The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies

Radu-Daniel Vatavu
MintViz Lab | MANSiD Research Center
University Ştefan cel Mare of Suceava
Suceava 720229, Romania
radu.vatavu@usm.ro

ABSTRACT

We introduce the dissimilarity-consensus method, a new approach to computing objective measures of consensus between users' gesture preferences to support data analysis in end-user gesture elicitation studies. Our method models and quantifies the relationship between users' consensus over gesture articulation and numerical measures of gesture dissimilarity, e.g., Dynamic Time Warping or Hausdorff distances, by employing growth curves and logistic functions. We exemplify our method on 1,312 whole-body gestures elicited from 30 children, ages 3 to 6 years, and we report the first empirical results in the literature on the consensus between whole-body gestures produced by children this young. We provide `C#` and `R` software implementations of our method and make our gesture dataset publicly available.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Gestural input**.

KEYWORDS

Gesture input; Gesture elicitation; Consensus; Growth curves; Logistic model; Children; Whole-body gestures; Dataset.

ACM Reference Format:

Radu-Daniel Vatavu. 2019. The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies. In *CHI Conference on Human Factors in Computing Systems Proceedings (CHI 2019), May 4–9, 2019, Glasgow, Scotland UK*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3290605.3300454

ACM ISBN 978-1-4503-5970-2/19/05…$15.00
https://doi.org/10.1145/3290605.3300454

1 INTRODUCTION

Gestures enable users to operate devices fast and intuitively by means of direct input on touchscreens [77], wrist control on smartwatches [22], head movements for augmented reality glasses [28], feet input for locomotion interfaces in virtual reality [70], mid-air hand shortcuts for peripheral interaction [55], and whole-body gesture input for video games [56]. However, designing effective gesture interaction and a rewarding user experience requires key knowledge about what gestures are intuitive [72], low-fatigue [30], efficient to perform [33,53], and straightforward to recall [43].

Understanding users' preferences for gestures they would like to use represents an important step towards effective gesture UI design. To this end, the gesture elicitation methodology [68,71,72] has proven immensely resourceful for designers to form an understanding of users' mental models of gesture interaction. Since its first implementation for multi-touch gestures [72], the methodology has been reapplied for a variety of gesture types and applications [7,13,20,31,41,52, 53,56,60,62,69], revealing users' preferences for interactive gestures and accumulating important design knowledge.

One of the outcomes of any gesture elicitation study is a set of recommendable gestures together with an estimation of users' consensus or agreement [68,71] as a measure of the expected intuitiveness of those gestures [72]. Consensus has been evaluated numerically by clustering the elicited gestures into classes of similar types [72], a procedure performed manually by a human and guided by a set of clustering criteria. Unfortunately, the criteria used to evaluate which gestures are similar vary from study to study [7,13,20,31,41,52,56,60,62], causing the magnitude of consensus to depend on the specific and subjective criteria chosen by the practitioner. While a subjective approach to the interpretation of gestures is indispensable when trying to understand the common meaning of equivalent, yet structurally different gestures, such as cultural gestures [4,19], or when applying the principles of somaesthetics [24,32,40] to inform movement-based interaction, the vast majority of gesture elicitation studies are conducted to find objective consensus in gesture articulation. For the later case, the use of subjective criteria to cluster elicited gestures may lead to

Figure 1. Thirty children performing whole-body gestures with variations in body pose, handedness, amplitude of movement, hand poses, etc. What criteria should be used to assess the similarity of any two body gestures in order to understand the consensus between these children's gestures? Unfortunately, this is where the practitioner's subjectivity intervenes with a direct influence on the magnitude of reported consensus; see the text for a numerical example.

different consensus results for the same data, a distressing outcome, as we are about to show with a numerical example. The alternative, which we introduce in this paper, is *a holistic approach to understanding the numerical relationship between gesture similarity and users' consensus over articulation*, a process that can be conducted entirely on a computer to deliver *objective magnitudes of consensus in just a few seconds.*

Subjectivity in Reporting Consensus in End-User Gesture Elicitation Studies: A Motivating Example

We provide an example to illustrate the dependence of the magnitude of reported consensus on the criteria employed to cluster elicited gestures into classes of similar types. We also use this example to introduce new readers to the principles of the end-user gesture elicitation methodology [72].

Consider a designer that wishes to understand children's preferences for whole-body gestures to symbolize a cat scratching (an action denoted in the following as the "referent," according to the terminology from Wobbrock *et al.* [72]) in order to inform a technique to detect such gestures effectively for a gesture-controlled video game. Following the steps of the gesture elicitation methodology [72], the designer assembles a group of children, e.g., N=30 children, representative of the target user group, presents them with the desired effect, i.e., a cat scratching, and asks each child to perform a gesture to generate that effect. At the end of the experiment, the designer has recorded 30 gestures, such as the ones illustrated in Figure 1. Clearly, not all the gestures will have the same articulation, but rather gestures will vary in terms of, e.g., the amplitude of movement, the use of the dominant, nondominant, or both hands, execution speed, repetition of movement, hand poses to suggest claws, body poses and facial expressions to suggest a cat, and so on. The designer wishes to understand how much consensus exists in the data they collected, preferably as a value between 0% and 100%, where 0% means no consensus (i.e., all the gestures are different) and 100% denotes perfect consensus.

Consensus has been computed in the literature as a two-step procedure: (1) elicited gestures are clustered into classes of similar types, and (2) the cardinalities of all the clusters are aggregated into a numerical measure of consensus, such as the Agreement Rate [68,72]. To implement the first step, the designer decides which gestures are "similar" by defining and employing a set of criteria to cluster the elicited gestures. And here is where subjectivity intervenes. Let's say that two "scratch like a cat" gestures are judged to be similar if they are performed with the same hand or with both hands "scratching" simultaneously. Using this *handedness* criterion, the designer evaluates the magnitude of consensus at 39.3%, meaning that, of all pairs of gestures, 39.3% are similar.[1] But what about body pose? Is it important if the child stands straight up, sits down on the floor, crouches, walks around, and so on, while scratching like a cat? With the *body pose* criterion added, consensus drops at 19.6%.[1] But what about the pattern of the hand(s) moving to symbolize scratching? Is

[1]This example reports actual consensus values from our data (N=30 children) using the Agreement Rate measure of Vatavu and Wobbrock [68].

scratching with the left hand followed by the right the same gesture as scratching with the right hand followed by the left and then the right hand again? By considering the *pattern* criterion, consensus drops further at 12.2%.[1] What about repetitive movements? Is scratching once the same gesture as scratching multiple times? With *repetition* as the fourth criterion, consensus drops at 4.7%,[1] a value that is 8 times smaller than the original magnitude of consensus computed with the *handedness* criterion only. What about the location where scratching takes place, the speed of the scratching movement, or the hand pose that symbolizes a claw? Should all these criteria, or others, be considered? The answer may depend on many factors (e.g., the application requirements, the resolution of the sensor, the goals of the investigation, etc.) but, it is clear that, by considering more criteria, the magnitude of consensus can eventually become 0% for this example. Our designer will compromise somewhere between 0% and 39.3%, favoring some criteria and dismissing others, but it is evident now how the choice of the clustering criteria affects the magnitude of reported consensus.

Contributions

We contribute an alternative approach to subjective clustering criteria and manual labeling of elicited gestures and introduce the "dissimilarity-consensus" method (abbreviated τ-C) by adopting *a holistic perspective on computing and understanding how consensus forms.* Moreover, the process is implemented entirely by a computer, which transforms long hours and even days of manual, subjective clustering into obtaining reliable, objective results in a matter of just a few seconds. Our method employs growth curves, modeled with logistic functions, that describe how fast consensus increases in response to an increase in the tolerance in gesture dissimilarity that is allowed when judging how similar two gestures are. Our practical contributions are as follows:

(1) We introduce the dissimilarity-consensus method (τ-C) for computing objective magnitudes of consensus in end-user gesture elicitation studies.
(2) We demonstrate the τ-C method on whole-body gestures elicited from children aged 3 to 6 years, a user group that we specifically chose to maximize the variance of elicited gestures (children at this age are still developing their motor and cognitive skills) and, thus, maximize the influence of clustering criteria on the magnitude of consensus reported with the traditional approach.
(3) We release software implementations in C# and R to compute consensus and visualize τ-C growth curves to support data analysis for end-user gesture elicitation studies towards accumulation of new gesture knowledge in our community. We also release our dataset of 1,312 whole-body gestures produced by 30 children to foster advances in designing gesture interaction for small children.

2 RELATED WORK

We relate in this section to prior work on gesture elicitation studies and review whole-body interaction for children.

