Compound Gesture Generation: A Model Based on Ideational Units

Yuyu Xu[1], Catherine Pelachaud[2], and Stacy Marsella[1]

[1] Northeastern University, Boston, MA 02115, USA
[2] CNRS - LTCI, Telecom ParisTech, France

Abstract. This work presents a hierarchical framework that generates continuous gesture animation performance for virtual characters. As opposed to approaches that focus more on realizing individual gesture, the focus of this work is on the relation between gestures as part of an overall gesture performance. Following Calbris' work [3], our approach is to structure the performance around ideational units and determine gestural features within and across these ideational units. Furthermore, we use Calbris' work on the relation between form and meaning in gesture to help inform how individual gesture's expressivity is manipulated. Our framework takes in high level communicative function descriptions, generates behavior descriptions and realizes them using our character animation engine. We define the specifications for these different levels of descriptions. Finally, we show the general results as well as experiments illustrating the impacts of the key features.

1 Introduction

Gestures play an important role in everyday communications. People use gestures to indicate location, describe an imaginary object, express attitudes, or regulate conversation flow [8]. Effective speakers use gestures as a tool to better convey ideas. For example, a clinician's use of gestures can impact the clinician-patient relation [6].

Our desire is to generate gestural performances for life-like virtual characters. At the level of individual gestures, gestures have a complex feature structure. There are the phases of gestural motion including the rest, preparation, stroke, holding and relax phases [8], as well as the form of the motion, its location and changing handshapes. However, our interest lies beyond realizing these features in individual gestures. We model an approach to integrating individual gestures' features into an overall fluid performance involving gesture sequences (a.k.a. gesture units [7,9]).

Specifically, the goal of this work is to realize gesturing for virtual characters that takes into account that human gesturing has a hierarchical structure that serves important demarcative, referential and expressive purposes [3]. Within a gesture performance, some features such as handshape, movement or location in space, may be coupled across gestures while other features serve at times a key role in distinguishing individual gestures, both physically and at the level of its

T. Bickmore et al. (Eds.): IVA 2014, LNAI 8637, pp. 477–491, 2014.

meaning, from one another. For example, the hands may go into a rest position between gestures to indicate the end of an idea, a change of handshape can serve to indicate the start of a new idea in the discourse or one gesture's location may serve to refer to a preceding gesture.

We layout an approach that uses this higher level of organization to realize gesture performances. This approach *a*) determines when and which features are common versus which ones must be distinguishable and *b*) addresses issues concerning the physical coordination or co-articulation between gestures within gesture units, including when gestures go into relax, rest positions or holds. Closely leveraging the work of Calbris [3], we use the concept of ideational unit, which Calbris argues structures the discourse and the kinesic segmentation of gestures. Specifically, ideational units serve to impose requirements on gestural features in an overall performance. Further, we use Calbris' work on the relation of form and meaning in gesture [3] to help inform how a gesture's expressivity is manipulated.

In this paper, we propose a model of this kinesic segmentation. We then describe an implementation that satisfies these requirements by selecting and flexibly modifying the performances of gestures. Finally we present results and suggest areas for further improvements.

2 Background Theory

To ground our model, we leverage research from Calbris [3], a French semiologist, that views shape and movement of communicative gestures as abstraction of physical objects and actions (e.g. putting the hands in the shape of a bowl as a sign of offering; using the back-front line to indicate temporal information). She also studied how gestures are organized one from other.

First, Calbris argues for a structuring of verbal and nonverbal behaviors into larger ideational units, a coupling of related ideas that can span multiple gestures for example. This coupling plays important demarcative functions as well as helping to convey meaning and is therefore critical to gesture specification and realization.

Specifically, consistency of aspects of the form or motion across gestures serve a demarcative function of illustrating a common overriding topic while changes in gestural form and movement can serve a demarcative function of indicating a topic shift. In other words, changes convey information. So going from one gesture to the next in one overriding conversational topic, a gesturer tends to not make changes in motions or handshapes that are not meaningful. To do otherwise risks undesired false implicatures, misinterpretations by an observer. Related to that, one can view a gesture's motion as a form of optimization, avoiding unnecessary exertion.

