The Perception of Finger Motions

Sophie Jörg*
Trinity College Dublin

Jessica Hodgins†
Carnegie Mellon University
Disney Research Pittsburgh

Carol O'Sullivan‡
Trinity College Dublin

Figure 1: *Conversational gesture used to compare the quality of animation methods (motion captured movements, close-up view).*

Abstract

In this paper, we explore the perception of finger motions of virtual characters. In three experiments, designed to investigate finger animations, we asked the following questions: When are errors in finger motion noticeable? What are the consequences of these errors? What animation method should we recommend? We found that synchronization errors of as little as 0.1s can be detected, but that the perceptibility of errors is highly dependent on the type of motion. Errors in finger animations can change the interpretation of a scene even without altering its perceived quality. Finally, out of the four conditions tested – original motion capture, no motions, keyframed animation and randomly selected motions – the original motion captured movements were rated as having the highest quality.

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

Keywords: Finger animation, perception of human motion, virtual characters, motion capture

1 Introduction

Hand and finger motions are omnipresent in daily life. We use our hands naturally to punctuate our speech or to handle complex tools. In fact, as psycholinguistic research shows, hand movements and finger gestures play a crucial role in communicating information and meaning [McNeill 1992]. Given the importance of such motions in real life, they are also likely to affect the perception of virtual characters, especially when those characters are very human-like or interact with people.

Motion capture has become a widespread technology to animate virtual characters for movies and games. The subtle motions of the fingers, nevertheless, are complex to capture. Due to their small scale, the high degrees of freedom of hands, and frequent occlusions, optical motion capture of fingers requires careful planning, a larger number of cameras to cover the same space, and laborious manual post-processing. Wired gloves, as an alternative, produce less accurate results and need to be calibrated regularly, thereby interrupting the capture session. Therefore, finger movements are often animated manually. It is also possible to capture them separately from the body and to subsequently synchronize them, which might lead to visible artifacts [Majkowska et al. 2006]. Little is known about the perception of such artifacts or about the effect of missing finger motions. We evaluate the perceptibility, saliency, and perceived quality of errors in finger motions. Although, strictly speaking, *finger* denotes four digits and excludes the thumb [Palastanga et al. 2006], for simplicity we use this term to refer to all five digits. Furthermore, we concentrate on the movements of all phalangeal joints and the carpometacarpal joint of the thumb, thus excluding the wrist.

There are three questions that we wish to address: Firstly, what threshold of error can be detected when body motions and finger motions are not adequately aligned in time? Animations below this just noticeable error in synchronization would not need to be considered. However, as humans have impressive abilities to recognize detail in human body motion, it is likely that they are very good in detecting erroneous gestures. Secondly, what are the consequences of errors on the interpretation of a scene? Because finger motions play an important role in the communication of meaning, incorrect movements might confuse the viewer and change the understanding of a scenario. We therefore evaluate the impact of finger movements on participants' interpretation of a scene. Lastly, which type of animation method produces the best perceived quality? We establish a ranking between no motion at all, motion captured movements, keyframed animation and randomly selected motion captured movements. For each question, we produced appropriate stimuli (see Figure 1) and ran a set of perceptual studies, described in Sections 3–5.

2 Related Work

People are remarkably skilled at recognizing human motion. The impressive ability to identify human walks within milliseconds based on only 12 point lights was first demonstrated by Johansson [1973]. As subsequently specified in several studies, very subtle details can be distinguished and interpreted differently [Cutting and Kozlowski 1977; Atkinson et al. 2004; Mather and Murdoch 1994]. Perceptual thresholds and the effects of motion manipulations for virtual characters have been evaluated by various researchers. Reitsma and Pollard [2003] showed that participants in

*e-mail: sophie.joerg@cs.tcd.ie
†e-mail: jkh@cs.cmu.edu
‡e-mail:carol.osullivan@cs.tcd.ie

their study were sensitive to velocity and acceleration errors created from editing motion capture sequences containing ballistic motion. McDonnell et al. [2009] analyzed the perception of desynchronized body motions in conversations.