End-User Gesture Elicitation Studies

Wobbrock *et al.* [71] introduced the elicitation methodology in 2005 as a practical implementation of a participatory design study to evaluate and maximize the "guessability" of symbolic input with an application to text entry and the EdgeWrite [73] alphabet. The first application to gestures (2009) was for multitouch input on tabletops [72]. Since then, many gesture elicitation studies have been conducted to unveil end-users' preferences for a variety of gesture types and gesture-controlled applications, such as video games [56], augmented reality [46], TV control [62,75], interaction with multiple displays [54], pairing devices [31], keyboard shortcuts [7,20], web applications [41], motion gestures on smartphones [52], interacting with drones [12], smart environments [36], on-skin input [8], elastic and deformable displays [60], input on smart rings [21], smartwatches [5], mid-air gesture input for connected cars [38], designing consistent gesture commands across devices [16], etc. Over the years, the original method [72] has been refined with an updated formula for computing agreement [20], measures of disagreement and co-agreement [68], adaptations to within-subjects and between-subjects designs [68,69] and using crowdsourcing to elicit users' preferences [1]. Ticklish aspects, such as the "legacy bias" [42], i.e., the influence of users' prior experience with interactive systems on elicited gestures, were addressed with methodological variations, such as production, priming, and partners [23,42], soft-constraints [53], and the "framed guessability" approach [10]. However, the essential part of computing consensus by manual clustering of gestures has remained unchanged.

Studies on Children's Whole-Body Gestures

In this work, we demonstrate our new dissimilarity-consensus method by applying it to whole-body gestures elicited from small children, ages 3 to 6 years. To connect our results with the literature on child-computer interaction, we review in this section prior work conducted to understand how children perform whole-body gestures.

One way to acquire data about children's gestures is through observational studies, where children are placed in interactive contexts and their actions observed. For example, Rahman *et al.* [49] examined children's perceptions of whole-body gesture interaction (Kinect) *vs.* touch input (iPad); Hoysniemi *et al.* [25,26] conducted a Wizard-of-Oz study to observe children's whole-body movements to control an avatar; and Connell *et al.* [13] collected gestures from 6 children, ages 3 to 8 years, in response to 22 referents regarding object manipulation, navigation, and spatial interaction tasks.

Two recent studies [17,29] revealed differences between children and adults' whole-body gesture articulations. Jain *et al.*'s [29] perceptual study showed that the gestures of children between 5 and 9 years old are perceived differently by human observers than the same gesture types performed by an adult. This result has implications for recognition techniques that should be tailored to children's specific ways to articulate gestures, but also to data synthesis of "child-like" movements [17] to support creation of realistic child characters for computer animations and video games.

These studies have contributed with important knowledge on how children, of various age groups, perform gestures. However, compared to the considerable advances in generic body gesture recognition and analysis for adults [2,11,47,48, 62,64,65,76], children's whole-body gesture performance is, unfortunately, still little understood today. Besides limited research, the lack of publicly-available datasets on gestures produced by children has also prevented new discoveries. A very recent (2018) initiative of Aishat *et al.* [3] was to release a Kinect dataset with 58 motions performed by 10 children (ages 5 to 9 years) and 10 adults. As a side contribution of this work, we also align to this initiative by releasing our dataset of gestures performed by 30 children, ages 3 to 6 years. In this context, our empirical results on the consensus between children's gesture articulations and our dataset come at a right time to stimulate more research in this direction.

3 THE DISSIMILARITY-CONSENSUS APPROACH TO ANALYZING GESTURE ELICITATION DATA

We present in this section our new method to compute the consensus between end-users' gestures. Our approach introduces a shift from gesture data analysis based on manual labeling and subjective clustering criteria, currently in-use in the community, to a holistic modeling of the numerical relationship between consensus and gesture dissimilarity.

Automated Computation of Consensus using Gesture Dissimilarity Functions

We start the presentation of our method by introducing a formula to compute consensus for a set of gestures by using a gesture dissimilarity function. Let g_i represent the gesture elicited from the i-th participant in response to some referent R. Let Δ denote a dissimilarity function that computes a real, positive number to characterize how dissimilar two gestures are. For example, Δ may be the Dynamic Time Warping (DTW) cost function, a popular and accurate technique for classifying time-series data [15,50], including gesture data of all kinds, e.g., stylus gestures [74], finger touch strokes [34,67], motion gestures [35,51,63], mid-air freehand gestures [6], and whole-body movement [14,58,59,64,65]. Let τ represent our tolerance for deciding when two gestures are "similar," i.e., gestures g_i and g_j are considered similar if $\Delta(g_i, g_j) \leq \tau$. With these definitions, we compute the consensus between N gestures $g_{i(=1..N)}$ elicited for referent R from N users as follows:

Definition: Consensus for referent R is the percent of all pairs of gestures that are evaluated to be similar,

$$C_R(\tau) = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} \left[\Delta(g_i, g_j) \leq \tau\right]}{\frac{1}{2}N(N-1)} [\cdot 100\%] \qquad (1)$$

where N is the number of participants from which gestures are elicited, and the expression in square brackets evaluates to either 1 or 0, depending whether it is true or false. $C_R(\tau)$ takes values in [0..100], where 0% denotes no consensus and 100% perfect consensus. For example, consider that $N = 4$ participants are elicited for their gesture preferences with respect to some referent and that the DTW dissimilarity computed on all the $(4 \times 3)/2 = 6$ pairs of participants gives the results shown in Table 1. If we choose $\tau = 1.00$, then consensus is $C_R = (1 + 0 + 1 + 0 + 1 + 0)/6 \cdot 100\% = 50\%$.

	P$_1$	**P$_2$**	**P$_3$**	**P$_4$**
P$_1$	0	0.25	1.85	0.93
P$_2$	0.25	0	2.13	0.78
P$_3$	1.85	2.13	0	1.10
P$_4$	0.93	0.78	1.10	0

Table 1. Mock-up example to illustrate computation of consensus. Using Eq. 1 and $\tau = 1.00$, consensus is 50%.

Computation of Consensus for Repeated Elicitation

Eq. 1 covers the case where only one gesture is elicited from each participant for referent R, which represents the original implementation of the gesture elicitation methodology [72]. However, more complex experimental designs may elicit more than one gesture per participant, a procedure known as "production" to force participants move beyond legacy-biased gestures [42]. In the following, we extend Eq. 1 to address such scenarios involving multiple, distinct gestures collected from the same participant for the same referent.

Let $g_{i,t}$ represent the t-th gesture proposal collected from the i-th participant for R. The extended formula becomes:

Definition: Consensus for referent R, under repeated elicitation, is the percent of all pairs of gestures, including their repetitions, that are evaluated to be similar,

$$C_R^{\star}(\tau) = \frac{\sum_{i=1}^{N} \sum_{j=i+1}^{N} \left[\zeta\left(\Delta(g_{i,t}, g_{j,u}) \,\forall t, u\right) \leq \tau\right]}{\frac{1}{2}N(N-1)} [\cdot 100\%] \qquad (2)$$

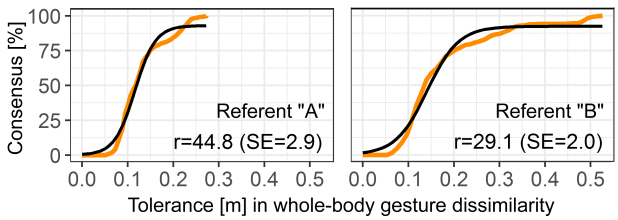

Figure 2. Growth curves quantify the relationship between consensus and the tolerance in gesture dissimilarity, τ, below which two gestures are considered similar. In this example, the growth rate of consensus for referent "A" is larger than for "B," showing faster reaching consensus for "A." *Note:* actual consensus in orange, logistic model fit in black.

where t and u index gestures proposed by participants i and j for referent R, and ζ is a new function that takes as input all the dissimilarity values computed for all the $t \times u$ combinations of $g_{i,t}$ and $g_{j,u}$ and returns a single, aggregated value. In this work, we implement and evaluate ζ with the `min`, `max`, and `avg` functions, corresponding to optimistic, pessimistic, and realistic computation of consensus. For example, consider that the four participants from Table 1 were elicited two more times. In this case, we would have a set of 12 gestures with three gestures from each participant. To compare the gestures of, say, participants P_1 and P_2, we need $3 \times 3 = 9$ evaluations of the dissimilarity measure Δ in Eq. 2. Suppose that the resulting dissimilarity values are, in ascending order, 0.23, 0.29, 0.35, 0.36, 0.51, 0.72, 0.89, 1.10, and 1.51. Depending on our choice of ζ, the aggregated dissimilarity between gestures elicited from P_1 and P_2 may be 0.23 for ζ=`min`, 1.51 for ζ=`max`, 0.66 for ζ=`avg`, and so on. After the aggregation, computation of consensus for a given τ proceeds similarly as in the previous example; see Table 1.

