The Kinematic Theory Produces Human-like Stroke Gestures

Luis A. Leiva[1,*] Daniel Martín-Albo[2] Réjean Plamondon[3]

[1] Sciling, Camí de Vera, s/n, 46022 València, Spain
[2] PRHLT Research Center, Universitat Politècnica de València, València, Spain
[3] Laboratoire Scribens, École Polytechnique de Montréal, QC, Canada

Abstract

We show that the Kinematic Theory produces synthesized stroke gestures that "look and feel" the same and hold the same statistical characteristics as human-generated gestures. Previous research in this vein has conducted such comparison from the classification accuracy performance, which is a legitimate though indirect measure. In this article, we synthesized two well-known public datasets comprising unistroke and multistroke gestures. We then compared geometric, kinematic, and articulation aspects of human and synthetic gestures, and found no practical differences between both populations. We also conducted an online survey involving 236 participants and found that it is very difficult to tell human and synthetic gestures apart. We can finally be confident that synthesized gestures are actually reflective of how users produce stroke gestures. In sum, this work enables a deeper understanding of synthetic gestures' production, which can inform the design of better gesture sets and development of more accurate recognizers.

Author Keywords: Gesture Synthesis; Bootstrapping; Kinematic Theory; Sigma-Lognormal Model; Strokes; Marks; Symbols; Unistrokes; Multistrokes; User Interfaces; Rapid Prototyping

1 Introduction

Over the past few years, touchscreen-based products like smartphones and tablets have notably increased the popularity of stroke gestures as commands and symbols. Put it simply, strokes gestures represent the movement trajectory of one or more contact points on a sensitive surface. User feedback is entered e.g. by means of a tablet, a touchscreen, or any other input device that can produce

*Corresponding author. Email: `name@sciling.com`

a temporal sequence of spatial coordinates representing specific actions (Leiva et al. 2014). Compared to traditional interactions, stroke gestures have the potential to lower cognitive load and the need for visual attention (Appert and Zhai 2009, Zhai et al. 2012). Stroke gestures also may improve the usability of UIs, by replacing standard shortcuts by more accessible triggers.

Training a high-quality gesture recognizer requires providing a large number of examples to enable good performance on unseen, future data. However, recruiting participants, data collection and labeling, etc. necessary for achieving this goal are usually time-consuming and expensive (Amini and Li 2013, Kohlsdorf and Starner 2013). Previous works have proposed to address this problem by generating synthetic samples (Anquetil et al. 2007, Djioua and Plamondon 2008b, Hollerbach 1981, Simard and LeCun 1991). Among these works, Djioua and Plamondon (2008a) presented a technique for synthesizing strokes using the kinematic theory of rapid human movements (Plamondon 1995a) and its associated Sigma-Lognormal model (Plamondon and Djioua 2006). This model has demonstrated to be the most accurate descriptor of human movements in a wide number of scenarios. For example, reproducing wrist movement and eye saccades (Plamondon 1995b), 2D and 3D arm movements (Leduc and Plamondon 2001), and more recently, stroke gestures (Almaksour et al. 2011). However, researchers have evaluated the gesture synthesis from the perspective of classification performance and not from the perspective of how "human-like" the synthesized gestures really are. A notable exception is a study by Galbally et al. (2012) involving 25 participants who examined the human likeness of synthetic handwritten signatures. While it was found a high degree of similarity between synthesized and human signatures, it is unclear whether this will hold for stroke gestures. In sum, a dedicated evaluation has been pending for too long.

In this article, we empirically demonstrate that the Sigma-Lognormal model ($\Sigma\Lambda$M) produces synthetic stroke gestures that hold similar statistical characteristics as human-generated gestures, evidencing thus that they "look and feel" the same. We used relative measures (Vatavu et al. 2013) comparing geometric, kinematic, and articulation aspects to support this assertion. We also conducted an online survey involving 236 participants and found that it is very difficult to tell human and synthetic gestures apart.