Despite the important role of hand and finger motions in our daily interactions [McNeill 1992; Napier 1980], we are not aware of previous studies that focus on the perception of finger motion manipulations. Cassell et al. [1994] highlighted the importance of correct gestures and hand motions for virtual humans. Nevertheless, research on embodied conversational agents that aims at generating natural gestures from speech does not address the motions of the fingers specifically [Cassell et al. 2001; Stone et al. 2004].

In the field of modeling and animation of hands, however, various new methods have been introduced recently. Sueda et al. [2008] and Albrecht et al. [2003] presented highly detailed models of human hands, which simulate their anatomical structure. ElKoura and Singh [2003] developed a procedural algorithm to generate complex animations of a hand playing the guitar. Grasping motions have been studied extensively due to their complexity and their importance in robotics [Liu 2009]. This work illustrates the level of activity in this area, and our studies aim to bridge the gap between research on finger animation and motion perception.

3 Just Noticeable Errors

In our first experiment, we show animations with four levels of errors in finger synchronization to determine the amount of error detectable by the viewer.

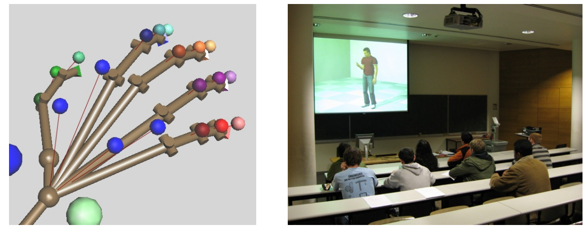
3.1 Stimuli

We generated four short sequences involving finger motions: counting, drinking from a bottle, pointing, and snapping. We chose the gestures to be as varied as possible. *Count* and *Snap* have cyclical repetitions in their main part, while *Point* has only the common three phases of gestures (preparation, main part, and retraction), and *Drink* has a more complex structure. Furthermore *Snap* and *Drink* include interactions (with a bottle or between the fingers). All are self-explanatory, very common and easy to recognize.

We simultaneously captured the body and finger motions of a male performer, using 19 markers attached on each hand and 48 on the body, with a 12 camera Vicon optical system. Three versions of each action were recorded to prevent the participants from becoming too familiar with the motions resulting in 12 distinct original gestures. The movements of the skeleton were then calculated (see Figure 2a) and displayed on a virtual human (see Figure 3). For each version and action, four levels of modification were generated. The keyframes of the finger joints of both hands were moved in time, so that their motions were delayed by 0.1, 0.2, 0.3, and 0.4 seconds (3, 6, 9, and 12 frames). Modified and unmodified motions were counterbalanced. As a result every participant watched 4 (actions) x 4 (modifications) x 3 (versions) x 2 (to counterbalance) = 96 motion clips.

3.2 Experimental Design

Twenty-two participants (4f, 18m, aged 18-35) viewed these clips in two groups and in different random orders on a large screen in a lecture theater (see Figure 2b). At the beginning of the experiment, we explained what motion capture is by showing a video of an actor in a motion capture suit and the corresponding virtual character performing the same motion. Then, for each clip, participants were asked: *Is this motion MODIFIED or NOT MODIFIED*? Between each movie a black screen was displayed for 3 seconds, during which participants noted their answer and were alerted to the next stimulus by a sound. The experiment took about 25 minutes in total. Each participant was compensated with a 5 euro book token.



(a) (b)

Figure 2: *(a) Hand skeleton calculated based on the marker positions (b) Group participating in the experiment*

3.3 Data Analysis

A two factor (motion type and level of synchronization error) analysis of variance (ANOVA) with repeated measures showed a main effect of motion type ($F(3,60)=15.4$, $p<0.001$) and of synchronization error ($F=(4,80)=17.6$, $p<0.001$). There was also an interaction effect between motion type and synchronization error ($F(12,240)=6.6$, $p<0.001$). We then performed a post-hoc analysis using standard Newman-Keuls tests for pairwise comparisons among means for each motion type.

