# Sleight of Hand: Perception of Finger Motion from Reduced Marker Sets

Ludovic Hoyet* Kenneth Ryall† Rachel McDonnell‡ Carol O'Sullivan§

Graphics, Vision and Visualisation Group, Trinity College Dublin

**Figure 1:** *Using inverse kinematics with 8 markers per hand enables us to simultaneously capture full-body and hand motion while retaining most of the fingers' perceived information. This figure depicts side-by-side comparisons of real movies and computer generated animations.*

## Abstract

Subtle animation details such as finger or facial movements help to bring virtual characters to life and increase their appeal. However, it is not always possible to capture finger animations simultaneously with full-body motion, due to limitations of the setup or tight production schedules. Therefore, hand motions are often either omitted, manually created by animators, or captured during a separate session and spliced with full body animation. In this paper, we investigate the perceived fidelity of hand animations where all the degrees of freedom of the hands are computed from reduced marker sets. In a set of perceptual experiments, we found that finger motions reconstructed with inverse kinematics from a reduced marker set of eight markers per hand are perceived to be very similar to the corresponding motions reconstructed using a full set of twenty markers. We demonstrate how using this reduced set of eight large markers enabled us to capture the finger and full-body motions of two actors performing a range of relatively unconstrained actions using a 13-camera motion capture system. This serves to simplify the capture process and to significantly reduce the time for clean-up, while preserving the natural biological movements of the hands relative to the actions performed.

**CR Categories:** I.3.7 [Computer Graphics]: Three Dimensional Graphics and Realism—Animation;

**Keywords:** finger animation, perception, motion capture

*e-mail:hoyetl@tcd.ie
†e-mail:ryallk@tcd.ie
‡e-mail:ramcdonn@cs.tcd.ie
§e-mail:carol.osullivan@cs.tcd.ie


I3D 2012, Costa Mesa, CA, March 9 – 11, 2012.
© 2012 ACM 978-1-4503-1194-6/12/0003 $10.00

## 1 Introduction

Gestures and finger motions play an important role in our everyday actions: we use our hands to interact with the environment as well as to convey and emphasize information in conversations. Adding such details to computer generated animations can therefore help to increase the naturalness, credibility and appeal of virtual characters. However, because of the complexity of the hand's anatomy and the consequential difficulties in capturing all its subtle movements, finger motions are often omitted from animations or manually created by animators. Two main techniques are typically used to capture finger motions: datagloves or optical systems. In both cases, wires or small optical markers limit the capture area and constrain the actions that can be captured. Therefore, hands and bodies are often captured during separate sessions and later spliced together.

In the case of optical systems, which can usually capture large areas, introducing hand capture requires moving the cameras closer to the actor in order to increase the resolution of the projected image of small markers ($< 6mm$), thereby drastically reducing the capture space. Furthermore, whereas a set of approximately 50 larger ($\approx 14mm$) markers is usually enough to capture the full body motion of a character reasonably accurately, the capture of finger motions introduces about 20 additional markers per hand. Because of the hand's complexity and the small size of the markers used (e.g., $6mm$ markers project a surface less than five times the size of $14mm$ ones), numerous occlusions and labeling errors occur during manual post-processing, thereby significantly increasing the workload on animators. It is because of these constraints that hand animations are often ignored or greatly simplified during busy production schedules for games or movies.

Reducing the number of markers required to capture plausible hand movements has several potential benefits. It increases the distance between markers, which allows the use of larger markers to simultaneously capture hands and bodies when the number and resolution of cameras are constraints. Furthermore, it greatly reduces the number of occlusions and labeling errors, thereby reducing manual post-processing time by almost an order of magnitude. Finally, and perhaps most importantly, humans rely on synchronization cues when perceiving communication and gestures [McNeill 2005; Giorgolo and Verstraten 2008]. Breaking this synchrony, either between auditory and visual information or between different visual channels (hands, arms, body or facial movements), can reduce the credibility

of virtual characters. Therefore, capturing multiple channels simultaneously ensures that the synchrony of the actions is preserved.

In this paper, we present a set of perceptual experiments to investigate the perceived fidelity of hand motions. We examine a variety of different action types, ranging from pointing and grasping to conversational gestures. Our aim is to find the optimal trade-off between perceived fidelity and the number of markers needed to capture finger motion. We show that finger motions reconstructed with inverse kinematics (IK) from a reduced marker set of eight markers per hand are perceived to be very similar to the corresponding motions reconstructed using a full set of twenty markers. Based on our results, we demonstrate that this set of eight $14mm$ markers per hand can be used to simultaneously capture the full-body and finger motion of two actors performing a range of relatively unconstrained actions across our full capture space (Figure 1). Our results will be of interest to those who have a constrained motion capture setup, or in the games industry, where reducing the time spent on post-processing by animators is a priority.

## 2 Related Work

Understanding and modeling the human hand has been addressed by many researchers from both computer graphics and biomechanics. Previous research has focused on designing and animating high quality anatomical hand models [Albrecht et al. 2003; Pollard and Zordan 2005; Tsang et al. 2005; Liu 2009; Sueda et al. 2008], which sometimes requires focusing in particular on some joints, such as the complex thumb Carpometacarpal Joint [Chang and Pollard 2008].