Dissimilarity-Consensus Growth Curves

The value chosen for the tolerance parameter, τ, can cause consensus to take any value from 0% to 100%, e.g., choosing a small, conservative value when judging how similar two gestures are will lead to a smaller magnitude of consensus than when employing a more permissive τ. Thus, working with a single τ value would be equivalent to clustering the elicited gestures according to a specific set of criteria, just like in the example from the introduction of this paper.

The alternative approach, which we propose in the following, is to adopt a holistic perspective, according to which the relationship between consensus and τ is modeled in the form of a growth curve. Figure 2 illustrates two examples of τ-C growth curves for two referents from our dataset (presented in the next section). When τ values are small, consensus is small as well, and it is difficult to differentiate the two referents by the magnitudes of their consensus. For τ less than 0.1 m, it appears that referent "B" has a slightly larger consensus than referent "A," but the difference is small and it may be that we are too conservative in our dissimilarity criterion to detect a true difference in consensus. For τ values larger than 0.4 m, both referents reach very high consensus, i.e., 100% for "A" and 94% for "B," but this time it may be that we are too permissive in our criteria (τ) to really understand differences in consensus. The big picture is provided only when we look at consensus overall as a function of τ, instead of taking snapshots at specific points.

To characterize the growth of consensus numerically, we need a model of the τ-C relationship. Because consensus cannot grow indefinitely as it is upper bounded by 100%, logistic growth, a technique commonly employed to model growth for populations that increase towards a maximum limit [61], seems appropriate to model the dissimilarity-consensus relationship. We thus employ the standard form of the logistic model which, expressed using our notations, is:

$$C_R(\tau) = \frac{C_\infty \cdot C_0}{C_0 + (C_\infty - C_0) \cdot \exp(-r \cdot \tau)} \tag{3}$$

where $C_\infty = \lim_{\tau\to\infty} C(\tau)$ and $C_0 = \lim_{\tau\to 0} C(\tau)$ are the upper and lower bounds of consensus and r is the growth rate. For the experiments reported in this work, we fit growth curves using the `R` library `growthcurver` by Sprouffske [57]. For a good fit, we want C_0 to be close to zero, C_∞ close to 100, and r to show a statistically significant fit at $\alpha = .05$. The growth rate r is our measure to characterize the overall numerical relationship between consensus and gesture dissimilarity.

Creation of Consensus Gesture Sets

According to Wobbrock *et al.* [72], the consensus set of recommendable gestures is formed by *"taking the largest groups of identical gestures for each referent and assigning those groups' gestures to the referent"* (p. 1087). The τ-C technique basically works by running an iterative clustering of the elicited gestures for multiple, continuously increasing values of the tolerance threshold τ. At any τ, the result is a binary matrix encoding similarity relationships between any two gestures, according to the evaluation of the expression $[\Delta(g_i, g_j) \le \tau]$. From this matrix, a set of optimally separable clusters and, specifically, the largest cluster of similar gestures can be determined automatically using techniques such as hill climbing or correlation clustering, among others, as demonstrated by the recent Crowdsensus system [1]. If the practitioner cannot decide on a single τ, the range values of τ can be sampled, e.g., in 10 or 50 points, and the corresponding binary matrices added together, providing thus an overall perspective of similarity relationships across all τ's, just like the logistic curve describes the relationship between dissimilarity and consensus. The same clustering techniques [1] are then applied directly to the sum matrix to identify the largest cluster and, correspondingly, the "winning" gestures. Also, as consensus is determined for each referent independently, the same type of outcomes as in previous work on gesture elicitation, e.g., selecting the same gesture for two distinct referents, are still possible with our technique.

Domain of Application

In this work, we demonstrate the τ-C method on whole-body gestures elicited from small children; see the next section. However, our method is general and applicable to any type of gestures, such as touch, multitouch, mid-air, free-hand, whole-body, or any combinations of these, as long as a numerical representation of those gestures is available and a dissimilarity function Δ can be defined to be used in Eqs. 1 and 2. For example, gestures are commonly represented as time series, where instantaneous measurements about the progression of a gesture in time are reported by some sensor. These measurements can be 2-D points for touch input [74], 3-D accelerations for motion gestures [63], hand poses for free-hand gestures [75], or poses of the whole body [64], as exemplified later in this paper. For time series, the Dynamic Time Warping function, to name just one example, has been proven to work remarkably well in practice [50,58].

4 EXPERIMENT

To demonstrate our method, we conducted a gesture elicitation experiment [72] to collect whole-body gestures.

Participants

Thirty (30) children, ages 3 to 6 years (M = 4.4, SD = 0.9), participated in our study. Half were boys, and the age distributions were similar for the two gender groups (M=4.4 years for boys and 4.5 years for girls, respectively). Parents' consent was obtained before the study. Children were divided into three age groups of equal size: (i) younger than 4 years, (ii) between 4 and 5 years old, and (iii) older than 5 years.

Apparatus

Children's whole-body gestures were captured with a Microsoft Kinect sensor $v1.8$ [39] that was connected to a 2.1 GHz Dual-Core PC running Windows 7 and our custom software application. All gestures were stored as skeleton data in XML format with 20 joints per body pose.

Task

Children stood at about 3 m in front of the Kinect sensor inside a circle with a diameter of 2 m delimited on the ground with white tape to prevent them from exiting the active sensing area; see Figure 1. Children were asked to produce body gestures in response to short instructions provided by our software with audio recordings, e.g., *"show how you throw a ball!"* or *"draw a flower in mid-air!"* There was no visual feedback in order not to influence children's body movements in any way. Instructions were played by a speaker in the form of a bear toy, representing an implementation of gamification [9], to keep children motivated during the study. Children were told to move as they wished in response to the instructions received from the bear. After the child confirmed that the instructions were understood, they produced the gesture, which was recorded by our application.

No.	Referent	Instructions received†
1	Throw ball	*Show how you throw a ball!*
2	Climb ladder	*Show how you climb a ladder!*
3	Slice carrots	*Show how you slice carrots!*
4	Angry bear	*Show how an angry bear looks like!*
5	Bird flying	*Show how a bird flies!*
6	Cat scratching	*Show how a cat scratches!*
7	Circle	*Draw a circle in mid-air!*
8	Square	*Draw a square in mid-air!*
9	Flower	*Draw a flower in mid-air!*
10	Applaud	*Applaud as hard as you can!*
11	Hands up	*Raise your hands up in the air!*
12	Stand on one foot	*Stand on one foot!*
13	Jump	*Jump as high as you can!*
14	Crouch	*Crouch!*
15	Turn around	*Turn around!*

†Provided to children in the form of audio recordings.

Table 2. The set of referents used to elicit body gestures.

We selected 15 referents to reveal children's metaphors of thought for various iconic movements, e.g., manipulation of objects, mimicking animal behavior, and drawing shapes in mid-air; see Table 2. Some of the gesture types were inspired by previous work, e.g., "throw a ball" [49], "jump" [3,25], "crouch" [25], "fly like a bird" and "climb" [3], while the others were newly designed to complement the gesture types from the literature. In contrast to other studies [3], we elicited children for the same referent multiple times to collect more gesture variation. Each referent was presented for three times, resulting in $15 \times 3 = 45$ trials. The order of the trials was randomized per participant. On average, data collection took about 7 minutes per child.

Completion Rate

The total number of expected gestures was 30 (children) $\times$ 15 (referents) $\times$ 3 (repetitions) = 1,350. The actual number of collected gestures was 1,312, corresponding to a completion rate of 97.2%. Missing data were caused by children not knowing how to move in response to some of the referents. Except "angry bear" (71.1%), all the other referents scored completion rates over 95.5%, very high given the small age of our children (3–6 years) and the low completion rates often reported in the literature for studies with children, e.g., 81.8% for touch input [66] or between 73% and 97% for stroke gesture input on smartphones [9]. Implementing gamification (the bear toy) and keeping the data collection procedure short (7 minutes on average) definitely helped.