This work may have a significant impact well beyond HCI, since now researchers and practitioners can be confident that synthesized gestures using $\Sigma\Lambda$M are actually reflective of how users produce stroke gestures. Taken together, our results enable a deeper understanding of synthetic gestures' production, which can inform the design of gesture interaction by (1) automatically augmenting current gesture sets with more human-like samples and, in consequence, (2) building more accurate gesture recognizers.

2 Related Research

Training data is the key factor to build a competitive gesture recognizer. For example, the Freehand Formula Entry System (Smithies et al. 2001) suggests

20–40 examples per symbol per user. Further, Koch et al. (2010) studied gesture recognition using a Wiimote controller and found that 120 training patterns of accelerometer-based data is a lower bound; below that threshold the error rate increased dramatically. Synthesizing new samples can thus be used to improve rapidly the recognition performance. For example, it has been shown that a classifier "resists" better when introducing new samples and it is able to re-estimate rapidly all its parameters and to improve its recognition performance (Almaksour et al. 2011, Plamondon et al. 2014).

2.1 Bootstrapping Gestures by Synthesis

In the literature, we can find the following competing systems aimed at creating synthetic gestures as a means to improve gesture recognizers. First, Gesture Script (Lü et al. 2014) allows developers to describe the structure of a stroke gesture and its parts. With this information, Gesture Script synthesizes new gesture samples by changing the relative scale of each part and their rotation angles. Unfortunately, Gesture Script can only deal with unistroke gestures that are performed in a unique way. Besides, having to provide too detailed information for each gesture can be time-consuming.

Second, MAGIC Summoning (Kohlsdorf and Starner 2013) and Gesture Follower (Caramiaux et al. 2014) provide the user with a means of generating synthetic gesture samples in 3D space. MAGIC Summoning performs local perturbations to a gesture's resampled points, whereas Gesture Follower introduces some variations to a gesture template using Viviani's curve formulation. Both approaches are promising, although artificial samples that are generated this way might perform poorly since they do not illustrate sufficient variation required for high-quality training (Plamondon et al. 2014). However, these prior projects put forward the ongoing importance of and interest in improving gesture recognition by acquiring large data samples.

Finally, and more relevant to this work, G3 (Leiva et al. 2016) produces synthetic stroke gestures by means of the Kinematic Theory. Concretely, G3 creates a model of a user-provided gesture example and introduces local and global perturbations to the model parameters. This results in realistic human-like gestures, although human likeness was measured indirectly, inter-mediated by classification accuracy performance.

2.2 Gesture Production Features

Stroke gesture production has been studied in different ways in the literature, including the consistency between and within users (Anthony et al. 2013), preferences between user populations (Kane et al. 2011), and the impact of input device (Tu et al. 2012). What has been missing, however, is a "fine-grained analysis of gesture articulations to support an understanding of how gestures vary relative to each other and to recognizers' canonical template forms" (Vatavu et al. 2013).

Even gesture recognition algorithms use features that can be potentially used as measures of gesture production. However, most measures lack descriptive power because they capture global characteristics about the gesture as a whole; such as gesture path length, average speed, or articulation time. Fortunately, the Gesture RElative Accuracy Toolkit (GREAT) (Vatavu et al. 2013) provides fine-grained features (to be introduced later) that reveal subtleties about the gesture production process. More specifically, these measures describe the way gestures unfold and what happens during their production in terms of their closeness to a reference form, analogous to MacKenzie et al.'s pointing accuracy measures (MacKenzie et al. 2001). We therefore used this toolkit to compare the production of synthesized and human-generated gestures.

3 A Kinematic Theory Primer

Many models have been proposed to study human movement production; e.g., models relying on neural networks (Bullock and Grossberg 1988), behavioral models (Thomassen et al. 1983), or models exploiting minimization principles (Flash and Hogan 1985). Among these, the kinematic theory of rapid human movements (Plamondon 1995a) provides a well-established and solid framework for the study of the production of human movements (Plamondon et al. 1993). This framework takes into account different psychophysiological features, such as the neuromuscular response time, and has been shown to outperform many other approaches; see Discussion and Implications. The ΣΛM (Plamondon and Djioua 2006) is the latest instantiation of this framework, and very recently has been used to explore gesture recognition. For example, it has been shown that synthesized gestures achieve a similar recognition accuracy as their human counterparts in terms of articulation speed, size of gesture vocabulary, and input device (Leiva et al. 2016). It should be noted that the word "rapid" in the Kinematic Theory name is due to historical reasons, as the first models were aimed at studying truly rapid movements, such as those involved in handwritten signatures. Lately it has been shown that ΣΛM generalizes to any type of movements (Plamondon and Djioua 2006). We should note that ΣΛM creates a model of a particular movement; it is not a user model but a handwriting model of a particular handwritten trajectory.