Additionally, the ideational units can be divided up into tighter rhythmic-semantic units by the type of change from one gesture to the next, change that provides information and performs a referential function. In other words, consecutive motions that have similar features or referential relations construct a rhythmic-semantic unit. For example, in enumeration there are similarities in the motion and form, contrast requires a certain symmetry and elaboration suggests repetition.

Second, Calbris argues that human conversation both verbally and non-verbally, is often grounded in metaphors, especially physical metaphors. So abstract concepts like an idea, an agreement or a relationship can be represented through gesture as physical objects or actions. Further, properties of, and actions on that abstract concept such as the importance of an idea, discarding an agreement or ending of a relationship can be indicated by gesture that suggests size, discarding or cutting of a physical object, respectively. Such metaphoric gestures are common in human speech and in some cultures a quite common form of gesture.

The physical metaphor that underlies a gesture imposes critical constraints on how the gesture is physically realized. So to convey the ending of a relationship, a speaker may gesture with a cutting motion that involves a vertical acceleration of the hand in a flat configuration with the edge of the hand leading the motion, as if it were a knife. Each of these features is critical, with the acceleration representing the knife's chopping motion, the flat palm representing the knife and the edge of the palm representing the knife's edge. So as a consequence, any animation system that realizes this behavior must obey these constraints that arise from the embodiment of virtual agent. In particular, attempts to fit the timing of the gesture to co-occurring speech or gestures that precede or follow must not override these constraints.

Similarly, we argue that the underlying metaphor can inform how we want to realize and manipulate the gesture's expressivity. For example, if the importance of an idea is conveyed by a two-handed gesture that frames an imaginary object, then a very important idea can be conveyed by framing a larger object. Similarly discarding an idea with prejudice might be indicated by a particularly strongly accelerated sweeping motion.

Finally the grounding in physical concepts and space means location matters across gestures. The delineation of a concept as an object in physical space means subsequent references to that object must refer back to that location.

3 Gesture Model Description

Our model for gesture works as a system of constraints, containing three categories based on hierarchical structuring of gesture performances. The first category, the one most central to this paper, is the constraints on gestural features within and across coupled gesture sequences, established by structuring into ideational units. The second category deals with the connection between individual gestures. The last category focuses on the constraints on individual gestures, such as timing of the stroke that are common to gesture work [17].

We first discuss the overall structuring of the model then we go into greater details into the constraints that operate on these structures.