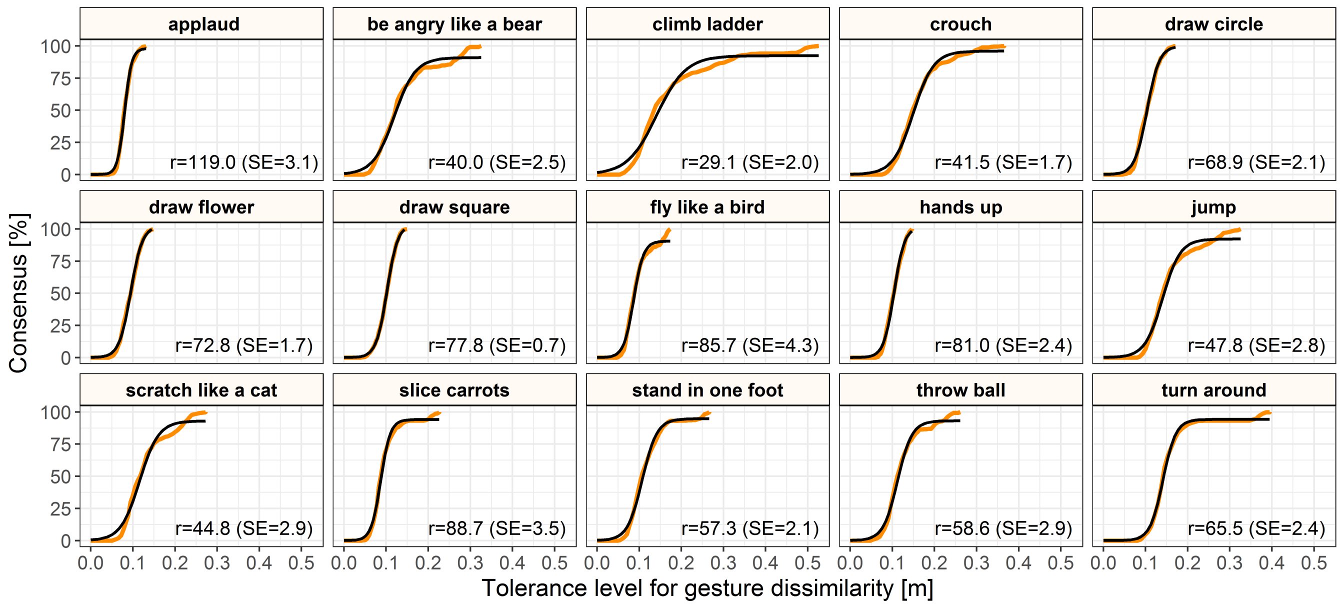

Figure 3. Growth curves (actual data in orange, logistic models in black) illustrating the dissimilarity-consensus relationship: the larger the tolerance τ allowed for judging two whole-body gestures as similar (x axis), the higher the consensus (y axis). ***Notes:*** **growth rates, r, and their standard errors are shown for each referent: the larger the growth rate, the faster consensus is reached for the same unit of τ; all growth rates showed a statistically significant fit at $p < .001$.**

5 RESULTS #1: VALIDATION OF THE LOGISTIC MODEL FOR CONSENSUS GROWTH CURVES

We start the presentation of our empirical results with an evaluation of the goodness of fit of the logistic function for modeling τ-C growth curves. At the same time, we also want to understand the effect of the gesture dissimilarity measure (Δ in Eqs. 1 and 2) and the aggregator function (ζ in Eq. 2) on consensus. To this end, we designed the validation procedure as a 3-way $15 \times 4 \times 3$ mixed design with the following independent variables:

(1) Referent, nominal, 15 conditions; see Table 2.
(2) Dissimilarity Δ, nominal, 4 conditions: DTW, Euclidean, Hausdorff, and modified Hausdorff, described next.
(3) Aggregator ζ, nominal, 3 conditions: `min`, `max`, and `avg`.

Dissimilarity Measures for Whole-Body Gestures

As mentioned earlier, one choice for the dissimilarity measure Δ is DTW due to its high accuracy and popularity in the gesture literature, including for whole-body gestures [14,58, 63–65], while the `avg` function seems a reasonable implementation of the aggregator function ζ for repeated elicitation that is neither pessimistic nor optimistic. These choices will be validated by the results of this section, as we compare multiple Δ and ζ conditions. Next, we provide a definition and a brief motivation for each dissimilarity measure that we employ in this work. For all following definitions, we consider a whole-body gesture g to be represented as a time series of body poses, $g = \{g_{[i]} \mid i = 1..n\}$, where n is the number of body poses and each pose is a set of 3-D points representing locations of body joints in space, $g_{[i]} = \left\{g^k_{[i]} \mid k = 1..20\right\}$. Note that we place the index i in square brackets to refer to a body pose and, thus, to differentiate from the notations regarding participants' indices used in Eqs. 1 and 2.

Dynamic Time Warping. DTW is a generic technique for matching time series data of all kinds that computes the optimum chronological alignment between two series using dynamic programming [15]. The optimum matching is computed using memoization, i.e., partial results are stored in a matrix Δ, so that cell $\Delta_{i,j}$ contains the optimum matching of the first i data points of the first series to the first j data points of the second. The final result is found in the bottom-right cell of the matrix, $\Delta_{n,m}$, where n and m are the lengths of the two series. In this work, we employ the normalized DTW measure by dividing the matching result $\Delta_{n,m}$ to the number of alignments performed during the matching process, $l_{n,m}$:

$$\Delta_{\mathrm{DTW}}(g, h) = \Delta_{n,m} / l_{n,m} \tag{4}$$

where the matrix $\Delta_{\cdot,\cdot}$ is defined recursively as follows:

$$\Delta_{i,j} = \begin{cases} \delta(g_{[1]}, h_{[1]}), \text{ if } i = 1 \land j = 1 \\ \Delta_{1,j-1} + \delta(g_{[1]}, h_{[j]}), \text{ if } i = 1 \land j > 1 \\ \Delta_{i-1,1} + \delta(g_{[i]}, h_{[1]}), \text{ if } i > 1 \land j = 1 \\ \min\left\{\Delta_{i-1,j-1}, \Delta_{i-1,j}, \Delta_{i,j-1}\right\} + \delta(g_{[i]}, h_{[j]}), \text{ otherwise} \end{cases}$$

and $\delta(g_{[i]}, h_{[j]}) = \frac{1}{20}\sum_{k=1}^{20} |g_{[i]}^k - h_{[j]}^k|$ represents the Euclidean distance between the i-th body pose of gesture g and the j-th body pose of gesture h, respectively. We divide the sum by 20 to make δ invariant to the number of points tracked on the human body by the Kinect $v1.8$ sensor.

Euclidean distance. The Euclidean distance between two gestures g and h computes the sum of the Euclidean distances between their corresponding body poses, $g_{[i]}$ and $h_{[i]}$, under the assumption of the same number n of body poses for both g and h. (An assumption not met in practice, but achieved with a resampling procedure; see next.)

$$\Delta_E(g, h) = \frac{1}{n}\sum_{i=1}^{n} \delta(g_{[i]}, h_{[i]}) \qquad (5)$$

where δ was introduced before. The Euclidean distance does not posses the matching flexibility of DTW, but is straightforward to implement and was found to work well with all kinds of gesture types represented as time series [63,65,74].

Hausdorff distance. A common and accurate technique employed in the Computer Vision community to match shapes, point sets, sketches, stroke-gestures, volumes, and images is to compute the maximum of the minimum distances between pairs of data points of the two sequences being matched, known as the "Hausdorff distance" [27]. This procedure can be easily extended to whole-body gestures as follows:

$$\Delta_H(g, h) = \max\{\text{Hausdorff}(g, h), \text{Hausdorff}(h, g)\} \qquad (6)$$

$$\text{Hausdorff}(g, h) = \max_{i=1,n}\left\{\min_{j=1,m} \delta(g_{[i]}, h_{[j]})\right\}$$

where δ is the Euclidean distance, and n and m represent the number of body poses of gestures g and h, respectively.

Modified Hausdorff distance. A variant of the Hausdorff distance, delivering more accurate results in practice [18], considers the average instead of the maximum aggregator:

$$\Delta_{MH}(g, h) = \max\{\text{m-Hausdorff}(g, h), \text{m-Hausdorff}(h, g)\} \qquad (7)$$

$$\text{m-Hausdorff}(g, h) = \frac{1}{n}\sum_{i=1}^{n} \min_{j=1,m} \delta(g_{[i]}, h_{[j]})$$

Gesture Preprocessing

Before computing the dissimilarity measures, we normalized all the gestures using the following steps:

(1) *Resampling of body poses.* This operation makes the dissimilarity computations independent of the sampling resolution of the sensor. We resampled all the gestures at 25 fps. For example, a gesture that took 3.32 seconds to produce was resampled into 83 body poses, uniformly spaced at $3.32/(83-1) = 0.04$ seconds apart (i.e., 25 fps).

(2) *Height normalization.* This operation makes the dissimilarity values independent of children's body sizes. We rescaled all the gestures so that to the body height of each child, standing up straight, was 1.0 m.

(3) *Translation to origin.* This operation makes the dissimilarity values independent of where the gesture is produced in space. For each gesture, we subtracted its centroid from each body joint, so that the new centroid was (0, 0, 0).