3.1 Gesture Reconstruction

At a high-level representation, ΣΛM assumes that a complex handwritten trace (e.g. a character, a digit, a word, a signature, or a gesture) is composed of a series of primitives[1] (circular arcs) connecting a sequence of virtual targets. This series of primitives conform the "action plan" of the user, which is fed through

[1]In gesture recognition, a "stroke" denotes the trajectory between two consecutive pen-down and pen-up events. In the Kinematic Theory, a "stroke" is what we call "primitive" in this article.

the neuromuscular network to produce a trajectory that leaves a handwritten trace; see Figure 1.

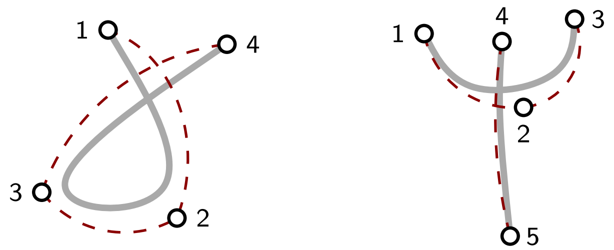

Figure 1: A gesture stroke (solid lines) is described by the temporal overlap of a series of primitives (dashed arcs) connecting a sequence of virtual targets (numbered circles). Each primitive is described by a lognormal velocity profile.

Mathematically, the magnitude of the velocity of the ith primitive is described by a lognormal-shaped function scaled in amplitude by a command parameter D_i and time-shifted by the time occurrence t_{0_i} of this command:

$$\begin{aligned} \|\vec{v}_i(t)\| &= D_i \Lambda(t; t_{0_i}, \mu_i, \sigma_i^2) \\ &= \frac{D_i}{\sigma_i \sqrt{2\pi}(t - t_{0_i})} \exp\left(\frac{-[\ln(t - t_{0_i}) - \mu_i]^2}{2\sigma_i^2} \right) \end{aligned} \tag{1}$$

where μ_i and σ_i define the variability of the neuromuscular execution of the ith motor command.

The trajectory that produces the human movement $\vec{v}(t)$ is computed as the temporal overlap of each primitive's velocity $\vec{v}_i(t)$:

$$\vec{v}(t) = \sum_{i=1}^{N} \vec{v}_i(t) = \sum_{i=1}^{N} \begin{bmatrix} \cos \phi_i(t) \\ \sin \phi_i(t) \end{bmatrix} D_i \Lambda(t; t_{0_i}, \mu_i, \sigma_i^2) \tag{2}$$

where the angular position $\phi_i(t)$ is obtained by:

$$\phi_i(t) = \theta_{s_i} + \frac{\theta_{e_i} - \theta_{s_i}}{2} \left[1 + \operatorname{erf}\left(\frac{\ln(t - t_{0_i}) - \mu_i}{\sigma_i \sqrt{2}} \right) \right] \tag{3}$$

being θ_{s_i} and θ_{e_i} the starting angle and the end angle of a given primitive, respectively.

Finally, the reconstruction of the original trajectory can be computed using the following compact notation (O'Reilly and Plamondon 2009):

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \sum_{i=1}^{N} \frac{D_i}{\theta_{e_i} - \theta_{s_i}} \begin{bmatrix} \sin \phi_i(t) & - & \sin \theta_{s_i} \\ -\cos \phi_i(t) & + & \cos \theta_{s_i} \end{bmatrix} \tag{4}$$