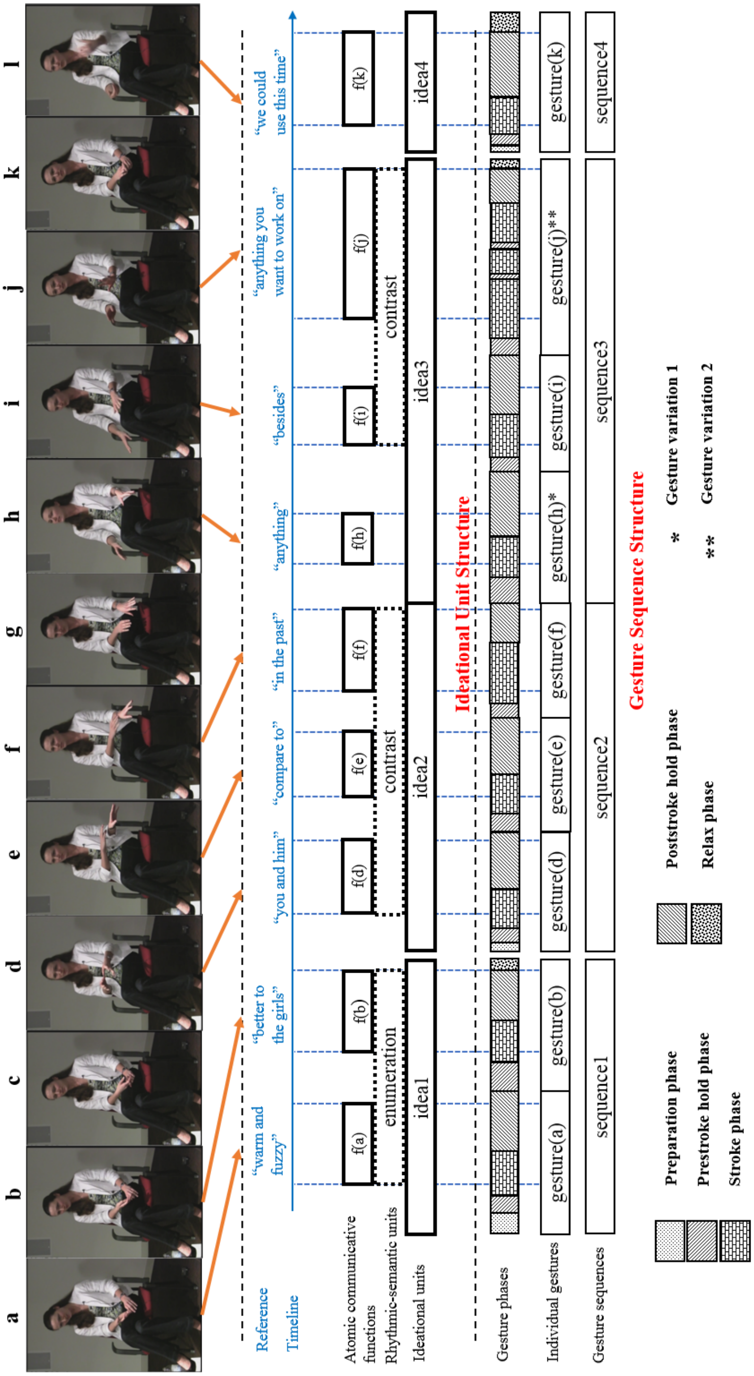


Fig. 1. Compound gesture model example

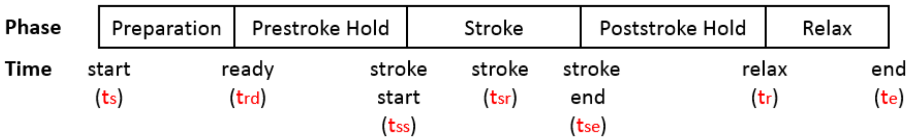


Fig. 2. *Gesture Phases*: $Preparation[t_s,\ t_{rd}]$ defines the period when gesture leaves body posture and arrives to gesture space. $PrestrokeHold[t_{rd},\ t_{ss}]$ is a pause before real stroke motion happens. $Stroke[t_{ss},\ t_{se}]$ is where gesture motion conveys its meaning, it's the most essential part of a gesture. $PoststrokeHold[t_{se},\ t_r]$ is a hold period after stroke phase is done. $Relax[t_r,\ t_e]$ is when gesture goes back to body posture from gesture space.

We use an annotated turn within a dialog, shown in Figure 1, to illustrate some of these constraints. The dialog is a part of a role play between a clinician and a person pretending to be a client suffering from depression. The clinician is saying she is willing to talk about anything except what the client's husband is complaining about. First section of the figure (above first horizontal dash separator) lists a sequence of images in correspondence to the phrases the clinician used. The images without an arrow connecting the phrases indicate the rest positions. Second section (between first and second horizontal dash separator) demonstrates the ideational unit structure which represents the utterance in ideation space. Third section (below second horizontal dash separator) presents the example's gesture sequence structure in gesture space. The lines between second section and third section indicates the one to one mapping between ideation space and gesture space.

In what follows, we use standard notions concerning the time markers and phases of gestures [8], as described in Figure 2.

3.1 Hierarchical Unit Structure

We refer to Calbris' first argument and define structures accordingly. As seen in Figure 1, the hierarchical structure of the gesture performance is composed of two major types of units.

- **Ideational Unit/Gesture Sequence.** As we see in Figure 1, an ideational unit can be a compound or coupling of atomic ideas (communicative functions) - the willingness to talk about anything except what the husband wants constitutes one ideational unit, it can also contain what Calbris [3] referred to as rhythmic-semantic units. For example, contrast or enumeration tightly relate atomic communicative functions and their associated gestures to each other. Ideational unit is in correspondence to gesture sequence (gesture unit), which is formed by one or more individual gestures.
- **Atomic Communicative Function/Individual Gesture.** This level is comprised of communicative functions that are associated with corresponding individual gestures [8], such as "anything", $f(h)$ in Figure 1.