Δ	ζ	C_0 †	C_∞ ‡	**Growth rate r**
Δ_{DTW}	`min`	0.18	96.55	82.92 ($p < .001$)
	`max`	0.37	95.07	47.26 ($p < .001$)
	`avg`	0.25	95.77	65.24 ($p < .001$)
Δ_E	`min`	0.13	96.62	76.33 ($p < .001$)
	`max`	0.33	95.33	45.67 ($p < .001$)
	`avg`	0.20	95.97	61.77 ($p < .001$)
Δ_H	`min`	0.29	96.00	61.36 ($p < .001$)
	`max`	0.69	92.96	32.42 ($p < .001$)
	`avg`	0.60	94.93	43.80 ($p < .001$)
Δ_{MH}	`min`	0.18	95.89	94.89 ($p < .001$)
	`max`	0.37	94.43	55.39 ($p < .001$)
	`avg`	0.26	95.34	74.99 ($p < .001$)
Average, all $\Delta \times \zeta$		**0.32**	**95.41**	·

† Values of C_0 closer to 0 show a better fit.
‡ Values of C_∞ closer to 100 show a better fit.

Table 3. Numerical indicators of the goodness of fit of the logistic growth model for all the $\Delta \times \zeta$ conditions.

Goodness of Fit of the Logistic Model

We computed the τ-C growth curves for all the fifteen referents and the twelve $\Delta \times \zeta$ combinations. Figure 3 illustrates the data for the Δ_{DTW} dissimilarity and the `avg` ζ aggregator: raw data is shown in orange and the fitted logistic model in black. (For space concerns, we skip the illustration of the other $\Delta \times \zeta$ combinations, but numerical goodness of fit results are shown in Table 3 for all $\Delta \times \zeta$, while all the growth curves are accessible from the companion web page of this paper, where we make our gesture dataset publicly available.) Figure 3 shows visually that the logistic model provides a good fit for all the referents. Table 3 confirms this intuition with the numerical results of the fit. The estimated values for C_0 (see Eq. 3) are close to zero ($M = 0.32$, $SD = 0.17$) and the values of C_∞ approach 100 ($M = 95.41$, $SD = 1.00$), both signs of a good fit. Moreover, all the growth rates r are significant at $p<.001$, which confirms the suitability of the logistic function to model the τ-C relationship.

The Effect of Gesture Dissimilarity and Aggregator

Next, we analyze the effect of the dissimilarity measure Δ and aggregator function ζ on growth rates r. Specifically, we want to know whether the magnitude of growth rates r and the ranking of gestures by overall consensus, as reflected by growth rates, are impacted by the choice of Δ and/or ζ.

A preliminary RM ANOVA indicated a significant effect of Δ on the magnitude of growth rates r ($F_{(2.078,29.090)} = 98.260$,

	ζ = min				ζ = max				ζ = avg			
	Δ_{DTW}	Δ_E	Δ_H	Δ_{MH}	Δ_{DTW}	Δ_E	Δ_H	Δ_{MH}	Δ_{DTW}	Δ_E	Δ_H	Δ_{MH}
Δ_{DTW}	1.000	.974	.972	.971	1.000	.977	.915	.935	1.000	.988	.959	.971
Δ_E	·	1.000	.953	.956	·	1.000	.923	.877	·	1.000	.940	.958
Δ_H	·	·	1.000	.971	·	·	1.000	.836	·	·	1.000	.940
Δ_{MH}	·	·	·	1.000	·	·	·	1.000	·	·	·	1.000
Average	**Pearson's $r_{(N=15)}$ = .966**				**Pearson's $r_{(N=15)}$ = .911**				**Pearson's $r_{(N=15)}$ = .959**			

Table 4. Pearson coefficients between growth rates for each $\Delta \times \zeta$ combination. *Note:* all correlations are significant at $p < .001$.

$p < .001$, $\eta_p^2 = .875$, Greenhouse-Geisser's $\hat{\epsilon} = .693$) as well as a significant effect of aggregator function ζ ($F_{(1.157,16.202)} = 68.509$, $p < .001$, $\eta_p^2 = .830$, $\hat{\epsilon} = .579$). However, this result is little informative, as differences *are* expected to be present in the magnitude of growth rates simply because of the different magnitudes delivered by different dissimilarity functions on the same data; see Eqs. 4, 5, 6, and 7. What we are actually interested in is whether the ratio of consensus (e.g., as in the expression "consensus for referent A growths twice faster than consensus for B") is preserved across Δ's and ζ's. To this end, we normalized growth rates r for each $\Delta \times \zeta$ condition into $[0..1]$ by applying a linear scale transform, i.e., $r = (r - \min(r))/(\max(r) - \min(r))$. This time, the same ANOVA test found no significant effects of either Δ ($F_{(3,42)}$=.891, p>.05, *n.s.*) or ζ ($F_{(1.144,16.020)}$=.909, p>.05, *n.s.*, $\hat{\epsilon} = .572$) on mean growth rates r, which builds up confidence that the relative consensus is preserved across Δ and ζ functions.

To further understand the impact of $\Delta \times \zeta$ on ranking referents by consensus, we also performed Pearson correlations between the growth rates computed for the $N = 15$ referents. Results are shown in Table 4. Pearson coefficients were very large: .966 on average for ζ=min, .911 for ζ=max, and .959 for ζ=avg, showing that the ranking of referents by consensus, as reflected by their growth rates r, is little affected by the choice of Δ and ζ. The largest correlation coefficients were obtained for ζ=min (the optimistic aggregator) and ζ=avg (the realistic aggregator). Based on this empirical evidence, we recommend Δ_{DTW} and ζ=avg for use in practice.

Manual Gesture Clustering *vs.* Automated Computation of Consensus using Growth Rates

As a final test, we computed Pearson correlation coefficients between the magnitudes of growth rates r obtained with Δ_{DTW} and the avg aggregator and agreement rates AR calculated using the formula of Vatavu and Wobbrock [68] after manual labeling and clustering of the elicited gestures using the criteria mentioned in the example from the Introduction. Because there is no extended formula of AR [68] for repeated elicitation, we computed agreement rates separately for each trial: AR_1, AR_2, and AR_3, respectively. Results showed statistically significant correlations between growth rates r and AR_1 (Pearson's $r_{(N=15)}$=.644, p<.01), AR_2 ($r_{(N=15)}$=.647, p<.01), and AR_3 ($r_{(N=15)}$=.650, p<.01), respectively.

6 RESULTS #2: CONSENSUS BETWEEN CHILDREN'S WHOLE-BODY GESTURES

In this section, we demonstrate the τ-C method by applying it to report the level of consensus between whole-body gestures produced by small children, ages 3 to 6 years. Overall, we report empirical results on 1,312 gestures, consisting in a total number of 48,299 body poses, elicited from N=30 children. Note that it is not our intention to be exhaustive in the analysis that we report in this section, but rather to demonstrate how the τ-C method can be applied in practice. Nevertheless, we do report, for the first time in the literature, empirical results on the way children produce whole-body gestures, such as an effect of the age group on the magnitude of consensus between their gesture articulations; see next.

Consensus Results

We computed τ-C growth curves for each referent from Table 2 using the Δ_{DTW} dissimilarity, the avg aggregator, and the consensus formula from Eq. 2; see Figure 3. We found that consensus growth rates varied between $r = 29.1$ (for the "climb ladder" referent) to $r = 119.0$ ("applaud") with a mean of 65.2 (SD = 23.4); see Figure 4a for all the referents listed in decreasing order of their consensus growth rates. The ratio between the largest and smallest growth rates was 4.0, i.e., consensus for "applaud" grew four times faster than for "climb ladder" for the same unit increase of τ. Referents that involved moving around (e.g., "jump", "crouch") or movement of several body parts at once (e.g., "stand on one foot," for which children used movements of both hands to keep stable equilibrium) achieved less consensus, revealed by smaller growth rates, than referents that involved stable body poses and movement of fewer body parts (e.g., "applaud", "slice carrots", or "hands up"). Also, we found that referents that necessitated instantiation of a metaphor into a motor response, such as "angry bear", "cat scratching" or "climb ladder," received less consensus compared to simple shapes drawn in mid-air, such as "flower", "square", and "circle."

The Effect of Age Group

Our analysis showed that consensus was reached 24.5% faster by 4-year-olds and 44.8% times faster by 5-year-olds compared to children aged between 3 and 4 years; see Figure 4b.