3.2 Gesture Synthesis

Previous works have demonstrated the connection between the distortion of the ΣΛM parameters and the intra-variability found in human handwriting (Djioua and Plamondon 2009). This way, it is possible to produce a synthetic sample (Leiva et al. 2016, Martín-Albo et al. 2014):

$$p_i^* = p_i + n_{p_i} \tag{5}$$

where $p_i = \{\, \mu_i, \sigma_i, D_i, \theta_{s_i}, \theta_{e_i} \}$ denote the ΣΛM parameters, with $n_{p_i} = \mathcal{U}(-n_i, n_i)$ being the noise applied to each primitive, according to a uniform distribution (i.e., a rectangular distribution with constant probability, Figure 2) centered around the expected human variability ranges (see Section 4.2). The different noise values, to be described in the next section, are set according to previous work (Galbally et al. 2012).

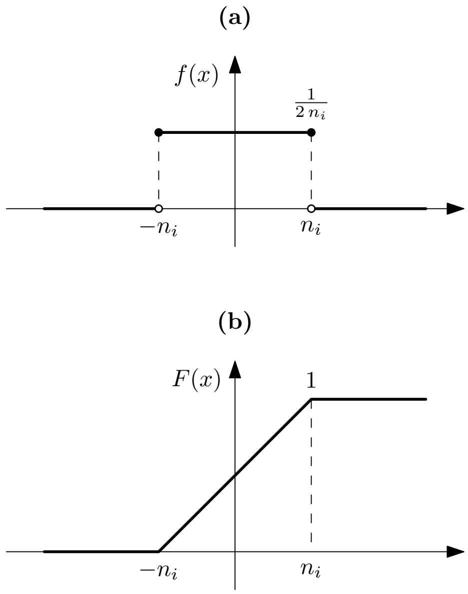

Figure 2: Probability density function (a) and cumulative distribution function (b) of the noise levels (n_i) applied to each primitive's parameter.

Concretely, the noise in the peripheral parameters (μ and σ) mimics a writer who instantiates the same gesture intention and executes it with an upper limb slightly different from one trial to another. The variations of these parameters involves smoothing and sharpening effects in the trajectory, which result in a modification of the overlaps between neighboring primitives, leading to smoother or sharper trajectories. Moreover, the noise in the control parameters (D, θ_s and θ_e) refers to temporal and spatial fluctuations in the action plan, reflecting, for example, attention changes from one trial to another. In particular, variations in the command parameter (D) introduces scaling perturbations, whereas variations in the directional parameters (θ_s and θ_e) introduces rotational perturbations. Combining both types of variations reflects real-life situations like performing the same movement under different psychophysiological conditions.

4 Evaluation

In order to confirm the fundamental assertion raised by this article's title, we replicated two well-known public datasets in HCI. Then, we conducted an online study to assess the subjective perception that non-expert human observers have of synthetic gestures. Finally, we used a number of relative measures (gesture descriptors) to study the differences between the original (human) and the artificial (synthetic) gesture samples.

4.1 Datasets

The following are two reference datasets in HCI to test unistroke and multistroke-based recognizers. Both datasets provide XML files containing gesture points with millisecond timestamps.

GDS: Available at https://depts.washington.edu/aimgroup/proj/dollar/xml.zip. Comprises 16 unistroke gesture classes, 5,280 samples in total (Wobbrock et al. 2007). Ten users (plus 1 pilot user) provided 10 samples per class at 3 articulation speeds (slow, medium, fast) using an iPAQ Pocket PC (stylus as input device).
For slow speed, users were asked to "be as accurate as possible;" for medium speed, users were asked to "balance speed and accuracy;" for fast speed, users were asked to "go as fast as you can".

MMG: Available at https://depts.washington.edu/aimgroup/proj/dollar/mmg.zip. Comprises 16 multistroke gesture classes,[2] 9,600 samples in total (Anthony and Wobbrock 2012). Twenty users provided 10 samples per class at 3 articulation speeds (same as in GDS) using either finger (half of the users) or stylus as input device on a Tablet PC. Speed definitions are the same as in the GDS dataset.

4.2 Gesture Reconstruction

Each gesture sample in both datasets was modeled with ΣΛM, which uses 6 parameters $(D, t_0, \mu, \sigma, \theta_s, \theta_e)$ per "primitive" (see Figure 1). For this, we used a ΣΛM parameter extractor (Martín-Albo et al. 2015) that provides the required set of parameters for Equations (1) to (4).