Because of the setup necessary to simultaneously capture fingers and full body motion, it is interesting to explore how to decrease the number of markers used to track finger motions with optical systems. Chang et al. [2007] used supervised feature selection to determine the minimal set of surface markers necessary for grasp classification of individual hand poses. Similarly, Jörg and O'Sullivan [2009] explored the correlations between the degrees of freedom of the hand and advised to focus on the capture of the thumb, the index and at least one additional finger. However, these correlations were based on the degrees of freedom (DoF) of the fingers, and not on the number of markers required to capture finger animation. Other researchers have investigated image-based finger motion capture, such as Wang and Popović [2009] who use a color glove to map the hand configuration to a database of hand poses. In the same vein, but for full-body motions, Chai and Hodgins [2005] reconstructed human motions from a reduced number of markers using a pre-recorded database.

Instead of simultaneously capturing hand and full-body motion, another solution would be to splice separately captured hand motions with full body animation. This tedious process commonly used by games and movie companies is usually done manually, but can be computed automatically in certain situations [Jin and Hahn 2005; Majkowska et al. 2006]. Furthermore, this technique requires skilled animators to ensure that the synchrony between the full body animation and the hand movements is maintained, since it has been proven that humans rely heavily on synchrony to communicate [McNeill 2005; Giorgolo and Verstraten 2008].

Despite the large amount of work on the perception of human motion in computer graphics [Hodgins et al. 1998; Reitsma et al. 2008; McDonnell et al. 2008; Ennis et al. 2010], as well as on the importance of correct gestures and hand motions for virtual humans [Cassell et al. 1994; Kipp et al. 2007], the perception of finger animation in particular has seldom been studied. To our knowledge, the only study was conducted by Jörg et al. [2010], who showed that synchronization errors between full-body and hand animations are easily detected, but the perception of these errors is highly dependent on the type of motion.

In computer animation, IK methods have been extensively used to drive the end effector position of articulated chains. Rijpkema and Girard [1991] described the IK control of fingers, coupled with high-level control of the hand for the grasp planning problem. El-Sawah et al. [2006] proposed to compute an offline error of the IK model, which is used at run time to solve the fingers' four DoF joint angles. To model the sympathetic motion between the fingers of the hand, ElKoura et al. [2003] presented an example-based finger IK solver used in combination with a procedural algorithm to control the fretting hand for a given piece of music.

In this paper, we are interested in evaluating the perceptual fidelity of finger animations computed using IK with a small number of markers. For these experiments, we simultaneously recorded high quality finger motions (numerous markers) and full-body motion. We then generated new finger animations using IK with different subsets of the high quality captured markers, and asked participants to compare these animations to animations generated with the full high quality marker set.

## 3 Methods and Setup

### 3.1 Motion Capture

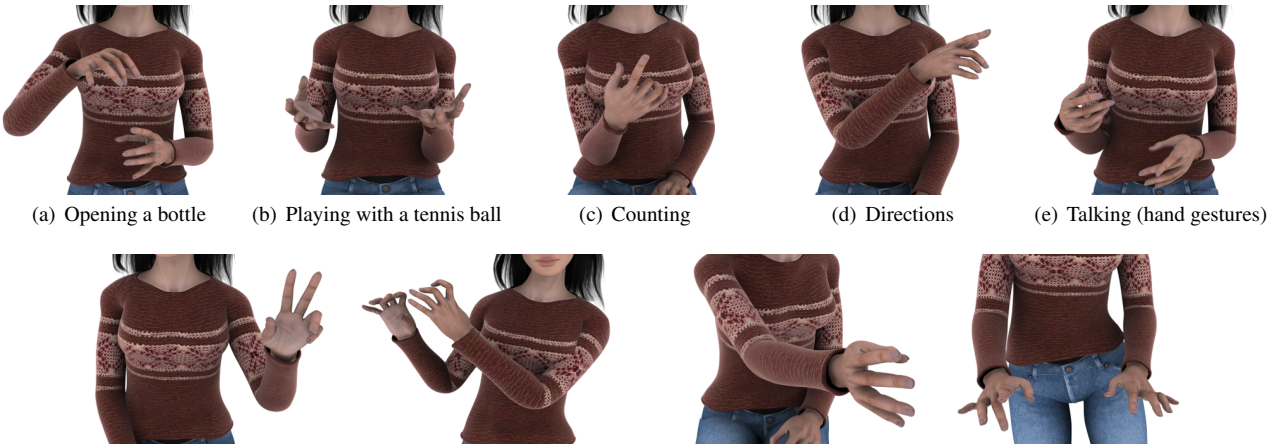
We simultaneously recorded the full body and finger movements of a non-professional female actor. Motion capture was conducted using a 13 camera Vicon optical system at 120Hz, where 51 markers were used to capture the full body motion of the actor, and 20 additional markers were used per hand to simultaneously capture the finger movements. Capturing the small markers positioned on the fingers ($4mm$) required us to carefully position our motion capture cameras around a small area ($\approx 1.5m \times 1.5m$).

To cover a large range of finger motions, we captured nine scenarios, with different levels of finger coordination and velocity: opening/closing a bottle, grasping, playing with a tennis ball, signing, talking, pointing, counting, typing on a keyboard, playing the flute (Figure 2). This set of motions included four of the six grasp types, selected from functional grasps for daily living [Edwards et al. 2002]. We also captured a range of motion sequence of both hands where each finger was moved independently.