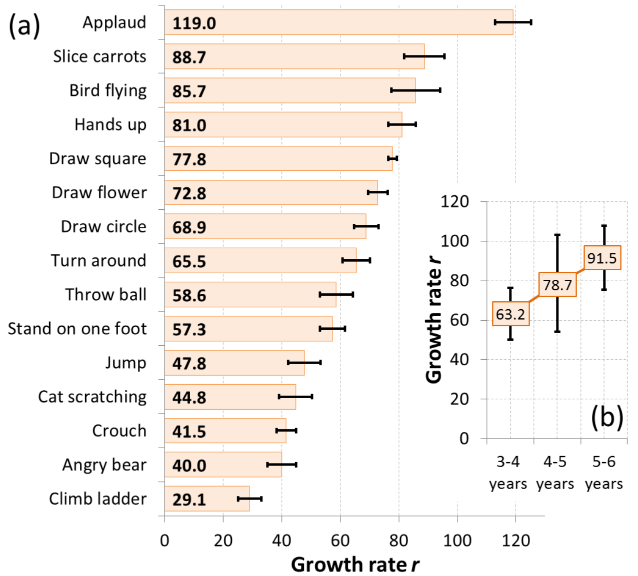

Figure 4. Consensus between whole-body gestures elicited from 30 children, expressed with growth rates: (a) per referent and (b) per age group. *Note:* error bars show 95% CIs.

A Repeated-Measures ANOVA revealed a significant effect of Age-Group on growth rates ($F_{(1.451,20.318)} = 5.034$, $p = .025$, $\eta_p^2 = .264$, Greenhouse-Geisser's $\hat{\epsilon} = .726$). Post-hoc tests (Bonferroni corrected) revealed significant differences between children with ages between 3 and 4 years and the 5-6-years-old group ($p < .001$). These results find support in the literature of child development and especially in the developmental theory of Piaget [44]: as children grow up, they acquire and refine their motor skills, integrate more metaphors of thought, and move beyond an egocentric perspective affecting their cognitive representations of the physical world [45].

Towards Further Discoveries

As mentioned before, it is not the goal of this paper to continue with detailed examinations of how small children perform whole-body gestures, although an exciting investigation. Instead, we provide our dataset in the community for free to foster such new discoveries; see the next section for details. Future investigations may include numerical analysis of children's whole-body gesture articulations, such as by employing the geometric, kinematic, and body-appearance set of measures and toolkit of Vatavu [64], or investigating correlations between consensus between proposed gestures and numerical measures of gesture articulation, such as gesture volume, quantity of movement, or body pose variation [64]. Designing robust gesture recognition techniques for whole-body gestures produced by children is another challenging future work direction, for which our large dataset can provide support for user-independent evaluation procedures.

7 GESTURE DATASET AND SOURCE CODE

We release our dataset composed of 1,312 whole-body gestures and 48,299 body poses elicited from 30 children free to download and use for research purposes. To our knowledge, our dataset is the only whole-body gesture data publicly available for children this young (3 to 6 years old). We also release source code in C# that reads the gesture data and implements the dissimilarity measures Δ and aggregator functions ζ employed in this work and R code to compute dissimilarity-consensus growth rates and visualize τ-C curves. All the resources are available from the companion web page of this paper at http://www.eed.usv.ro/~vatavu

8 CONCLUSION AND FUTURE WORK

We introduced and evaluated in this paper a new approach to computing and understanding consensus in end-user gesture elicitation studies by adopting a holistic perspective on the numerical relationship between consensus and gesture dissimilarity. We hope that our theoretical contribution, empirical results, and practical source code will benefit designers and practitioners in the need of an objective assessment of the magnitude of consensus for their gesture studies.

At the same time, we acknowledge the need for more work that is needed to incorporate into our method the interpretative dimensions of sociocultural impact, such as cultural gestures with the same meaning, yet structurally different appearance [4] or approaches based on somaesthetics [4,19,37], that cannot be addressed at this moment by our automated technique, which is blind to such aspects. We also suggest further investigations of the middle ground between manual and automated approaches, which we believe will foster valuable methodological developments. Another direction for future work is applying the τ-C method for other types of gestures. Although we demonstrated τ-C for whole-body movements, our technique is general and applicable to any type of gestures, such as touch [20,72], mid-air [46,75], or ring gestures [21], to name just a few. Also, we have barely scratched the potential for understanding how small children produce whole-body gestures. But we hope that our large whole-body gesture dataset will foster new discoveries, advancing our present understanding regarding how small children perform whole-body movements.

ACKNOWLEDGMENTS

This work was supported by a grant of the Ministry of Research and Innovation, CNCS-UEFISCDI, project no. PN-III-P1-1.1-TE-2016-2173 (TE141/2018), within PNCDI III. We thank G. Cramariuc for help with data collection and the anonymous reviewers for very useful suggestions.

REFERENCES

[1] Abdullah X. Ali, Meredith Ringel Morris, and Jacob O. Wobbrock. 2018. Crowdsourcing Similarity Judgments for Agreement Analysis in End-User Elicitation Studies. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. ACM, New York, NY, USA, 177–188. https://doi.org/10.1145/3242587.3242621

[2] Saad Ali and Mubarak Shah. 2010. Human Action Recognition in Videos Using Kinematic Features and Multiple Instance Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 2 (Feb. 2010), 288–303. https://doi.org/10.1109/TPAMI.2008.284

[3] Aishat Aloba, Gianne Flores, Julia Woodward, Alex Shaw, Amanda Castonguay, Isabella Cuba, Yuzhu Dong, Eakta Jain, and Lisa Anthony. 2018. Kinder-Gator: The UF Kinect Database of Child and Adult Motion. In *EG 2018 - Short Papers*, Olga Diamanti and Amir Vaxman (Eds.). The Eurographics Association. https://doi.org/10.2312/egs.20181033

[4] Dane Archer. 1997. Unspoken Diversity: Cultural Differences in Gestures. *Qualitative Sociology* 20, 1 (March 1997), 79–105. https://doi.org/10.1023/A:1024716331692

[5] Shaikh Shawon Arefin Shimon, Courtney Lutton, Zichun Xu, Sarah Morrison-Smith, Christina Boucher, and Jaime Ruiz. 2016. Exploring Non-touchscreen Gestures for Smartwatches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3822–3833. https://doi.org/10.1145/2858036.2858385

[6] Ilhan Aslan, Andreas Uhl, Alexander Meschtscherjakov, and Manfred Tscheligi. 2014. Mid-air Authentication Gestures: An Exploration of Authentication Based on Palm and Finger Motions. In *Proceedings of the 16th International Conference on Multimodal Interaction (ICMI '14)*. ACM, New York, NY, USA, 311–318. https://doi.org/10.1145/2663204.2663246

[7] Gilles Bailly, Thomas Pietrzak, Jonathan Deber, and Daniel J. Wigdor. 2013. Métamorphe: Augmenting Hotkey Usage with Actuated Keys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 563–572. https://doi.org/10.1145/2470654.2470734

[8] Idil Bostan, Oğuz Turan Buruk, Mert Canat, Mustafa Ozan Tezcan, Celalettin Yurdakul, Tilbe Göksun, and Oğuzhan Özcan. 2017. Hands As a Controller: User Preferences for Hand Specific On-Skin Gestures. In *Proceedings of the 2017 Conference on Designing Interactive Systems (DIS '17)*. ACM, New York, NY, USA, 1123–1134. https://doi.org/10.1145/3064663.3064766

[9] Robin Brewer, Lisa Anthony, Quincy Brown, Germaine Irwin, Jaye Nias, and Berthel Tate. 2013. Using Gamification to Motivate Children to Complete Empirical Studies in Lab Environments. In *Proceedings of the 12th International Conference on Interaction Design and Children (IDC '13)*. ACM, New York, NY, USA, 388–391. https://doi.org/10.1145/2485760.2485816

[10] Francesco Cafaro, Leilah Lyons, and Alissa N. Antle. 2018. Framed Guessability: Improving the Discoverability of Gestures and Body Movements for Full-Body Interaction. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 593, 12 pages. https://doi.org/10.1145/3173574.3174167

[11] Ziyun Cai, Jungong Han, Li Liu, and Ling Shao. 2017. RGB-D datasets using Microsoft Kinect or similar sensors: A survey. *Multimedia Tools and Applications* 76, 3 (Feb 2017), 4313–4355. https://doi.org/10.1007/s11042-016-3374-6

[12] Jessica R. Cauchard, Jane L. E, Kevin Y. Zhai, and James A. Landay. 2015. Drone & Me: An Exploration into Natural Human-Drone Interaction. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 361–365. https://doi.org/10.1145/2750858.2805823

[13] Sabrina Connell, Pei-Yi Kuo, Liu Liu, and Anne Marie Piper. 2013. A Wizard-of-Oz Elicitation Study Examining Child-defined Gestures with a Whole-body Interface. In *Proceedings of the 12th International Conference on Interaction Design and Children (IDC '13)*. ACM, New York, NY, USA, 277–280. https://doi.org/10.1145/2485760.2485823

[14] Suranjith De Silva, Michael Barlow, and Adam Easton. 2014. An Evaluation of DTW Approaches for Whole-of-Body Gesture Recognition. In *Proceedings of the 28th International BCS Human Computer Interaction Conference on HCI 2014 - Sand, Sea and Sky - Holiday HCI (BCS-HCI '14)*. BCS, UK, 11–21. https://doi.org/10.14236/ewic/hci2014.2