As can be seen from the results summarized in Figure 3, for the Count motion, there were significant differences between the original and the motion with an error of 0.1 seconds as well as between the versions desynchronized by 0.1 and by 0.2 seconds. In general, the error was detected more often for increasing error levels. A similar result, albeit to a lesser degree, was found for the Drink motion, where the synchronization errors could be recognized for errors bigger than 0.3 seconds. However, for the Snap motion a delay of 0.4 seconds was not detected at a significant rate, whereas a smaller error of 0.1 seconds was. Finally, for the Point gesture there was no significant effect of desynchronization level at all.

A reason for the observed results for Snap and Count could be the cyclic nature of those motions. Count has a longer cycle than Snap, which explains why the percentage of clips rated as unmodified does not increase significantly within the analyzed error levels. The recognition rate was the highest for Snap, which had the highest velocity and energy of the fingers. This observation suggests that velocity or energy violations might be a factor in the perception of errors in motions. It would be very difficult to carry out a snapping action with desynchronized finger movements, whereas it is physically easy to perform a pointing action with delayed finger motion. This fact might explain why there was no significant effect of the synchronization errors at all for the Point gesture.

3.4 Summary

Our first experiment provides strong evidence that perceptibility thresholds for this type of error are significantly affected by the type of action being performed. We illustrated that the correct timing of hand motions can be very important, as people were able to recognize errors of only 0.1 seconds (3 frames). However, no single threshold for the perceptibility of errors in synchronization was found, as for one test motion even errors of 0.4 seconds went undetected.

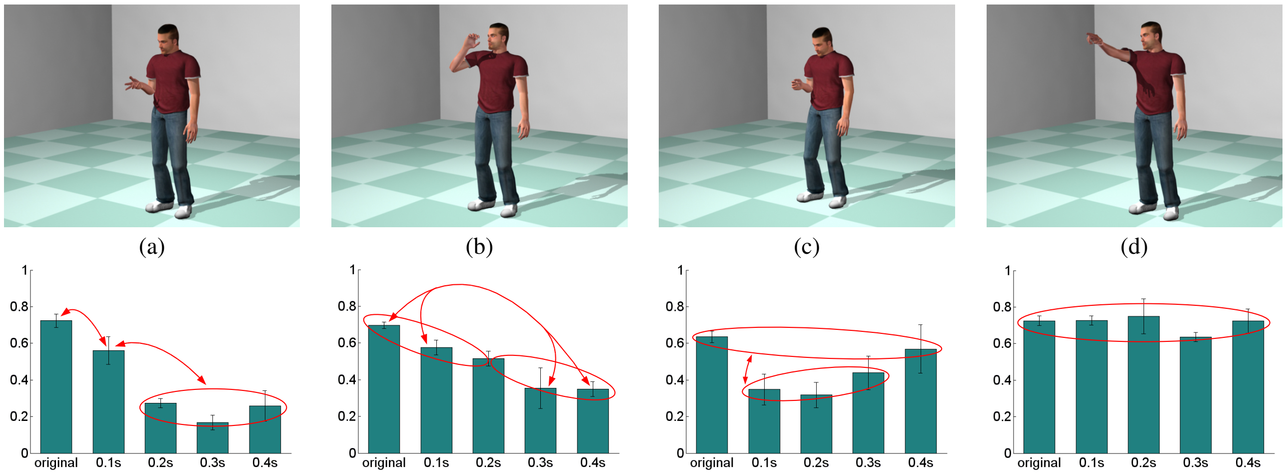


Figure 3: *Screenshots and results for: (a) Count, (b) Drink, (c) Snap, and (d) Point. The graphs show the percentage of motions rated as unmodified (y-axis) for the original motion and an increasing level of error (x-axis). Values that are not significantly different from each other are circled, whereas significant differences are indicated with arrows. Error bars represent one standard error of the mean.*

Figure 4: *Scene from the Computer Crash Vignette*

4 Interpretation of Altered Finger Movements

Our second experiment investigates the impact of incorrect finger motions on the interpretation of a scenario by asking high-level questions.

4.1 Stimuli

We generated a short movie with emotional content to study the impact of erroneous finger movements on the interpretation of a story. Our scenario depicts an angry reaction to a *Computer Crash* (see Figure 4). Our setup consisted of 18 cameras covering a 3m x 4m capture area. The actor was wearing a suit with 48 markers on the body and 19 markers on each hand, which resulted in 86 markers in total. The motion captured male actor was instructed that, when his computer freezes while typing something important, he should become increasingly frustrated and first try to get a response through clicking the mouse, then trying ctrl-alt-del, before finally throwing the monitor to the ground.

