DADA Version Longue

Description des taches

Rémi Ronfard, 25 mars 2015

WP1 Kinesics

Task11 – Choosing representations for full body motion ; representation learning ; transfer learning ; parameterisation of kinesic components. This can include neuro-muscular variables ; also the choice of kinematic trees rooted at the head ; also the grouping of kinematic variables into synergies ; etc.

One classical distinction is between world-frame positions and joint angle variables In our case, we are making a strong statement that we will study kinesic variables for all joints relative to the rigid body frame associated with the actor. This could be the ground floor position of the actor plus a rigid body position associated with the actor’s head. Thus proxemic variables could be footsteps and head movements ; kinesic variables could be all other joint angles or joint positions in world coordinates.

Task12 – Learning « bi-dimensional» models of actions and moods from a mocap dataset of six actions (walk, carry object, knock on door, throw object, lift object, move object) in 8 moods (neutral, happy, afraid, angry, anxious, sad, proud, shameful). Ideally, we would like to separate the action components from the mood components of motion and extrapolate the moods to other actions, and the actions to other moods.

Task13 – Learning models of gesture and facial expressions in dialogue situations – based on INRIA’s previous work on visual prosody, we would like to learn joint models of gesture and speech prosody. Ideally, this should be done without MOCAP data, using only audio and video processing, possibly enhanced with depth (kinect).

Task14 – Learning parameterized kinesic models – because our models are contextual, conditioned by the proxemic components, we shuld be able to change their velocity, amplitude, direction and phase (in the way of « parametric HMM models »). This is challenging and needs to be investigated.

WP2 Proxemics

Task 21 : Group behaviors during multi-way conversation ; turn-taking ; synchronization of actors in dialogue ; implementation of conversational phenomena (interruption, pause, re-run, repetition, etc.)

Task 22 : Group behaviors during stage movements – implementation of advanced « steering behaviors » such as follow, flee, separate, join, merge, enter stage, exit stage, etc.

Create better models by taking gaze and head movement into account ; Physical models of social interactions ; implementation of action-reaction chains between actors ;

Task 23 : High level models of actions and reactions, timing, implementation of theatrical effects. Provoke surprise and/or expectation, etc. Models of tri-partite interaction between two actors and the audience in dialogue and in movement. Theatrical cheating techniques. ?

Task 24 – Combination of statistical and procedural models ; smooth transitions between action keeping the same mood; between moods keeping the same action ; smooth transitions between dialogue and action ; adaptation to proxemic contexts.

WP3

Task31 Specification of a dramatic language of verbs (actions, speech acts, movements) and adverbs (moods, attitudes, dramatic effects) for directing actors ; define cues as synchronisation points between actors ; define parallel and sequential behaviors ; etc.

Part of this language will be devoted to stage blocking / movement

Part of this language wil be devoted to dialogue

Task32 Compilation of the language into a finite state machine and/or Petri net ; allowing real-time execution of the dramatic score ; user interface for directing actors by sketching stage floor plans and composing the dramatic score ; one line per actor per motion component (proxemic behaviors, kinesic actions, kinesic moods, speech acts, etc.)

Task33 Real-time execution of the dramatic score ; real-time combination of proxemic (procedural) and kinesic components of motion ; non-deterministic motion generation ; synchronization to cues ; real-time skinning and advanced 3D animation ; integration of physically-based secondary animation (skin, hair, clothes, etc.)

This includes integration of the GRETA BML realizer with IMAGINE animation ; and real-time integration of the statistical models of motion with the procedural animation components.

Task 34 User adaptation– allow director to add corrections and let system learn the differences using imitation learning or active learning or other related techniques ; etc.

WP4

Task42 Writing scenes with didascalia

Dialogue scenes with groups of 2, 3 and 4 actors using a choice of didascalia

Silent stage movements, as in opera synched to music, using a choice of didascalia

Alternations of dialogue and stage movements in theatre scenes with 2 actors

Task43 Validation of the interaction

Is the dramatic language adequate ? useful ? efficient ?

Is the dramatic score interface adequate ? useful ? efficient ?

Is the stage floor plan sketching tool adequate ? useful ? efficient ?

Task42 Validation of the animation

* Dialogue scenes
* Silent stage movements of groups of 2, 3 and 4 actors, as in opera synched to music

END OF TASK DESCRIPTIONS

References

1) If needed, we could try to learn models of gesture and proxemics from video along the lines of

Sergey Levine, Philipp Krahenbuhl, Sebastian Thrun, Vladlen Koltun. Gesture Controllers. ACM SIGGRAPH 2010. 

Using our previous work in actor and action recognition, including

Daniel Weinland, Remi Ronfard, Edmond Boyer. Automatic Discovery of Action Taxonomies from Multiple Views. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) - 2006. 

D Weinland, E Boyer, R Ronfard. Action recognition from arbitrary views using 3d exemplars. International Conference on Computer Vision, 2007. ICCV 2007.

V Gandhi, R Ronfard. Detecting and naming actors in movies using generative appearance models. Computer Vision and Pattern Recognition (CVPR), 2013.

2) Learning head-driven kinesic animation could be done following the line of

Sreenivasa, Soueres, Laumond, Berthoz ; Steering a humanoid robot by its head. Intelligent Robots and Systems, 2009.

3) Combining proxemics and kinesics components could be done along the lines of

Hironori Mitake, Kazuyuki Asano, Takafumi Aoki, Marc Salvati, Makoto Sato, Shoichi Hasegawa : 'Physics-driven Multi Dimensional Keyframe Animation for Artist-directable Interactive Character', Computer Graphics Forum, Vol.28, No.2, pp.279-287, 2009.

References

* ["Natural Voices"](http://www.nextup.com/attnv.html) from AT&T ($35.00)
* ["Verbose"](http://www.nch.com.au/verbose/) from NCH ($29.99)
* ["TextAloud"](http://nextup.com/ta3purchase.html) from NextUp ($29.95)
* ["RealSpeak"](http://sites.fastspring.com/nextup/product/realspeak?store=parameters) from Nuance/ScanSoft ($45.00)
* ["Acapela"](http://sites.fastspring.com/nextup/product/acapela) from Acapela Group ($35.00)
* ["Cepstral"](http://www.nextup.com/redir/cepstral.html) from Cepstral LLC ($29.99)

There are also:

* ["CrazyTalk6 PRO"](http://www.reallusion.com/purchase_ct.asp) from Reallusion ($149.95)
* ["IVONA Voices"](http://www.ivona.com/voices.php) from ivona ($45)
* ["Dragon Premium"](http://www.nuance.com/for-individuals/by-product/dragon-for-pc/index.htm) from Nuance ($199.99)
* ["VoiceText"](http://www.neospeech.com/tts-engine.aspx) from NeoSpeech (about $74.95)
* ["FonixTalk"](http://www.fonixspeech.com/tts.php) from Fonix Speech (about $45.00)
* ["NaturalReader"](http://www.naturalreaders.com/order_information.htm) from NaturalSoft ($199.50)
* ["Cerevoice"](https://www.cereproc.com/) from Cereproc ($400)

# HMM-based Speech Synthesis System (HTS)

<http://hts.sp.nitech.ac.jp/>