3.2 System of Constraints

We consider three types of constraints depending on if they act within or across ideational units. While most of the constraints come from section 2 and literature review about gestures [8,7,16,17,2], some of the constraints come from our own observations of human data.

Within and Across Ideation Units/Gesture Sequences This category of constraints are based on the arguments from section 2. It defines the constraints for gesture performance according to the structure of the ideational units.

– *To serve the purpose of demarcating ideas, hands tend to rest or relax at the end of ideational units (see first section in Figure 1), but do not go to rest between gestures within ideational units.* Whether hands rest between gestures within ideational units depends on whether there is sufficient timing between gestures (for example, notice in Figure 1 between $gesture(f)$ and $gesture(h)$, there's no time left to go back to rest position.)
– *Consecutive gestures should have distinctive features within ideational unit.* In effect, they should be visibly distinguishable one from the other with notable provisos: features that are not relevant to what current gesture is conveying, persist from previous gestures (for example, notice in Figure 1, handshape persists).
– *References between gestures can be through similarities in (i.e., constraints on) the shape, movement or location features.* The reference can be local, e.g. in the case of enumeration, the reference is within ideational unit though potentially it can also be across ideational units. In the case of referencing impacting movement: the movement in the subsequent (referring) case may likely be more abstract (loose repetition of the movement). A gesture that refers to other gestures by physical location should have their stroke-end be at the location of the referred gesture and less critically the referring gesture shape should be the shape of the referred stroke-end gesture. See Figure 1, $gesture(h)$ is a metaphoric gesture that is referred to by $gesture(j)$.

Within An Ideational Unit This category describes the co-articulation constraints between gestures, to be more specific, gesture phase sequence construction between two individual gestures. An example can be found in third section of Figure 1.

– *Co-articulation between gestures eliminates some phases* [8]. Given two gestures, $gesture1$ and $gesture2$, usual co-articulation has the following phase sequence:
$Stroke(gesture1)$, $PoststrokeHold(gesture1)$,
$Preparation(gesture2)$, $[PrestrokeHold(gesture2)]$, $Stroke(gesture2)$
Note that the phase specified inside the square bracket usually can be skipped depending on the timing constraints (i.e. if there is not enough time). The gesture can go to a relax phase instead of a hold.
– *The co-articulation from previous gesture to a beat gesture should ensure that the beat gesture should have a preparation phase that allows it to be distinguishable from the previous one.* The dynamic property of the preparation

phase of the beat should come mainly from the beat gesture. This is important when 2 gestures phases (relax phase of previous gesture and preparation phase of following gesture) need to blend with each other; or to co-articulate with one another. Beat gestures often have a quick and accelerated preparation phase: an anticipation of the dynamic quality of the stroke phase of the beat gesture. This constraint comes from our own data analysis.
- *When the same gesture is repeated to emphasize a point, it tends to be repeated with a different expressivity quality* [2]: the motion between t_{ss} and t_{se} of the repeated gesture have a larger spatial extent; the movement between the t_{ss} and t_{se} is larger (e.g. the hands draw a larger circle); and the velocity of the movement increases.
- *Gesture co-articulation fundamentally has to obey the dynamic constraints of the motion (again based on our own observations).* Otherwise it will end up in unnatural results that draws the observer's attention and interpretation away from the stroke of the gesture itself. For example, the transition motion speed between gestures can't be very different from the gesture speed before the transition and after, which might cause unexpected acceleration or de-acceleration.

Individual Gestures This category of constraints deals with individual gestures, most of them are common to the gesture literature [8,17], but we also incorporate the concept of embodiment into physical gestures [3].

- *Gesture with a referential content can be adjusted on the fly according to content's physical properties* [3]. For example, representing an abstract concept as a physical object might have its size property, either big or small, be adjusted to reflect the importance of the idea. Gestures that have such variations can be found in Figure 1 marked with $*$.
- *Relax position is different from rest position* [16]. At relax position, the hands still stay in the gesture space. The hands may go toward the body a bit and the hand shape takes a more open relax shape.
- *Although hold are often described as being necessary for co-articulation, they can also be used for emphasis* [8]. Their use is highly dependent on timing constraints between consecutive gesture strokes.
- *Strokes have to occur on or slightly before stress words* [17].