We created two versions of the 30 seconds long scene – one original and one where we delayed the motions of the finger joints of both hands by 0.5 seconds. The motions were then displayed on a skinned model in a rendered environment representing an office at night.

4.2 Experimental Design

The unmodified version was shown to a group of 47 participants (11f, 36m), while the modified version was shown to a second group of 69 participants (20f, 49m). In each case the movies were displayed on a large projected screen in a lecture theater. All participants were naïve as to the purpose of the experiment and all had some experience with computers. After watching the vignette, they wrote down their answers to a number of questions of three types: the first type required them to rate their empathy with the character on a 5-point scale (e.g., *How angry would you rate the character in the animation?*); the second were factual questions (e.g., *How often does the character try to move the mouse?*) and the third type required them to describe the scene in free form text. Finally, they were shown the vignette a second time and asked to pay attention to the quality of the animation, rating it on a scale of 1 to 5 and justifying their scores.

4.3 Data Analysis

Using between groups ANOVA to analyze the five-point scale response data, we found that the error did not introduce any significant difference in viewer empathy toward the character or to the factual questions. Contrary to our expectation, we did not find significant differences between groups for the quality rating. In fact, participants from both groups even made comments on how realistic the finger animations were. The most interesting results were found from examining the written descriptions. In the unmodified version, 21% of participants mentioned the character using ctrl-alt-del compared to only 1% of the participants who saw the desynchronized version. Furthermore, 40% mentioned the computer freezing in the unmodified version, but only 15% when the finger motion was altered; mostly, participants in this group deduced that his anger was due to something annoying he saw on screen or because he was frustrated when he could not do something correctly. This suggests that the modifications to the finger motions subtly but definitely altered the meaning of the vignette.

4.4 Summary

Our second experiment shows that a change in the finger motions of a virtual character altered the interpretation of a scene without changing the perceived quality of the animation. This result shows the importance of correct hand animation to make sure that the intended interpretation of a scenario is conveyed, even if viewers do not actively report that the motion is erroneous.

5 Comparing Finger Animation Methods

In our third experiment we use a pairwise comparison task to evaluate the perceived quality of four techniques to create finger animations.

5.1 Stimuli

A male actor was recruited and asked to narrate a common topic, in our case the content of a television series. His body and finger motions were captured simultaneously, using 20 markers attached on each hand and 48 on the body, with a 13 camera Vicon optical system. From a 2 minutes 20 seconds motion captured narration, we chose 4 scenes of 3–6 seconds each, mostly consisting of beats and cohesive gestures that accompany speech. Finger motions in these type of gestures are quite subtle. They do not exhibit a specific meaning as the motions chosen in our first experiment, but are used to punctuate and structure the speech.

For each scene we created four types of finger animations. In the *original* condition the motion captured finger movements were displayed unaltered while in the *no hands* condition the fingers were kept immobile in a natural resting position throughout the complete clip. To create the *keyframe* condition, we animated each scene using up to six keyframes per hand. A skilled animator can create completely realistic hand motions with keyframes, the only limiting factor being time. We wanted to compare methods used in a typical production process with limited resources. Basic animations with only a few poses can be generated reasonably quickly. Therefore, we limited the number of keyframes to six per scene for each hand and the number of poses to three: one with the fingers in a relaxed position, one where only the index was extended while the remaining fingers were kept relaxed, and one with all fingers extended. We then animated each clip, while aiming to replicate the motion captured movements as accurately as possible within these constraints. The motion between the keyframes was interpolated using flat tangents. Lastly, the *random* condition examined if small movements, even when incorrect, are preferred over the unnatural case of completely stiff fingers. We cut the keyframe curves of the finger joints of our full animation in two parts and interchanged the two halves. As the original scene contains mainly subtle finger motions, we obtained small random motions of the fingers.

Each animation was displayed and rendered with the same character and environment as in our first experiment. In addition to the long shots, we rendered close-ups, in which the character was visible from the hips to the neck (see Figure 1). We obtained 4 (scenes) x 2 (shot types) x 4 (conditions) = 32 different animations, which we showed in pairs, keeping the clip and shot type constant.