The captured motion of the actor was then mapped onto a skeleton, where joint angles were computed and used to drive the virtual characters in Autodesk 3ds Max. Beforehand, the marker trajectories were filtered at 20Hz using a Butterworth filter to remove high frequency noise without losing the subtle details of the finger motion.

### 3.2 Hand animation

To animate the virtual character's hands, we used a similar skeleton to those previously described [Lee and Kunii 1995; Wu and Huang 2001; El-Sawah et al. 2006]. Most models use only 2 DoF (flexion and abduction) for the finger's Metacarpophalangeal (MCP) joint, which does not capture the small but subtle tilting due to the movements of the metacarpal bones. We found that using a 3 DoF joint for the finger's MCP joint of the high quality animations captured these small deformations and produced more natural looking poses when using a full set of markers. Therefore, each of the four fingers has 5 DoF, with 3 DoF for the MCP joints and 1 DoF for the flexion/extension of each of the two other joints. The thumb has 5 DoF, with 3 DoF for the Carpometacarpal (CMC) joint and 1 DoF for the flexion/extension of each of the two other joints (Figure 3). The metacarpal bones are however treated as a single rigid body.



(a) Opening a bottle (b) Playing with a tennis ball (c) Counting (d) Directions (e) Talking (hand gestures)

(f) Signing (g) Playing the flute (h) Grasping (i) Typing

**Figure 2:** *Scenarios captured for the experiments presented in this paper.*

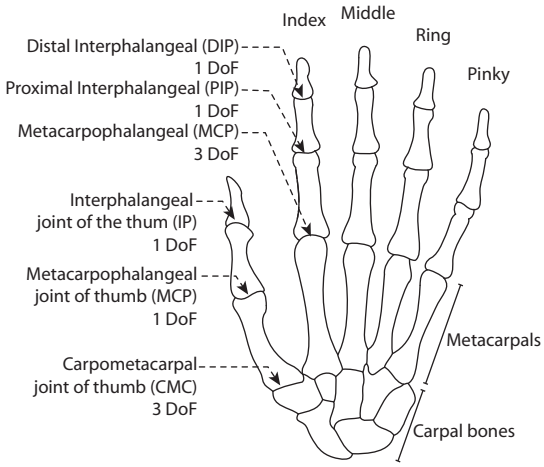



**Figure 3:** *Human hand skeleton and its main joints.*

To generate our gold standard finger motions, we used Forward Kinematics (FK) to drive each of the fingers and the thumb. Four markers were used to reconstruct each finger. The range of motion previously captured was used to compute the length of each segment in the skeleton of the actor. Each of the four fingers and thumb configurations was then reconstructed with FK using the trajectories of the corresponding four captured markers.

We also designed different models that mapped decreasing numbers of markers to the virtual hands. These simpler models are based on Inverse Kinematics (IK) with only two markers used to drive the animation of one given finger/thumb. Information about the length of each knuckle is crucial to generate natural looking poses with IK, as different knuckle lengths will generate different joint configurations. As this information is not present when using only two markers, we used biomechanical data from [Greiner 1991] to compute the average length of each knuckle depending on the length of the corresponding finger/thumb. We purposely recomputed the hand skeleton dimensions for each model in order to evaluate the performance of each reduced marker set on the whole animation process.

In the case of computing the configuration of a finger per frame, there are too many DoF to find a unique configuration using IK. Therefore, we need to reduce the number of DoF. In order to do so, we based our simplifications on information from previous literature (which leaves us with 3 DoF per finger) to allow the computation of an unique configuration:

1. In our IK model we use only two DoF (flexion and abduction) for the MCP joint [Lee and Kunii 1995; Wu and Huang 2001; El-Sawah et al. 2006], as using two markers would produce an infinite number of configurations.

2. A human finger has the property that it is almost impossible to move the DIP joint without moving the PIP joint, and vice versa. Therefore, we used a joint angle dependency discovered early in hand animation between the Proximal Interphalangeal (PIP) and the Distal Interphalangeal (DIP) joints to remove the DIP joint DoF [Rijpkema and Girard 1991]:

$$\theta_{DIP} = \frac{2}{3}\theta_{PIP} \quad (1)$$

Due to the greater kinematic complexity of the thumb, we found that computing its 3DoF configuration from two markers created abnormal configurations. We found that removing the Interphalangeal (IP) joint DoF helped to avoid these unnatural configurations. The following experiments will show that this simplification did not affect the perceived fidelity of animations.

## 4 Experiment 1

In this experiment, we wished to evaluate the perceived fidelity of finger animations created from a range of marker sets using traditional IK methods, compared to the corresponding high quality animations generated using FK with a full marker set. Our aim is to find an optimal trade-off between perceived fidelity and the number of markers used to capture finger motion.

### 4.1 Finger Animation Models

To differentiate between these different marker sets, we use a $x$F-$y$T notation, which stands for $x$ markers used for the animation of the four Fingers and $y$ markers used for the animation of the Thumb. The unused markers were discarded from the model.