Then, each primitive's parameter was perturbed with a different amount of noise, according to Equation 5, using a uniform distribution (Figure 2) that spans the expected human variability ranges: $n_\mu = n_\sigma = 0.1$, $t_0 = 0.005$, $n_D = 0.15$, $n_{\theta_s} = n_{\theta_e} = 0.06$. These values were estimated according to 6,400 signature samples from 400 users (Galbally et al. 2012), acquired in 5 different Spanish universities under very controlled conditions. Furthermore, these samples were captured in 4 acquisition sessions over a 6-month time span, which proves them robust to estimate both inter- and intra-session variability. We should note the impossibility of replicating the same conditions with the datasets we analyzed in

[2]Actually, 13% of the gestures in the MMG dataset are unistrokes.

this article.[3] Finally, after parameter perturbation, the Cartesian coordinates (x,y) were retrieved using Equation 4. This final step therefore produces a synthetic gesture sample. Figure 3 shows an example of synthesized gestures and their human counterparts.

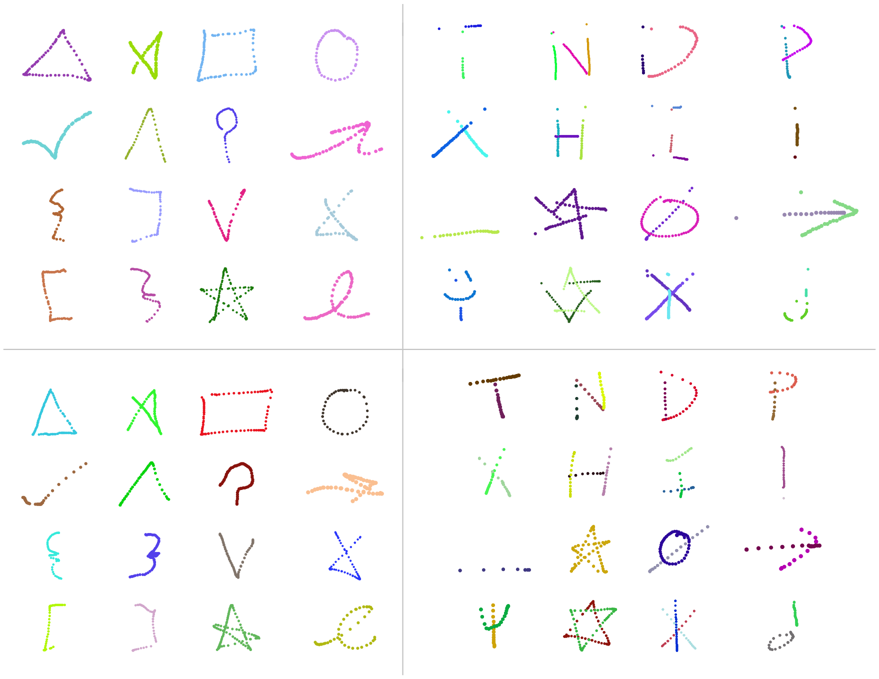

Figure 3: Examples of human (top row) and synthetic (bottom row) stroke gestures from the GDS (left column) and MMG (right column) datasets.

4.3 Visual Appearance Study

The set of MMG + GDS (both human and synthetic samples) was made available through an online study.[4] This amounts to $(5{,}280 + 9{,}600) \times 2 = 29{,}760$ gestures in total. The study was advertised in social networks, online chats, and blogs, ensuring that users had no expert knowledge on gesture recognition. Eventually, 236 users took part in the study.

Each user was presented with 10 samples drawn at random, half of which were synthetic, and the user had to click on a button to mark whether the gesture shown was human or machine-generated (Figure 4). The maximum time permitted to assess each gesture was 4 seconds at most. This was so because the overall objective of this experiment was not making a detailed and profound

[3] To accomplish this goal, each gesture should be annotated both at the stroke *and* at the "primitive" level (Figure 1). This information is unfortunately not available in these datasets.
[4] Available at [URL]

analysis of each gesture, but estimating the general visual appearance of gesture samples after a short inspection. This experiment thus aimed for recreating as much as possible the settings used by Galbally et al. (2012). A "Skip this guess" button allowed the user to load a different sample when she was unsure if it was synthetic or human.