5.2 Experimental Design

Twenty-four participants (10f, 14m, aged 21-45) in 6 groups were shown pairs of clips on a large screen. We compared all conditions against each other, resulting in 6 (condition pairs) x 4 (scenes) x 2 (shot types) = 48 pairs of clips. Both, the order of the 48 pairs and the order of the two clips within a pair, were randomized for each group. For each clip, we asked: *In which clip was the motion of better quality?* Similar to the procedure in the first experiment, a black screen was displayed for 3 seconds between each movie, during which participants noted their answer and were alerted to the next stimulus by a sound. The experiment took about 20 minutes in total and each participant was compensated with a 5 euro book voucher.

During debriefing it became clear that many participants misinterpreted one of the scenes, in which the performer moved his middle finger independently of his other fingers. Many participants reported this particular motion as unnatural, so we removed it from the evaluation.

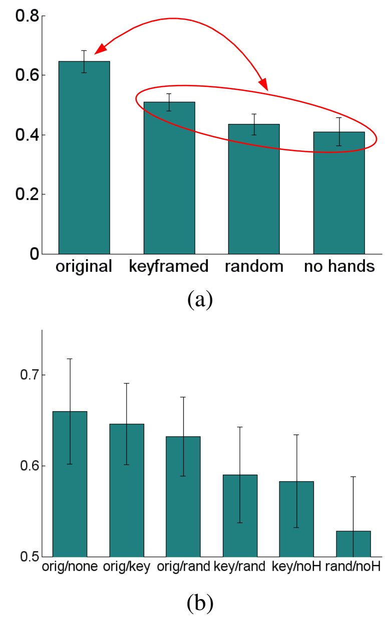


Figure 5: *(a) Overall preferences of animation method. Values that are not significantly different from each other are circled, whereas significant differences are indicated with arrows. (b) Pairwise preferences of animation method. The choices for the three pairs that include the original condition are significant. In both graphs error bars represent one standard error of the mean.*

5.3 Data Analysis

We merged the results of the three remaining scenes and shots and carried out statistical analysis in order to verify the significance of overall preferences. The results are summarized in Figure 5. As expected, the original motion was preferred overall. In total, participants rated the original clip 64.6% of the time as being of better quality. The keyframed, random, and no hands conditions were selected 50.9%, 43.5%, and 41.0% of the time, respectively. T-tests for comparing the means of dependent samples showed that the original condition was checked significantly more often than each of the three other conditions (all $p<0.05$), with no significant differences between the keyframed, random, and no hands animation. The difference between the keyframe animation and the no hands condition failed to reach significance.

The analysis of the pairwise preferences with T-tests for comparing the means provided consistent results, as can be seen in Figure 5 (b), with the choices for the three pairs that include the original condition being significant.

Contrary to our expectations, the viewpoint, i.e., long shot over close-up, had no significant effect on the perceived quality of the clips. We found this result noteworthy, as it signals that finger motions may be equally salient in near and far shots (up to a certain limit).

5.4 Summary

Of the four conditions tested, the original motion-captured movements were judged as being qualitatively the best, which is what we expected. However, although we had predicted further preference differences between the three other conditions, these were not significant. A closer look at the individual participants showed that, out of those who had strong preferences for one condition over another, there was not always consensus amongst the participants on which condition had the better quality. For example two participants judged the no hands condition to have the best quality consistently throughout the whole experiment (selected at least 5 times out of 6 for each pairwise comparison). Furthermore, as previously explained, we found that even captured motions can look unnatural if the performed motion is not as expected. In such cases, the idealized motions of keyframed animations might be preferred.

6 Conclusion and Future Work

With these three experiments we provide the first set of insights into the perception of anomalous finger movements on virtual humans. We show that very subtle errors can still be detected, such as a 0.1s desynchronization error. Furthermore, we demonstrate the difficulty of producing finger animation in an automatic way, as none of our three tested animations reached the perceived quality of motion captured movements. However, this realism is very important in order to convey the intended meaning of a scenario, as exposed in our second experiment.