**Gold Standard (GS) - 16F-4T:** This highest quality model (Gold Standard) will be compared against all other sets. The full marker set that was captured is used here: 4 markers on the thumb and on

| Effect | F-Test | Post-hoc |
|---|---|---|
| method | $F_{2,28} = 62.457, p < 0.00001$ | 8F-4T perceived as more similar to GS than all others |
| scenario | $F_{8,112} = 21.235, p < 0.0001$ | Some scenarios perceived more similar to GS, on average |
| method× scenario | $F_{16,224} = 13.224, p < 0.0001$ | 8F-4T: bottle scenario perceived significantly different to all others |
| | | 4F-2T: motions without independence between fingers (ball, talking, grasp and sign) rated as more similar to the GS than others |
| | | Static: motions with little finger movement (grasp, sign and direction) rated as more similar to the GS than others |

| Effect | F-Test | Post-hoc |
|---|---|---|
| method | $F_{2,28} = 50.674, p < 0.00001$ | 4F-4T was perceived significantly less similar to GS than the two others |
| motion | $F_{4,56} = 3.3071, p < 0.05$ | Flute motion perceived significantly less similar to GS than everything else |
| motion× method | $F_{8,112} = 4.1654, p < 0.0005$ | Motions using 6F-4T and 6F-2T never rated significantly different |
| | | Motions using 4F-4T always significantly different from other two (except ball) |

**Table 1:** *Significant results for the two-way ANOVA with between subject factors for experiment 1 (top) and experiment 2 (bottom). Factors:* method*,* scenario *and* motion*.* $A \times B$ *represents an interaction between factors* $A$ *and* $B$*.*

(a) (b) (c)

**Figure 4:** *The different sets of markers used for the first experiment: a) full marker set 16F-4T (20 markers) for the gold standard animations, b) 8F-4T (12 markers) and c) 4F-2T (6 markers).*

each of the four fingers (Figure 4.a). These markers were used to drive the joint configurations using FK (Section 3.2).

**8F-4T:** In this model, 2 markers were used per finger, positioned on the base and tip of each finger, and 4 markers for the thumb (12 markers in total, Figure 4.b). We reconstructed each of the four fingers using IK (Section 3.2), but computed the thumb with FK to evaluate separately the effect of the IK reconstruction of the four fingers on the perception of the animation.

**4F-2T:** In a lot of everyday motions, especially talking, hand movements do not always exhibit independent motions between fingers. In this model, we drive only two fingers from markers positioned on the base and tip of each, in order to evaluate the effect of dependencies between fingers. We expected this model to achieve good performance for motions where the fingers do not move independently of each other. Six markers in total are therefore used per hand to drive the thumb, Index (I) and Pinky (P) fingers using IK (Figure 4.c). We compute the joint configuration of the Middle (M) and Ring (R) fingers by evenly positioning them relative to the index and pinky fingers and linearly interpolating joint angles:

$$\forall k \in (MCP, PIP, DIP) \quad \begin{array}{l} \theta_{k,M} = \frac{1}{3} * (2 * \theta_{k,I} + \theta_{k,P}) \\ \theta_{k,R} = \frac{1}{3} * (\theta_{k,I} + 2 * \theta_{k,P}) \end{array} \quad (2)$$

**Static:** Game companies often use hand motions or poses selected from a database to animate the hands of characters. We wished to evaluate if under any circumstances a completely static hand, such as one manually posed by an animator, could be perceived to be the same as a high quality hand animation. For each captured motion, we manually selected a hand configuration, from the corresponding GS motion, that conveyed information about the desired action. This often corresponded to a peak in the action (e.g., stressing a comment, pointing at an object). For some actions which did not exhibit peaks, such as in the case of playing the flute or typing on a keyboard, we selected a plausible hand configuration from the GS motion. We chose not to randomly select a hand pose from the GS animation in order to ensure that the static pose would convey as much information as possible, as an animator would. We expect the absence of subtle hand movements to drastically lower the fidelity of these animations when compared to the GS.

### 4.2 Stimuli and Task

For this experiment, we chose a female virtual character which roughly matched the morphology of the actor. A viewpoint was chosen which displayed the character from waist to neck only in order to ensure that participants focused on the fingers. A white background was chosen to provide good contrast with the character. All movies were generated with high quality rendering (Figure 2) at 1280×800 and 30Hz. We generated 72 movies: 9 scenarios × 2 examples × 4 methods (GS, 8F-4T, 4F-2T and static). The 9 scenarios correspond to motion captured scenarios (Figure 2), while the two examples are two different takes of the same scenario. Each movie clip lasted 5s.

We chose a 2-Alternated Forced Choice protocol (2AFC) to evaluate the perceptual fidelity of the reduced marker set animations compared to the GS animations. This protocol allows us to directly compare two movies and to ask participants to give a binary answer (*same* or *different* in our case). In this experiment, participants viewed pairs of movies, where the first movie of each pair was always the reference GS animation and the second movie was one of the four methods (GS, 8F-4T, 4F-2T and static). Participants were instructed that the first movie was always the GS animation in order to evaluate the faithfulness of the second movie. After each trial, they were asked to answer if the second movie was the same as or different than the first movie, using the left or right mouse button, randomly interchanged per participant to avoid left-right bias.