4 Framework

Our gesture generation system follows the SAIBA ([11]) framework guidelines. The whole process is comprised of two main modules, behavior planning and behavior generation. Behavior planning module takes in high level communicative function descriptions as input using function markup language (FML) and generates behavior descriptions using behavior markup language (BML). Behavior realization module takes in BML descriptions outputs a description of the character's animation in terms of joint rotations.

4.1 Behavior Planning

The behavior planning module contains communicative function derivation and behavior mapping processes that maps ideational units (described by FML) to gesture behaviors (described by BML).

```
<fml>
....
<ideational id="i3">
  <function type="container" id="c3" location="center" action="delineate-set"
        modifier="round" attribute="large" reference="anything"/>
  <rhythmic-semantic id="d3" type="elaboration" nucleus="c3"/>
    <function type="movement" action="discard" source="c3" target="right"
        reference="what he wants"/>
    <function type="reference" action="indicate" location="c3" reference="anything"/>
    <function type="reference" id="r1" action="indicate" location="front" reference="you"/>
    <function type="emphasis" reference="you want to work on"/>
  </ rhythmic-semantic >
</ideational>
...
</fml>
<bml>
...
<gesture-sequence rest="sp1:T117" constraint-rest="true" constraint-handiness="true"
        constraint-handshape="true">
  <gesture move-lexeme="sweep-dome" hand-lexeme="flat" palm-orient="down"
        extent="large" location="center" stroke="sp1:T92"/>
  <gesture move-lexeme= "push" hand-lexeme="flat" palm-orient-right="down"
        palm-orient-left= "right" extent="large" location= "right" stroke="sp1:T98"/>
  <gesture move-lexeme="sweep-dome" hand-lexeme="flat" palm-orient="down"
        extent="large" location="center" stroke="sp1:T104"/>
  <gesture move-lexeme= "forward" hand-lexeme="flat" palm-orient="oblique-forward"
        location="front" stroke="sp1:T108">
    <gesture-overlay move-lexeme="forward-down" hand-lexeme="flat" palm-orient="down"
location="front" stroke="sp1:T110"/>
  <gesture/>
</gesture-sequence>
...
</bml>
```

Fig. 3. Example FML and BML snippet

Our FML descriptions use extended FML specifications to reflect our compound model and particularly to reflect the ideational unit structure. We add three type of tags: <ideational>, <rhythmic-semantic> and <function>. They are in correspondence to ideational unit, rhythmic-semantic unit and atomic communicative function respectively. <function>'s attributes can be found in Table 1. Note that <function> not only defines communicative function, but also specifies actions which usually only appear in BML domain. The reason is the actions in FML are not the real physical actions, but rather a description of the communicative function in terms of an underlying embodied metaphor.

Our BML descriptions extend standard BML specifications and add <gesture-sequence>, <gesture> and <gesture-overlay> to reflect the gesture sequence structure. <gesture-sequence> is the behavior space mapping of an ideational unit, its attributes for this tag represent the features within and across ideational

Table 1. FML <function>'s attributes

Specification	Description	Example
id	identity name for the function	c1, c2
type	type of communication function	reference, emphasis, movement, container
location	location of metaphoric object	center, left, right, up, down, front, back
action	action	listing, delineate, move, discard
modifier	modifier for the action	finger, round
attribute	attribute for metaphoric object	large, further
reference	reference to utterance	warm and fuzzy

unit. Similarly, <gesture> is the behavior space mapping of the atomic utterance content, its attributes reflect features within the same communicative function unit. <gesture-overlay> describes repeated gestures happen inside an unitary gesture which is a result of derivative process from features needed for gesture performance. An example snippet of FML and BML description can be found in Figure 3.

We use a rule based system approach (as in [15,13,4,14]) as our behavior planner and extend its rules to support features in the model. The key role for these rules is to map from FML descriptions to BML descriptions. As part of this mapping, the rules need to resolve the referential content's location accordingly to their ids (see Table 1). Also they must map FML aspects like emphasis to behavioral manifestations like repetitions, such as in the following rule example.[1]

```
emphasis_function
if
   fcn($functionType,$modifier,$start1,$end1,$priority1)
   fcn(emphasis,$modifier1,$start1,$end2,$priority2)
   check $end1 < $end2
then
   fcn($functionType,overlay,$end1,$end2,$priority1)
```

4.2 Behavior Realization

The behavior realization module generates the animation. The animation platform we use, SmartBody [21], supports both procedural and data driven techniques, but in the case of gesture generation, we rely on key-frame or mocap animations to get natural and smooth results. Based on the animation framework, we developed an algorithm that animates the BML output descriptions provided by the previous behavior planning module.

