Requirements for a gesture specification language
A comparison of two representation formalisms

Alexis Heloir and Michael Kipp

firstname.surname@dfki.de

Fig. 1. This figure depicts three gestures which are challenging for existing specification languages: secondary wriggling movement (a), contact for time points (b) and duration (c).

Describing human movement is a challenging task, given the many degrees of freedom of the human body. When using embodied agents in a human-computer interface context, the need for such a description often arises so that e.g. gestures can be formally specified to be then faithfully rendered by an animation engine.
The design of a gesture description language is determined by three factors: the producer (authoring module) of the language wants it to be expressive and easy to use, the consumer (animation module) wants it to be complete, precise, error-free, and convenient to interpret and an underlying theory usually directs the language design.
The behavior markup language (BML) [1] offers such a specification. However, the current version of BML focusses on the problem of temporal synchronization between modalities, whereas the question of how to describe the surface form of a gesture is still open. In order to get a better understanding for how BML must be extended toward a complete specification of gestural form, we compare two powerful existing formalisms for specifying human gestures. The first one, MURML, has been designed to specify coverbal gestures for an embodied conversational agent [2]. The second one has been designed to describe French sign language [3] and will be called LV in the further discourse. Both models have a similar theoritical background: sign language phonology. MURML bases some gesture description elements on HamNoSys [4]. LV is based on the Movement-Hold model by Liddel et Johnson [5].

Both languages use a tree-like representation to represent longer stretches of movement as a decomposition of smaller movement units. In MURML an *utterance* is the top-level unit that packages both speech and accompanying body movement. In LV, the *sign* is the top-level unit that packages a sequence of movements, naturally without reference to spoken language.

As an example, an iconic gesture, depicted in Fig. 1 (c), is encoded in both languages. In this gesture, the left hand traces a circle on the palm of the right hand. There is continuous contact between the index finger tip and the palm. This contact specification challenges both languages as highlighted in Fig. 2. Now follows A brief discussion of the two languages.

MURML offers a powerful mechanism to specify arbitrary complex motion trajectories with one or two hands. However, the resulting representations are deeply nested and therefore, hard to interpret. Also, the *parallel* tag seems to be used for two purposes: first, to structure the relationship of two hands moving in parallel, and second, to package the atomic

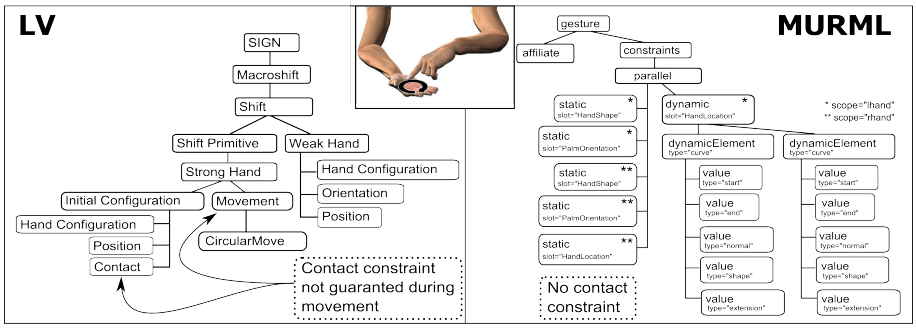

Fig. 2. Although powerful, presented specifications fail to convey some aspects of gesture.

motion of a single hand.
LV uses a flatter representation for movement but is less generic than MURML. However, it offers a feature to represent *contact gestures*. Such gestures can be critical for animation engines to render if it is unclear whether contact is desired or not. On a more design-centric aspect, LV carries some elements that are too dependant to its underlying theory (e.g. grammatical modifier element). Furthermore, we believe that a specification language should provide a way to specify arm swivel, which is an easily recognisable expressivity clue. As stated in the following summarizing table, none of the presented language take such parameter into account.

Aspect	LV	MURML	Conclusion
Complexity of representation	Relatively flat structure	Deep, nested structure	Use flat structure
Redundancy	Start/end poses	Start/end poses in dynamic constraints	Move to a pose-based representation
Arm swivel	---	---	Add feature for hand-arm configuration
Conceptual problems	Theory-dependent "modifier" element	Functional overloading of "parallel" tag	Rename "parallel", exclude "modifier"
Contact over period of time	Specify for start and end poses	---	Add static contact constraints
Secondary movement	Predefined set of secondary motions	Generic way to define e.g. hand shape changes	Allow both (e.g. provide MURML templates)

Fig. 3. Summary of some limitations we would like to address in future BML extension.

References

1. Kopp, S., Krenn, B., Marsella, S., Marshall, A.N., Pelachaud, C., Pirker, H., Thórisson, K.R., Vilhjálmsson, H.: Towards a common framework for multimodal generation: The behavior markup language. In: Proc. of IVA-06. (2006)
2. Kranstedt, A., Kopp, S., Wachsmuth, I.: Murml: A multimodal utterance representation markup language for conversational agents. Proceedings of the Workshop Embodied Conversational Agents, Autonomous Agents and Multi-Agent Systems (AAMAS02), Bologna, Italy (2002)
3. Losson, O., Vannobel, J.M.: Sign specification and synthesis. In: Proc. of Gesture Workshop. (1999)
4. Prillwitz, S.L., Leven, R., Zienert, H., Zienert, R., T.Hanke, Henning, J.: HamNoSys. Version 2.0. International Studies on Sign Language and Communication of the Deaf (1989)
5. Liddell, S.: American sign language : The phonological base. Sign language studies **64** (1989) 195–278