The task included 108 pairs of movies: 9 scenarios × 2 examples × 3 methods (8F-4T, 4F-2T or static) × 2 repetitions. In order to test the difficulty of the task, participant accuracy, and to avoid a response bias, we also included one repetition of the GS animations against themselves, to which we expected participants to respond as "same". This brought the number of movie pairs to 126. Participants then viewed in random order these 126 pairs of movies. To train for the task, they first viewed some examples of pairs of movies that were not present in the experiment. Fifteen volunteers took part in this experiment (5F-10M). All participants were naïve

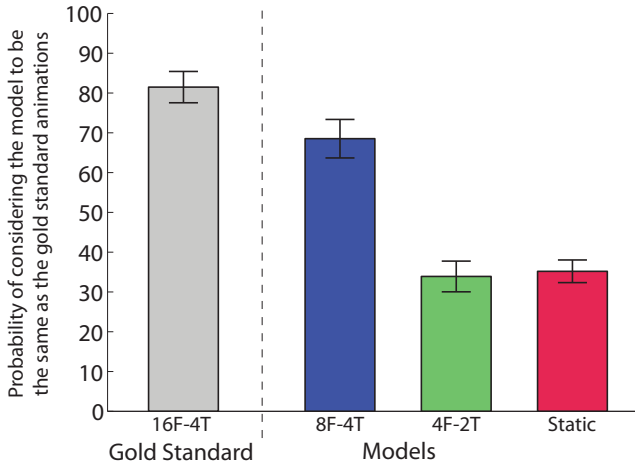



**Figure 5:** *Each bar represents the percentage of times participants answered "same" for the given models, when compared to the GS (vertical bars depict standard error).*

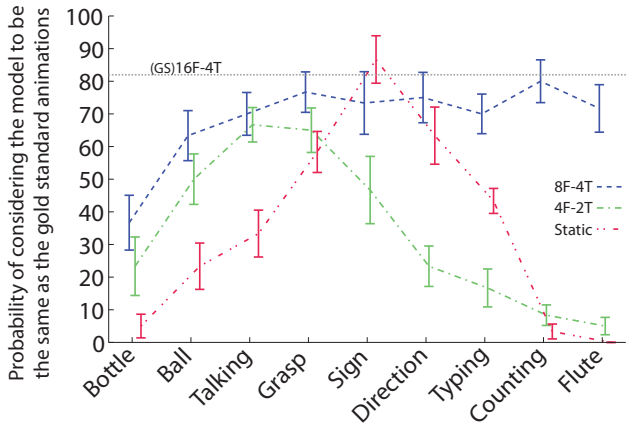



**Figure 6:** *Interaction between method and scenario in experiment 1. The average ranking of the Gold Standard animations against themselves (grey line) is provided for comparison.*

to the purpose of the experiment, came from various educational backgrounds, and received a book voucher for their efforts. During the experiment, participants were comfortably seated at their preferred distance from a 24-inch screen.

### 4.3 Results

Participant ratings were averaged over the two repetitions and the two examples of each scenario. We then performed a two-way repeated measures ANalysis Of VAriance (ANOVA) on this data with within-subjects factors *scenario* and *method* (8F-4T, 4F-2T or static). Post-hoc analysis was performed using standard Newman-Keuls tests for comparison of means. Significant effects are presented in Table 1 (top).

We found that *finger animations which used IK to compute the motion of the four fingers were considered to be perceptually similar to the GS animations*, in most cases (Figure 5). The 8F-4T model was considered to be the same as the GS animations 69% of the time, significantly more often than the two other models. In comparison, participants considered the GS to be the "same" as itself only 81% of the time, which shows the difficulty of the task. Therefore, *removing 8 markers from the fingers still retains most of the perceivable motion cues*. Figure 6 shows that all 8F-4T motions (besides the bottle opening) were perceived to be very similar to the GS. The difference for the bottle opening could be related to the fact that the IK solution tends to flatten the fingers, while the fingers computed with FK curl appropriately around the cap of the bottle.

Unsurprisingly, the results also demonstrated that the 4F-2T model does not allow the independence of motion between fingers that can be observed during directed hand motions, such as pointing or counting (Figure 7). However, *the 4F-2T model performs well when the displayed motions do not exhibit independence of motion between fingers*, such as in our examples of talking, grasping an object or playing with a tennis ball (Figure 6).

Finally, the absence of finger movements using a static hand pose can lead to a high perceptual similarity to captured animations when these animations do not display a lot of finger movement, such as in the case of direction, grasping and signing. This was surprising as we felt that participants would notice the absence of subtle secondary motions present in the GS. Therefore, *using a static pose from a database or one manually designed by an animator can produce natural animations in certain circumstances*.

**Figure 7:** *Left: GS animation. Right: 4F-2T. For some scenarios, the 4F-2T model fails to capture the independence between fingers (top: counting), but reconstructs the motion well in certain circumstances (bottom: talking).*

In order to evaluate if participants were answering significantly above chance level (i.e., significantly different from 50%), we carried out a single sample t-test for each method. We found that participants were anwering "same" above chance level for the 8F-4T method ($t(14) = 3.827, p < 0.005$), but were anwering "different" above chance level for the other methods (4F-2T: $t(14) = -4.170, p < 0.001$, Static: $t(14) = -5.195, p < 0.0005$).

## 5 Experiment 2

The previous experiment demonstrated that 12 markers per hands perform very well. In practice, this configuration would be difficult to capture in a large volume with large markers. Therefore, in this experiment we wished to determine if we could further decrease the number of markers while retaining most of the fidelity of the finger animation. The aim is to find an optimal marker configuration which balances the number of markers with the perceived quality of the motion.