Behavior realization has two steps. <gesture-sequence> defines the selection of individual gesture animations with given all the constraints as requirements

[1] Variables in the example rule refer to Table 1.

(these animations will combine to be the final animation sequence). Then the blending techniques are used to adjust individual gesture animation to achieve certain constraints or variations, such as to modify gesture so it can refer to location of previous gesture or depict a metaphoric object's physical size.

We assume a motion database with each gesture motion tagged with labels including communicative function, type of action, handedness, handshape, action modifier such as big and small. We run Algorithm 1 to realize step one as mentioned above.

We use $M = \{m\}$ to define our gesture motion database and each motion has a set of tags $T_m(m) = \{tag\}$, given input BML behavior set $B = \{b\}$ with each behavior with tag $T_b(b) = \{tag\}$, we are trying get a final animation sequence $A_{final} = \{...\}$. $ConstraintFunc$ applies the constraints defined in the section 3, for example, consecutive gestures should have the same handshapes. $ChooseBest$ looks at possible animation sequences $VectorA$ and tries to find the best one. Here, we simplify $ChooseBest$ by just hand picking the best one[2].

```
motion subset for behavior b M'[b] = {φ}
for i ← 1 to behaviorSetSize do
   for j ← 1 to motionSetSize do
      if T_m(j) is a subset of T_b(i) then
         Append(M'[b], M(j))
      end if
   end for
end for
VectorA = []
for i ← 1 to behaviorSetSize do
   temporary animation sequence A = {φ}
   motionSubSetSize = size(M'(i))
   for j ← 1 to motionSubSetSize do
      if A is empty then
         A ← M'(i)(j)
      end if
      a =last element of A
      meetConstraints = ConstraintFunc(M'(i, j), a)
      if meetConstraints is TRUE then
         Append(A, M'(i, j))
      else
         Prune M'(i, j) or a by priority and hands go back to rest position
      end if
   end for
   Append(VectorA, A)
end for
A_final = ChooseBest(VectorA)
```

Algorithm 1. Animation Sequence Generation Algorithm

[2] We visually choose the one that is the smoothest and the most natural in the overall performance.

Then we utilize the SmartBody blending controller, based on a barycentric parametric blending technique [5], so that it can flexibly adjust individual gesture animations. Similar gesture animations are grouped offline to define an animation blend space and adjustment parameters are extracted (automatically by calling SmartBody's API). During run-time execution, control parameters are given as an input from the BML descriptions and used to calculate the weights for each motion inside the blend. Take *sweep-dome* gesture for example, there can be two similar motions, one depicts a small object with modifier label as "small" and one depicts a large object with modifier label as "large". If BML descriptions input a sweep-dome gesture with a modifier "medium", we interpret that in parameter space, infer the distance needed for two hands and use it to compute the weights for motions. Similarly we can do that for adjusting locations of the hand.

5 Related Work

Many techniques have been explored to generate gestures, differing in terms of input data used, underlying model and framework. Many researchers have focused on gesture generation. ACE [10] focuses on deictic and iconic gestures, it takes text input and looks for specific words in order to display associated gestures with timing based on prosody analysis. NVBG [13] is a rule based system that uses the communicative intent embedded in the surface text as well as the agent's cognitive processing such as internal goals and emotion states. Cerebella [15,14] used an improved communication function inference mechanism along with prosodic analysis. Kopp et al. [12] based their system on the *Sketch Model* [20] and can create gestures from arbitrary form specifications and handle co-articulations. Kipp et al. [9] introduced a system that generates gestures in particular styles based on probabilistic reproduction of data captured from a human subject, and an extension that includes dynamics [18].

Among all these works our method resembles [9] the most in terms that we all deal with a sequence of gesture movements that go beyond the structure of individual gestures. The difference is they capture the regularities indirectly through data analysis. In our approach, we are explicitly modeling the constraints and features that are carried on within and in between ideational units.