[15] Hui Ding, Goce Trajcevski, Peter Scheuermann, Xiaoyue Wang, and Eamonn Keogh. 2008. Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures. *Proceedings of the VLDB Endowment* 1, 2 (Aug. 2008), 1542–1552. https://doi.org/10.14778/1454159.1454226

[16] Tilman Dingler, Rufat Rzayev, Alireza Sahami Shirazi, and Niels Henze. 2018. Designing Consistent Gestures Across Device Types: Eliciting RSVP Controls for Phone, Watch, and Glasses. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 419, 12 pages. https://doi.org/10.1145/3173574.3173993

[17] Yuzhu Dong, Aishat Aloba, Sachin Paryani, Lisa Anthony, Neha Rana, and Eakta Jain. 2017. Adult2Child: Dynamic Scaling Laws to Create Child-like Motion. In *Proceedings of the 10th International Conference on Motion in Games (MIG '17)*. ACM, New York, NY, USA, Article 13, 10 pages. https://doi.org/10.1145/3136457.3136460

[18] Marie-Pierre Dubuisson and Anil K. Jain. 1994. A modified Hausdorff distance for object matching. In *Proceedings of 12th International Conference on Pattern Recognition*, Vol. 1. 566–568. https://doi.org/10.1109/ICPR.1994.576361

[19] Jane L. E, Ilene L. E, James A. Landay, and Jessica R. Cauchard. 2017. Drone & Wo: Cultural Influences on Human-Drone Interaction Techniques. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 6794–6799. https://doi.org/10.1145/3025453.3025755

[20] Leah Findlater, Ben Lee, and Jacob Wobbrock. 2012. Beyond QWERTY: Augmenting Touch Screen Keyboards with Multi-touch Gestures for Non-alphanumeric Input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. ACM, New York, NY, USA, 2679–2682. https://doi.org/10.1145/2207676.2208660

[21] Bogdan-Florin Gheran, Jean Vanderdonckt, and Radu-Daniel Vatavu. 2018. Gestures for Smart Rings: Empirical Results, Insights, and Design Implications. In *Proceedings of the 2018 Designing Interactive Systems Conference (DIS '18)*. ACM, New York, NY, USA, 623–635. https://doi.org/10.1145/3196709.3196741

[22] Jun Gong, Xing-Dong Yang, and Pourang Irani. 2016. WristWhirl: One-handed Continuous Smartwatch Input Using Wrist Gestures. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology (UIST '16)*. ACM, New York, NY, USA, 861–872. https://doi.org/10.1145/2984511.2984563

[23] Lynn Hoff, Eva Hornecker, and Sven Bertel. 2016. Modifying Gesture Elicitation: Do Kinaesthetic Priming and Increased Production Reduce Legacy Bias?. In *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '16)*. ACM, New York, NY, USA, 86–91. https://doi.org/10.1145/2839462.2839472

[24] Kristina Höök, Martin P. Jonsson, Anna Ståhl, and Johanna Mercurio. 2016. Somaesthetic Appreciation Design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3131–3142. https://doi.org/10.1145/2858036.2858583

[25] Johanna Höysniemi, Perttu Hämäläinen, and Laura Turkki. 2004. Wizard of Oz Prototyping of Computer Vision Based Action Games for

Children. In *Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community (IDC '04)*. ACM, New York, NY, USA, 27–34. https://doi.org/10.1145/1017833.1017837

[26] Johanna Höysniemi, Perttu Hämäläinen, Laura Turkki, and Teppo Rouvi. 2005. Children's Intuitive Gestures in Vision-based Action Games. *Commun. ACM* 48, 1 (Jan. 2005), 44–50. https://doi.org/10.1145/1039539.1039568

[27] Daniel P. Huttenlocher, Gregory A. Klanderman, and William J. Rucklidge. 1993. Comparing Images Using the Hausdorff Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 9 (Sept. 1993), 850–863. https://doi.org/10.1109/34.232073

[28] Shoya Ishimaru, Kai Kunze, Koichi Kise, Jens Weppner, Andreas Dengel, Paul Lukowicz, and Andreas Bulling. 2014. In the Blink of an Eye: Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass. In *Proceedings of the 5th Augmented Human International Conference (AH '14)*. ACM, New York, NY, USA, Article 15, 4 pages. https://doi.org/10.1145/2582051.2582066

[29] Eakta Jain, Lisa Anthony, Aishat Aloba, Amanda Castonguay, Isabella Cuba, Alex Shaw, and Julia Woodward. 2016. Is the Motion of a Child Perceivably Different from the Motion of an Adult? *ACM Transactions on Applied Perception* 13, 4, Article 22 (July 2016), 17 pages. https://doi.org/10.1145/2947616

[30] Sujin Jang, Wolfgang Stuerzlinger, Satyajit Ambike, and Karthik Ramani. 2017. Modeling Cumulative Arm Fatigue in Mid-Air Interaction Based on Perceived Exertion and Kinetics of Arm Motion. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 3328–3339. https://doi.org/10.1145/3025453.3025523

[31] Christian Kray, Daniel Nesbitt, John Dawson, and Michael Rohs. 2010. User-defined Gestures for Connecting Mobile Phones, Public Displays, and Tabletops. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '10)*. ACM, New York, NY, USA, 239–248. https://doi.org/10.1145/1851600.1851640

[32] Wonjun Lee, Youn-kyung Lim, and Richard Shusterman. 2014. Practicing Somaesthetics: Exploring Its Impact on Interactive Product Design Ideation. In *Proceedings of the 2014 Conference on Designing Interactive Systems (DIS '14)*. ACM, New York, NY, USA, 1055–1064. https://doi.org/10.1145/2598510.2598561

[33] Luis A. Leiva, Daniel Martín-Albo, Réjean Plamondon, and Radu-Daniel Vatavu. 2018. KeyTime: Super-Accurate Prediction of Stroke Gesture Production Times. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. ACM, New York, NY, USA, Article 239, 12 pages. https://doi.org/10.1145/3173574.3173813

[34] Luis A. Leiva, Daniel Martín-Albo, and Radu-Daniel Vatavu. 2017. Synthesizing Stroke Gestures Across User Populations: A Case for Users with Visual Impairments. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 4182–4193. https://doi.org/10.1145/3025453.3025906

[35] Jiayang Liu, Lin Zhong, Jehan Wickramasuriya, and Venu Vasudevan. 2009. uWave: Accelerometer-based Personalized Gesture Recognition and Its Applications. *Pervasive and Mobile Computing* 5, 6 (Dec. 2009), 657–675. https://doi.org/10.1016/j.pmcj.2009.07.007

[36] Yihua Lou, Wenjun Wu, Radu-Daniel Vatavu, and Wei-Tek Tsai. 2017. Personalized gesture interactions for cyber-physical smart-home environments. *Science China Information Sciences* (July 2017), 60:072104. https://doi.org/10.1007/s11432-015-1014-7

[37] Dan Mauney, Jonathan Howarth, Andrew Wirtanen, and Miranda Capra. 2010. Cultural Similarities and Differences in User-defined Gestures for Touchscreen User Interfaces. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems (CHI EA '10)*. ACM, New York, NY, USA, 4015–4020. https://doi.org/10.1145/1753846.1754095

[38] Keenan R. May, Thomas M. Gable, and Bruce N. Walker. 2017. Designing an In-Vehicle Air Gesture Set Using Elicitation Methods. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17)*. ACM, New York, NY, USA, 74–83. https://doi.org/10.1145/3122986.3123015

[39] Microsoft. 2018. Kinect for Windows. http://kinectforwindows.org/

[40] Jin Moen. 2005. Towards People Based Movement Interaction and Kinaesthetic Interaction Experiences. In *Proceedings of the 4th Decennial Conference on Critical Computing: Between Sense and Sensibility (CC '05)*. ACM, New York, NY, USA, 121–124. https://doi.org/10.1145/1094562.1094579

[41] Meredith R. Morris. 2012. Web on the Wall: Insights from a Multimodal Interaction Elicitation Study. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*. ACM, New York, NY, USA, 95–104. https://doi.org/10.1145/2396636.2396651

[42] Meredith R. Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. 2014. Reducing Legacy Bias in Gesture Elicitation Studies. *interactions* 21, 3 (May 2014), 40–45. https://doi.org/10.1145/2591689

[43] Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of Pre-designed and User-defined Gesture Sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1099–1108. https://doi.org/10.1145/2470654.2466142

[44] Jean Piaget. 2001. *The Psychology of Intelligence (2nd Ed.)*. London: Routledge.

[45] Jean Piaget and BÃďrbel Elisabeth Inhelder. 1969. *The Psychology of the Child*. New York: Basic Books.