### 5.1 Finger Animation Models

As the metacarpal bones of the fingers are quite rigid, we expected to be able to discard the markers at the base of the middle and ring fingers. Removing these markers, close to the corresponding markers of the index and pinky fingers, would allow us to use larger markers and thereby improve the marker recognition. Furthermore, the configuration of the two markers on the Interphalangeal joint of the thumb being incompatible with bigger markers, we also wished to evaluate if computing the thumb using IK with 2 markers affects the perception of the whole hand movements. Therefore, this experiment will evaluate additional marker configurations (Figure 8):

**6F-4T:** We removed the 2 markers at the base of the middle and ring fingers from the 8F-4T model (Figure 8.a), and recomputed them by linearly interpolating the position of the index and ring base markers, using linear interpolation in a similar manner to Equation 2. We expected this model to perform as well as the 8F-4T model.

**6F-2T:** This model is similar to the 6F-4T model, but uses only 2 markers to reconstruct the thumb (Figure 8.b). The goal is to evaluate if this simpler thumb model degrades the perceived quality of the finger animation.

**4F-4T:** This model is similar to the 4F-2T model but uses the thumb computed with FK (Figure 8.c) to test if the poor performance of the 4F-2T model was in fact due to the lack of independence between fingers, and not due to the simplified thumb.

(a) (b) (c)

**Figure 8:** *The three subsets of markers used for the second experiment: a) 6F-4T (10 markers), b) 6F-2T (8 markers) and c) 4F-4T (8 markers).*

### 5.2 Stimuli and Task

The previous experiment showed that the 8F-4T model was perceived to be very similar to the GS animations, but that removing the independence between fingers was not as efficient. For this experiment, we selected the 5 examples with the highest difference between the 8F-4T and 4F-2T models. As we showed that a static pose can be as effective as an animated hand in some cases, we discarded the motions for which the static pose performed well from this selection. We then used one example of each of the five motions: ball, direction, typing, counting and flute. We therefore generated 15 additional movies: 5 examples $\times$ 3 methods (6F-4T, 6F-2T and 4F-4T), with the same settings as before.

Fifteen new volunteers took part in this experiment (5F-10M). We chose the same 2AFC protocol as before, where participants viewed 30 pairs of movies in random order. Each pair depicted one generated animation (method) and the corresponding GS reference: 5 examples$\times$ 3 methods (6F-4T, 6F-2T or 4F-4T) $\times$ 2 repetitions. The first movie of each pair always presented the GS reference animation. Participants were then asked to answer if the second movie

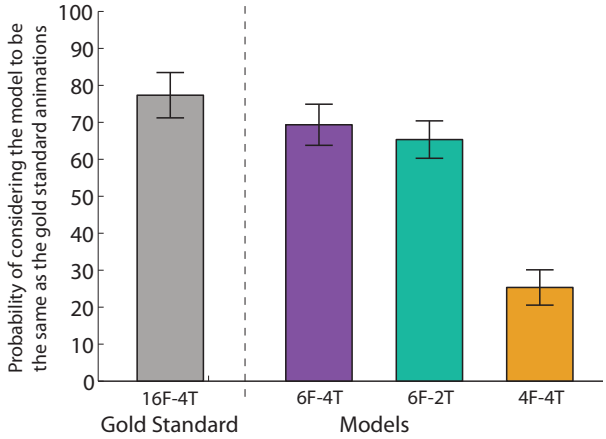



**Figure 9:** *Each bar represents the percentage of times participants answered "same" for the given models, when compared to the GS.*

was the same as or different than the first movie, in order to evaluate the perceptual fidelity of the generated animation compared to the reference GS. They first viewed some examples of movies that were not present in the experiment. As before, we also included one repetition of the GS animations against themselves, bringing the number of pairs to 35, to avoid a response bias and to test the difficulty of the task as well as participant accuracy.

### 5.3 Results

Participant ratings were averaged over the two repetitions of each example. We then performed a two-way repeated measures ANOVA on this data with within-subjects factors *example* and *method* (6F-4T, 6F-2T and 4F-4T). Significant effects are presented in Table 1 (bottom).

We found that *removing two markers from the base of the ring and middle fingers still produces animations which are perceived to be similar to the gold standard*. These animations were considered to be the same as the GS in 69% (6F-4T) and 65% (6F-2T) of cases (Figure 9). Because of the difficulty of the task, the repetition of the GS against itself was considered to be the same only in 77% of the cases, as in the first experiment. We also found that the simplification of the thumb model from 6F-4T to 6F-2T did not significantly change the perception of the finger motions.

As in the first experiment, the model which did not display independence between fingers (4F-4T) was perceived to be different from the GS, even though it had a higher quality thumb. Figure 10 shows that this model performed well only for the ball motion, which did not display independent movement of fingers.

As before, we carried out t-tests to evaluate if participants were answering significantly above chance level (i.e., significantly different from 50%). We found that participants were answering "same" above chance level for the 6F-4T and 6F-2T methods (resp. $t(14) = 3.477, p < 0.005$ and $t(14) = 3.031, p < 0.01$), but were answering "different" above chance level for the 4F-4T method ($t(14) = -5.174, p < 0.0005$).

**Cross-experiment analysis** In order to evaluate the effect of removing the markers at the base of the middle and ring fingers, as well as the effect of decreasing the number of markers on the thumb, we carried out a between-groups analysis on the results of the five motions that were used in both experiments.