6 Results

To illustrate the approach, we annotated five videos drawn from a simulated role play corpus, an example of which is depicted in Figure 1. FML descriptions were created for these videos (see Figure 3) to provide input for our framework which then generated gesture performances. In order to examine the impact of *within* and *across ideational units* constraints, we removed individual constraints one by one to check its impact. We created side by side comparison videos, one with all constraints active (on), the other missing one constraint, as can be seen from link

(a) Frame 1 **(b)** Frame 2 **(c)** Frame 3 **(d)** Frame 4

Fig. 4. A sample performance from an animation video: the above sequence shows the key strokes for the performance generated for: "Okay, let's just backup for a second then, what's happening right now? So when we were talking, you find yourself kind of drifting off, what's going on?"

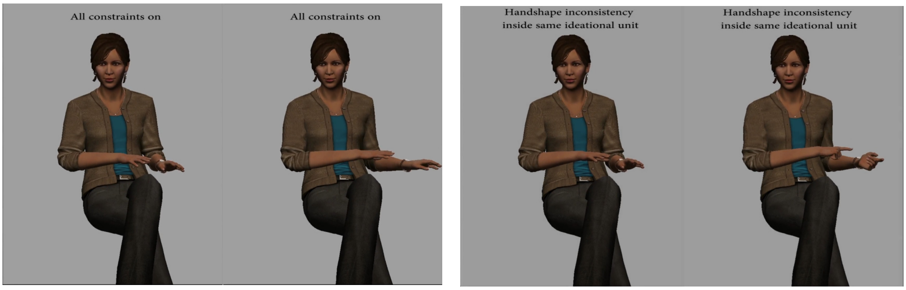
`http://youtu.be/A-3Ic-zCqnM`.[3] Alternatively, you can also find an example performance depicted Figure 4 and a comparison example from Figure 5.

Studies. We also did studies to specifically test the impact of the "hands going to rest or relax position constraint at the end of ideational units but not going to rest position within ideational units" constraint. We consider two conditions here, one is hands never go to rest or relax position and the other is hands always go to rest position after finishing individual gestures. Leveraging Amazon Mechanical Turk [1], we ask the participants to select which video they think is closer to how humans gesture and give their strength of preference (scaled from 1 to 5 and 5 is strongest) after watching the comparison video. We randomize both the order between video pairs and the overall order of videos being watched by participants. Each comparison video is assigned to 50 workers.

First study is designed to test the first condition, we only use three examples since two of them only contain one ideational unit, which won't be able to show the constraint impact across the ideational unit. Video with all constraints active is preferred with a percentage of 71.1%, strength of preference 3.79, while video with hands never go to rest or relax position is preferred with a percentage of 29.9%, strength of preference 3.11. Binomial test is run on preferences with significant value $p < 0.001$. The result of the study shows strong impact of the constraints.

Second study covers the second condition, all five examples are used. Video with all constraints on and video with hands always go to rest position has a preference rate of 40.6% and 59.4% respectively, although people who pick the

[3] The video is organized as follows: it first provides five overall results, then four comparison videos, followed by videos presenting adjustment of metaphoric size of an object and gesture referential location using parametric blending technique, finally we present a video showing the repeated gesture inside individual gesture used for emphasizing a point.



(a) All constraints on **(b)** Inconsistent handshapes

Fig. 5. A comparison example: each image shows two frames of the performance. Left image has all constraints active and right image disables one constraint - the handshape consistency within the same ideational unit.

video with all constraints on has a strength of preference value of 3.46 which is higher than 3.25 when they pick the video with hands always go to rest position. Binomial test is run on preferences with significant value $p < 0.001$. The gestures in the video are located closed to the rest position. So visually, there are not much differences between both videos. The rule needs to be further verified on other examples.

We haven't done study yet for other constraints like handshapes and handiness consistency. However, based on our visual inspections, it was fairly obvious to the authors that violating these constraints caused visual awkwardness.

7 Discussion

In this paper we present a novel sequential gesture model that looks at ideational structures to provide guidelines for the gesture performances.

The main contribution in turns of implementation includes creating behavior planning rules to map from FML descriptions to BML descriptions and a behavior realization algorithm. Results and studies show that these constraints play an important role in natural human gesturing that without it the performance would not look right.