[46] Thammathip Piumsomboon, Adrian Clark, Mark Billinghurst, and Andy Cockburn. 2013. User-defined Gestures for Augmented Reality. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13)*. ACM, New York, NY, USA, 955–960. https://doi.org/10.1145/2468356.2468527

[47] Ronald Poppe. 2007. Vision-based Human Motion Analysis: An Overview. *Computer Vision and Image Understanding* 108, 1-2 (Oct. 2007), 4–18. https://doi.org/10.1016/j.cviu.2006.10.016

[48] Ronald Poppe. 2010. A Survey on Vision-based Human Action Recognition. *Image and Vision Computing* 28, 6 (June 2010), 976–990. https://doi.org/10.1016/j.imavis.2009.11.014

[49] Mohd. Salihan Rahman, Nazlena Mohamad Ali, and Masnizah Mohd. 2013. A Study on the Naturalness of Gesture-Based Interaction for Children. In *Third International Visual Informatics Conference on Advances in Visual Informatics - Volume 8237 (IVIC'13)*. Springer-Verlag New York, Inc., New York, NY, USA, 718–728. https://doi.org/10.1007/978-3-319-02958-0_65

[50] Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. 2012. Searching and Mining Trillions of Time Series Subsequences Under Dynamic Time Warping. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12)*. ACM, New York, NY, USA, 262–270. https://doi.org/10.1145/2339530.2339576

[51] Jaime Ruiz and Yang Li. 2011. DoubleFlip: A Motion Gesture Delimiter for Mobile Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 2717–2720. https://doi.org/10.1145/1978942.1979341

[52] Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined Motion Gestures for Mobile Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 197–206. https://doi.org/10.1145/1978942.1978971

[53] Jaime Ruiz and Daniel Vogel. 2015. Soft-Constraints to Reduce Legacy and Performance Bias to Elicit Whole-body Gestures with Low Arm

Fatigue. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3347–3350. https://doi.org/10.1145/2702123.2702583

[54] Teddy Seyed, Chris Burns, Mario Costa Sousa, Frank Maurer, and Anthony Tang. 2012. Eliciting Usable Gestures for Multi-display Environments. In *Proceedings of the 2012 ACM International Conference on Interactive Tabletops and Surfaces (ITS '12)*. ACM, New York, NY, USA, 41–50. https://doi.org/10.1145/2396636.2396643

[55] Gözel Shakeri, John H. Williamson, and Stephen Brewster. 2017. Novel Multimodal Feedback Techniques for In-Car Mid-Air Gesture Interaction. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17)*. ACM, New York, NY, USA, 84–93. https://doi.org/10.1145/3122986.3123011

[56] Chaklam Silpasuwanchai and Xiangshi Ren. 2014. Jump and Shoot!: Prioritizing Primary and Alternative Body Gestures for Intense Gameplay. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 951–954. https://doi.org/10.1145/2556288.2557107

[57] Kathleen Sprouffske. 2018. *growthcurver: Simple Metrics to Summarize Growth Curves*. https://CRAN.R-project.org/package=growthcurver R package version 0.3.0.

[58] Eugene M. Taranta II, Amirreza Samiei, Mehran Maghoumi, Pooya Khaloo, Corey R. Pittman, and Joseph J. LaViola Jr. 2017. Jackknife: A Reliable Recognizer with Few Samples and Many Modalities. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM, New York, NY, USA, 5850–5861. https://doi.org/10.1145/3025453.3026002

[59] Eugene M. Taranta II, Thaddeus K. Simons, Rahul Sukthankar, and Joseph J. Laviola Jr. 2015. Exploring the Benefits of Context in 3D Gesture Recognition for Game-Based Virtual Environments. *ACM Transactions on Interactive Intelligent Systems* 5, 1, Article 1 (March 2015), 34 pages. https://doi.org/10.1145/2656345

[60] Giovanni Maria Troiano, Esben Warming Pedersen, and Kasper Hornbæk. 2014. User-defined Gestures for Elastic, Deformable Displays. In *Proceedings of the 2014 International Working Conference on Advanced Visual Interfaces (AVI '14)*. ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/2598153.2598184

[61] Anastasios Tsoularis and James Wallace. 2002. Analysis of logistic growth models. *Mathematical Biosciences* 179, 1 (2002), 21–55. https://doi.org/10.1016/S0025-5564(02)00096-2

[62] Radu-Daniel Vatavu. 2012. User-defined Gestures for Free-hand TV Control. In *Proceedings of the 10th European Conference on Interactive TV and Video (EuroITV '12)*. ACM, New York, NY, USA, 45–48. https://doi.org/10.1145/2325616.2325626

[63] Radu-Daniel Vatavu. 2013. The Impact of Motion Dimensionality and Bit Cardinality on the Design of 3D Gesture Recognizers. *International Journal of Human-Computer Studies* 71, 4 (April 2013), 387–409. https://doi.org/10.1016/j.ijhcs.2012.11.005

[64] Radu-Daniel Vatavu. 2017. Beyond Features for Recognition: Human-Readable Measures to Understand UsersâĂŹ Whole-Body Gesture Performance. *International Journal of HumanâĂŞComputer Interaction* 33, 9 (2017), 713–730. https://doi.org/10.1080/10447318.2017.1278897

[65] Radu-Daniel Vatavu. 2017. Smart-Pockets: Body-Deictic Gestures for Fast Access to Personal Data during Ambient Interactions. *International Journal of Human-Computer Studies* 103, C (July 2017), 1–21. https://doi.org/10.1016/j.ijhcs.2017.01.005

[66] Radu-Daniel Vatavu, Gabriel Cramariuc, and Doina Maria Schipor. 2015. Touch Interaction for Children Aged 3 to 6 Years: Experimental Findings and Relationship to Motor Skills. *International Journal of Human-Computer Studies* 74 (2015), 54–76. https://doi.org/10.1016/j.ijhcs.2014.10.007

[67] Radu-Daniel Vatavu, Bogdan-Florin Gheran, and Maria Doina Schipor. 2018. The Impact of Low Vision on Touch-Gesture Articulation on Mobile Devices. *IEEE Pervasive Computing* 17, 1 (Jan. 2018), 27–37. https://doi.org/10.1109/MPRV.2018.011591059

[68] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 1325–1334. https://doi.org/10.1145/2702123.2702223

[69] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2016. Between-Subjects Elicitation Studies: Formalization and Tool Support. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3390–3402. https://doi.org/10.1145/2858036.2858228

[70] Eduardo Velloso, Dominik Schmidt, Jason Alexander, Hans Gellersen, and Andreas Bulling. 2015. The Feet in Human–Computer Interaction: A Survey of Foot-Based Interaction. *ACM Comput. Surv.* 48, 2, Article 21 (Sept. 2015), 35 pages. https://doi.org/10.1145/2816455

[71] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the Guessability of Symbolic Input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI EA '05)*. ACM, New York, NY, USA, 1869–1872. https://doi.org/10.1145/1056808.1057043

[72] Jacob O. Wobbrock, Meredith R. Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1083–1092. https://doi.org/10.1145/1518701.1518866

[73] Jacob O. Wobbrock, Brad A. Myers, and John A. Kembel. 2003. EdgeWrite: A Stylus-based Text Entry Method Designed for High Accuracy and Stability of Motion. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)*. ACM, New York, NY, USA, 61–70. https://doi.org/10.1145/964696.964703

[74] Jacob O. Wobbrock, Andrew D. Wilson, and Yang Li. 2007. Gestures Without Libraries, Toolkits or Training: A $1 Recognizer for User Interface Prototypes. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology (UIST '07)*. ACM, New York, NY, USA, 159–168. https://doi.org/10.1145/1294211.1294238

[75] Ionut-Alexandru Zaiţi, Ştefan Gheorghe Pentiuc, and Radu-Daniel Vatavu. 2015. On Free-Hand TV Control: Experimental Results on User-Elicited Gestures with Leap Motion. *Personal and Ubiquitous Computing* 19, 5–6, 821–838. https://doi.org/10.1007/s00779-015-0863-y

[76] Mihai Zanfir, Marius Leordeanu, and Cristian Sminchisescu. 2013. The Moving Pose: An Efficient 3D Kinematics Descriptor for Low-Latency Action Recognition and Detection. In *Proceedings of the 2013 IEEE International Conference on Computer Vision (ICCV '13)*. IEEE Computer Society, Washington, DC, USA, 2752–2759. https://doi.org/10.1109/ICCV.2013.342

[77] Shumin Zhai, Per Kristensson, Caroline Appert, Tue Andersen, and Xiang Cao. 2012. Foundational Issues in Touch-Surface Stroke Gesture Design: An Integrative Review. *Foundations and Trends in Human-Computer Interaction* 5, 2 (Feb. 2012), 97–205. https://doi.org/10.1561/1100000012