bodily nonmanuals:
  use prepackaged sequences of skeleton adjustments
  ("pseudo-morphs")

Current Software Package — JASigning

- Java-based system, with Animgen native library via JNI
- SiGML-driven avatar panel:
  - in web pages, controlled by Javascript
  - in desktop apps, downloadable via JNLP
  - SiGML input: direct text or via a URL
  - real-time operation – allows interactive apps as long as required SiGML sequences can be promptly generated

`http://vhg.cmp.uea.ac.uk/tech/jas/std/`

Improvements to Current Animation System

- A few HamNoSys 4 manual features not covered — e.g. orientation-relative motion, brushing contacts
- Better collision detection (primarily for hands)
- More natural motion dynamics
- HamNoSys ambiguity resolution (labour of Sisyphus!)
  — issues about level of precision in SL transcription

Authenticity Issues

- Photorealism? — Not obviously essential
- Important to use qualified graphic designer to define avatar's visible features
- Do well-made morphs support a sufficient range of facial expressiveness?
- Morphs for speech
  - usually driven entire synthetically in our system
  - we have also used morph weights generated from speech data

Dicta-Sign Project

- Enhance/extend SiGML to support higher levels of precision and expressiveness in synthetic signing.
- Dicta-Sign is collecting a parallel corpus of SL material.
- We depend on higher level modelling of SL performance based on corpora such as this to determine how to exploit the extended SiGML Model.

PDTS SiGML

- Influenced by Liddell-Johnson model of SL phonetics
- PDTS = Posture - Detention - Transition - Steady shift
- expressive range is superset of HamNoSys
- allows looser specifications than HamNoSys
- permit continuous rather than discrete ranges for features such as orientation, distance, location, timing
- allow explicit control of previously hidden avatar configuration parameters — scale and location of signing space

PDTS SiGML — Issues

- SL-specific notation vs. general notation for human action
- Do nonmanual aspects of SL performance naturally fit the rigid alternating timing framework (P/T/P/T/...)?
- Need to define higher-level models of linguistic and non-linguistic communication and their mappings to the PDTS model.
- How is PDTS content to be generated?
  HamNoSys's attractiveness as an input for synthetic signing systems is increased by the the fact that there are already significant corpora of HamNoSys transcriptions — and tools.