The second study also identified a key issue with how we implemented the model. Although our model posits the distinctions between relax and rest pose within the ideational unit, we didn't realize it in our implementation, that might be the reason of the failure for the second study. To be more specific, our implementation uses a hold which is often utilized to emphasize a point, instead of relaxing after t_{se} of a gesture during the transition to the other gesture within the same ideational unit, this causing an unnatural result.

For the future work, we are hoping to test our sequential gesture model on different behavior realizers such as Greta [19].

Acknowledgments. We would like to thank Teresa Dey for all the animations provided for our motion database.

References

1. Amazon: Mechanical turk, `https://www.mturk.com/mturk/`
2. Calbris, G.: The Semiotics of French Gestures, vol. V. Indiana University Press (1990)
3. Calbris, G.: Elements of Meaning in Gesture. Gesture studies. John Benjamins Publishing Company (2011)
4. Cassell, J., Vilhjálmsson, H.H., Bickmore, T.: Beat: The behavior expression animation toolkit. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001, pp. 477–486. ACM, New York (2001)
5. Feng, A.W., Xu, Y., Shapiro, A.: An example-based motion synthesis technique for locomotion and object manipulation. In: ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games, Costa Mesa, CA (March 2012)
6. Hall, J.A., Harrigan, J.A., Rosenthal, R.: Nonverbal behavior in clinician patient interaction. Applied and Preventive Psychology 4(1), 21–37 (1995)
7. Kendon, A.: Gesture: Visible Action as Utterance. Cambridge University Press (2004)
8. Kendon, A.: Some relationships between body motion and speech. Pergamon Press (1972)
9. Kipp, M., Neff, M., Kipp, K., Albrecht, I.: Towards natural gesture synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 15–28. Springer, Heidelberg (2007)
10. Kopp, S., Jung, B., Leßmann, N., Wachsmuth, I.: Max - a multimodal assistant in virtual reality construction. Künstliche Intelligenz 17(4), 11 (2003)
11. Kopp, S., Krenn, B., Marsella, S.C., Marshall, A.N., Pelachaud, C., Pirker, H., Thórisson, K.R., Vilhjálmsson, H.H.: Towards a Common Framework for Multimodal Generation: The Behavior Markup Language. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 205–217. Springer, Heidelberg (2006)
12. Kopp, S., Wachsmuth, I.: Synthesizing multimodal utterances for conversational agents. Computer Animation and Virtual Worlds 15(1), 39–52 (2004)
13. Lee, J., Marsella, S.C.: Nonverbal behavior generator for embodied conversational agents. In: Gratch, J., Young, M., Aylett, R.S., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, pp. 243–255. Springer, Heidelberg (2006)
14. Lhommet, M., Marsella, S.C.: Gesture with meaning. In: Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (eds.) IVA 2013. LNCS, vol. 8108, pp. 303–312. Springer, Heidelberg (2013)
15. Marsella, S., Xu, Y., Lhommet, M., Feng, A.W., Scherer, S., Shapiro, A.: Virtual character performance from speech. In: Symposium on Computer Animation, pp. 25–35 (2013)
16. McNeill, D.: Hand and Mind: What Gestures Reveal about Thought. University of Chicago Press (1992)
17. McNeill, D.: Gesture and Thought. University of Chicago Press (2005)

18. Neff, M., Kipp, M., Albrecht, I., Seidel, H.P.: Gesture modeling and animation based on a probabilistic re-creation of speaker style. ACM Transactions on Graphics (2008)
19. Niewiadomski, R., Bevacqua, E., Mancini, M., Pelachaud, C.: Greta: an interactive expressive ECA system. In: International Conference on Autonomous Agents and Multiagent Systems, pp. 1399–1400 (2009)
20. de Ruiter, J.P.: The production of gesture and speech. Cambridge University Press (2000)
21. Thiebaux, M., Marsella, S., Marshall, A.N., Kallmann, M.: Smartbody: Behavior realization for embodied conversational agents. In: Proceedings of the 7th International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS 2008, vol. 1, pp. 151–158. International Foundation for Autonomous Agents and Multiagent Systems, Richland (2008)