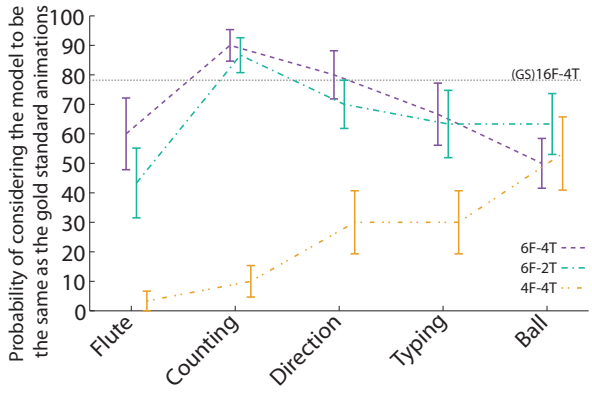



**Figure 10:** *Interaction between method and scenario in experiment 2. The average ranking of the Gold Standard animations against themselves (grey line) is provided for comparison.*

In this analysis, we compared pairs of different models: 8F-4T vs. 6F-4T, 8F-4T vs. 6F-2T, and 4F-2T vs. 4F-4T. The two first pairs were used to evaluate if removing the two base markers of the middle and ring fingers alters the perception of the produced animations. The last pair was used to evaluate the effect of removing the two markers on the Interphalangeal joint of the thumb.

The results did not show any main effect or interaction in any pair. Therefore, *removing the two markers from the base of the ring and middle fingers did not affect the perceived quality of the motions*. Also, *using only two markers for the thumb does not influence the perceived motion quality*.

## 6 Discussion

In this paper, we showed that people are not highly sensitive to all the subtle details of finger animation. Using Inverse Kinematics (IK) with a small number of markers produced animations that were perceived to be very similar to those generated using Forward Kinematics (FK) with a larger number of markers. From our observations, we have compiled a list of guidelines for finger motion capture:

- For the majority of cases, a simple 8 marker hand model (6 for the fingers, 2 for the thumb) should produce sufficiently high quality motions. This solution is particularly useful when simultaneously capturing full-body and finger movements in large capture areas, due to the fact that it allows the use of large markers.
- For motions where finger curvature is highly important, IK may flatten the fingers too much. In this case, we recommend using FK with a full set of markers in a small capture area.
- When independence between fingers is not important and processing time is limited, capturing only the thumb, index and pinky fingers (2 markers each) should produce reasonable results.
- If the fingers are only displaying secondary motion, a well chosen static pose may be adequate.

From a biomechanical point of view, the thumb is much more complex than the other fingers. However, we demonstrated that using either 4 or 2 markers for its reconstruction did not produce significantly different results. Perhaps this may have been due to the saliency of the fingers, rather than the quality of the thumb. Future eye-tracking studies would help to determine the most salient areas.

(a) (b)

**Figure 11:** *Motion capture setup with a smaller number of markers: a) glove used to capture the finger movements (6F-2T) and b) full body setup for one actor.*

Based on our results, we recommend using the simple 8 marker hand model (6 for the fingers, 2 for the thumb) to allow for the simultaneous capture of full-body and finger movements. This model provides a good trade-off between the perceived fidelity and the number of markers required to capture finger motion, and also captures the important synchrony which is used to convey information [McNeill 2005; Giorgolo and Verstraten 2008]. This solution should prove more efficient and economical than adding hand motion as a post-process.

A smaller number of bigger markers simplifies the motion capture process, in particular the tedious hand labeling post-process. Based on our experience, we are confident that labeling this simpler marker set reduces the manual post-processing time by almost an order of magnitude. Furthermore, the simpler marker set allows for more freedom of movement and more comfort than the larger set. However, in some cases, big markers could cause more problems than smaller ones when capturing interactions with objects or between actors (e.g., grasping hands).

Decreasing the number of markers can affect the quality of the captured hand motions. For instance, the finger configuration computed with IK is highly dependent on the finger bone length. In some cases, the finger length might need to be adjusted, to avoid unwanted popping of the finger because of slight marker movements. Computing the fingers' configuration using IK is also highly dependent on the length of each finger bone, which was computed from biomechanical data in our implementation. In order to improve the IK algorithm, a simple solution would be to roughly measure the different segments of the actor during the capture session. This would improve the biomechanical data while adding few constraints to the capture session.

To compute a unique IK configuration, we used simplifications from previous literature, such as the DIP/PIP dependency. Although this simplification is valid most of the time, it can however be broken by some actions, such as certain contacts with the environment or other fingers. More sophisticated algorithms could therefore be used to improve the finger reconstruction, such as example-based IK, but the quality of the resulting animation would depend significantly on the quality and number of examples. Furthermore, the construction of this database would still require a tedious high quality hand motion capture process.

Our animations also included hand-object interactions. However, we did not display the objects in the rendered movies to avoid the disturbing effect of incorrect contacts between hands and objects, which can occur even for high quality animations because of small differences between the real and virtual hand skeleton and objects. Our results can however be used as an input for hand-object contact algorithms.

**Validation** We conducted a two-person motion capture session where we simultaneously captured full-body and hand motion to test our proposed model in a complex setup (Figure 1). We used eight $14mm$ markers per hand for each actor (Figure 11), and 13 Vicon cameras to capture at 120Hz an area of approximately $6m \times 3m$. Our capture area limitation was only related to the physical space of the motion capture room, and we are confident that this setup would extend to a larger area, if available. For this test session, we captured various motions, such as conversing while walking, moving objects in cooperation and pushing each other. This capture setup was confirmed to be effective even while capturing two actors simultaneously, and we are confident that it would extend to multiple actors.

## Acknowledgements

We wish to thank the reviewers for their comments, and the participants in our experiments. This work was sponsored by Science Foundation Ireland as part of the Captavatar, NaturalMovers and Metropolis projects.