Many new questions now need to be answered. For example, it would be interesting to determine which features of a motion contribute to the perceptibility of synchronization errors. Do velocity, contacts, energy violations, or familiarity play a role? Are those results transferable to other virtual characters, e.g., more or less realistic ones? We know from Hodgins' study [1998] that this might not be the case. Would further factors, such as the quality of the rendering, the lighting, or the complexity of the scene, influence the results? Speech and gestures are highly correlated, so would the presence of audio and lip synchronization therefore alter our results? Finally, from our experiments we cannot conclude that all finger motions are interpreted the same way, even if they are not judged as being of different quality or as being erroneous. Further work is clearly needed to understand the complexity of human perception of finger animation.

Acknowledgments

We would like to thank Moshe Mahler and Justin Macey from Carnegie Mellon University for their help in creating the motion capture data for the two first experiments and the models. Furthermore, we thank all the subjects who participated in our perceptual experiments! This project has been supported by the Embark Postgraduate Scholarship Scheme of the Irish Research Council for Science, Engineering, and Technology (IRCSET), the National Science Foundation, and Science Foundation Ireland.

References

ALBRECHT, I., HABER, J., AND SEIDEL, H.-P. 2003. Construction and animation of anatomically based human hand models. In *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 98–109.

ATKINSON, A. P., DITTRICH, W. H., GEMMELL, A. J., AND YOUNG, A. W. 2004. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception 33*, 6, 717–746.

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 *Proceedings of SIGGRAPH 1994*, ACM Press / ACM SIGGRAPH, 413–420.

CASSELL, J., VILHJÁLMSSON, H. H., AND BICKMORE, T. 2001. Beat: the behavior expression animation toolkit. In *Proceedings of SIGGRAPH 2001*, ACM Press / ACM SIGGRAPH, 477–486.

CUTTING, J. E., AND KOZLOWSKI, L. T. 1977. Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society 9*, 5, 353–356.

ELKOURA, G., AND SINGH, K. 2003. Handrix: animating the human hand. In *SCA '03: Proceedings of the 2003 ACM SIGGRAPH/Eurographics Symposium On Computer Animation*, 110–119.

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

JOHANSSON, G. 1973. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics 14*, 2, 201–211.

LIU, K. C. 2009. Dextrous manipulation from a grasping pose. *ACM Transactions on Graphics 28*, 3 (Aug.), 3:1–3:6.

MAJKOWSKA, A., ZORDAN, V. B., AND FALOUTSOS, P. 2006. Automatic splicing for hand and body animations. In *SCA '06: Proceedings of the 2006 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, 309–316.

MATHER, G., AND MURDOCH, L. 1994. Gender discrimination in biological motion displays based on dynamic cues. *Proceedings of the Royal Society of London, Series B 258*, 273–279.

MCDONNELL, R., ENNIS, C., DOBBYN, S., AND O'SULLIVAN, C. 2009. Talking bodies: Sensitivity to desynchronization of conversations. In *APGV '09: Proceedings of the 6th Symposium on Applied Perception in Graphics and Visualization*.

MCNEILL, D. 1992. *Hand and Mind: what gestures reveal about thought*. The University of Chicago Press.

NAPIER, J. 1980. *Hands*. New York: Pantheon Books.

PALASTANGA, N., FIELD, D., AND SOAMES, R. 2006. *Anatomy and Human Movement – Structure and Function*, 5 ed. New York: Basic Books.

REITSMA, P. S. A., AND POLLARD, N. S. 2003. Perceptual metrics for character animation: sensitivity to errors in ballistic motion. In *ACM Transactions on Graphics 22*, no. 3 (July), 537–542.

STONE, M., DECARLO, D., OH, I., RODRIGUEZ, C., STERE, A., LEES, A., AND BREGLER, C. 2004. Speaking with hands: creating animated conversational characters from recordings of human performance. In *ACM Transactions on Graphics 23*, no. 3 (Aug.), 506–513.

SUEDA, S., KAUFMAN, A., AND PAI, D. K. 2008. Musculotendon simulation for hand animation. *ACM Transactions on Graphics 27*, 3 (Aug.), 3:1–3:8.