## References

ALBRECHT, I., HABER, J., AND SEIDEL, H.-P. 2003. Construction and animation of anatomically based human hand models. In *Proc. of SCA 2003*, 98–109.

CASSELL, J., PELACHAUD, C., BADLER, N., STEEDMAN, M., ACHORN, B., BECKET, T., DOUVILLE, B., PREVOST, S., AND STONE, M. 1994. Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In *Proc. of the 21st annual conference on Computer graphics and interactive techniques*, 413–420.

CHAI, J., AND HODGINS, J. K. 2005. Performance animation from low-dimensional control signals. *ACM Trans. Graph. 24*, 686–696.

CHANG, L., AND POLLARD, N. 2008. Method for determining kinematic parameters of the in vivo thumb carpometacarpal joint. *IEEE Trans. on Biomedical Engineering 55*, 7, 1897 – 1906.

CHANG, L., POLLARD, N., MITCHELL, T., AND XING, E. 2007. Feature selection for grasp recognition from optical markers. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2007*, 2944 –2950.

EDWARDS, S. J., BUCKLAND, D. J., AND MCCOY-POWLEN, J. D. 2002. *Developmental & Functional Hand Grasps*. Thorofare, New Jersey: Slack Incorporated.

EL-SAWAH, A., GEORGANAS, N., AND PETRIU, E. 2006. Finger inverse kinematics using error model analysis for gesture enabled navigation in virtual environments. In *IEEE International Workshop on Haptic Audio Visual Environments and their Applications, 2006*, 34 –39.

ELKOURA, G., AND SINGH, K. 2003. Handrix: animating the human hand. In *Proc. of SCA 2003*, 110–119.

ENNIS, C., MCDONNELL, R., AND O'SULLIVAN, C. 2010. Seeing is believing: body motion dominates in multisensory conversations. *ACM Trans. Graph. 29*, 91, 91:1–91:9.

GIORGOLO, G., AND VERSTRATEN, F. 2008. Perception of speech-and-gesture integration. In *Proc. of the International Conference on Auditory-Visual Speech Processing 2008*, 31–36.

GREINER, T. M. 1991. Hand anthropometry of US Army personnel. *Security, ADA244533*, 434.

HODGINS, J. K., O'BRIEN, J. F., AND TUMBLIN, J. 1998. Perception of human motion with different geometric models. *IEEE Trans. on Visualization and Computer Graphics 4*, 4, 307–316.

JIN, G., AND HAHN, J. 2005. Adding hand motion to the motion capture based character animation. *ISVC 2005*, 17–24.

JÖRG, S., AND O'SULLIVAN, C. 2009. Exploring the dimensionality of finger motion. In *Proc. of the 9th Eurographics Ireland Workshop (EGIE 2009)*, 1–11.

JÖRG, S., HODGINS, J., AND O'SULLIVAN, C. 2010. The perception of finger motions. In *Proc. of the 7th Symposium on Applied Perception in Graphics and Visualization*, 129–133.

KIPP, M., NEFF, M., KIPP, K. H., AND ALBRECHT, I. 2007. Towards natural gesture synthesis: Evaluating gesture units in a data-driven approach to gesture synthesis. In *Proc. of the 7th international conference on Intelligent Virtual Agents*, 15–28.

LEE, J., AND KUNII, T. 1995. Model-based analysis of hand posture. *IEEE Computer Graphics & Applications 15*, 5, 77 – 86.

LIU, C. K. 2009. Dextrous manipulation from a grasping pose. *ACM Trans. Graph. 28*, 59:1–59:6.

MAJKOWSKA, A., ZORDAN, V. B., AND FALOUTSOS, P. 2006. Automatic splicing for hand and body animations. In *Proc. of SCA 2006*, 309–316.

MCDONNELL, R., LARKIN, M., DOBBYN, S., COLLINS, S., AND O'SULLIVAN, C. 2008. Clone attack! perception of crowd variety. *ACM Trans. Graph. 27*, 26:1–26:8.

MCNEILL, D. 2005. *Gesture and Thought*. University of Chicago Press.

POLLARD, N. S., AND ZORDAN, V. B. 2005. Physically based grasping control from example. In *Proc. of SCA 2005*, 311–318.

REITSMA, P., ANDREWS, J., AND POLLARD, N. 2008. Effect of character animacy and preparatory motion on perceptual magnitude of errors in ballistic motion. *In Proc. of Eurographics '08*.

RIJPKEMA, H., AND GIRARD, M. 1991. Computer animation of knowledge-based human grasping. In *Proc. of the 18th annual conference on Computer graphics and interactive techniques*, 339–348.

SUEDA, S., KAUFMAN, A., AND PAI, D. K. 2008. Musculotendon simulation for hand animation. *ACM Trans. Graph. 27*, 83:1–83:8.

TSANG, W., SINGH, K., AND EUGENE, F. 2005. Helping hand: an anatomically accurate inverse dynamics solution for unconstrained hand motion. In *Proc. of SCA 2005*, 319–328.

WANG, R. Y., AND POPOVIĆ, J. 2009. Real-time hand-tracking with a color glove. *ACM Trans. Graph. 28*, 63:1–63:8.

WU, Y., AND HUANG, T. 2001. Hand modeling, analysis and recognition. *IEEE Signal Processing Magazine 18*, 3, 51 –60.