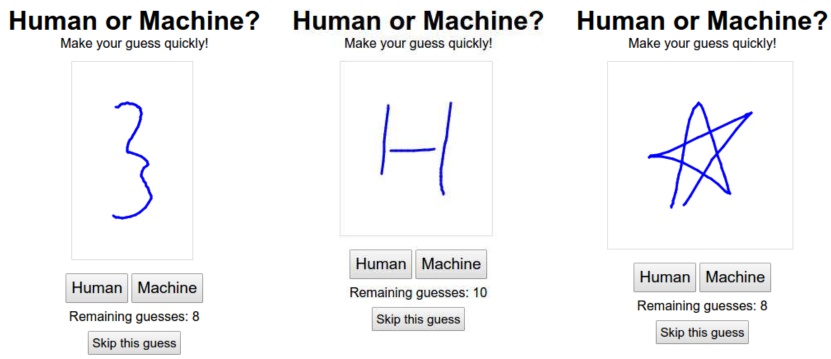

Figure 4: Screenshots of the application to conduct the visual appearance study. Can you guess which ones are synthetic and which ones are human-generated?

4.3.1 Results

We define the guessing accuracy (ACC) as the user's capability to distinguish between human and synthetic samples; i.e., how many samples were successfully classified by the user. Then, two types of errors can be committed in this study (Galbally et al. 2012): (1) a synthetic gesture is mistaken with a real sample, measured by the False Real Rate (FRR, Type I error); and (2) a real gesture is marked as synthetic, measured by the False Synthetic Rate (FSR, Type II error). These error rates are presented in Table 1.

Guessing accuracy		False Real Rate		False Synthetic Rate	
Mean	**95% CI**	**Mean**	**95% CI**	**Mean**	**95% CI**
49.65	[47.19, 52.11]	29.84	[27.33, 32.35]	20.40	[18.19, 22.60]

Table 1: Accuracy and error rates (in %, mean and 95% confidence intervals) of the users who took part in the online study (N=236 participants). The sum of False Real Rate (Type I error) and False Synthetic Rate (Type II error) yields the overall error rate (100 - accuracy).

From the results presented in the table above, we can see that half of the gestures (49%) were misclassified. Simply put, it is not easy to distinguish one type of gestures over the other. A paired two-sample t-test (two-tailed alternative hypothesis) revealed that there is no difference between the distributions of classified and misclassified gestures [$t(235) = 0.420, p = .674$, n.s.] proving thus the human-like appearance of synthetic samples. Figure 5 shows both distributions.

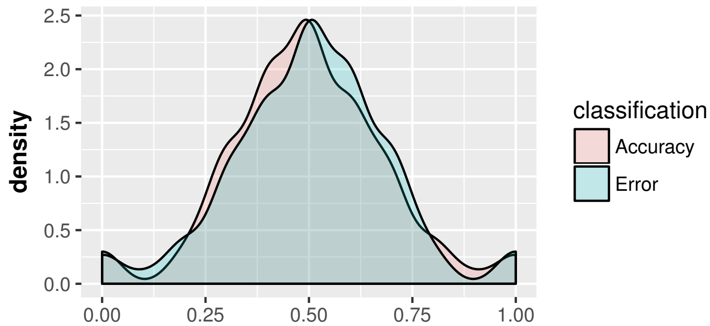

Figure 5: Distributions of classified and misclassified samples.

It should be noticed that a random guesser (e.g. by flipping a coin) would achieve an ACC of 50%. It should also be noticed that both error rates are comparable (FRR=29.8% and FSR=20.4%) as just one misclassified gesture leads to a 10% error rate in our study (each user is shown 10 samples). This means that the number of mistaken real and synthetic samples is similar, while it is true that participants misclassified slightly more human gestures as synthetic than the other way around. In sum, participants could not tell human and synthetic gestures apart. We conclude therefore that the visual appearance of the synthetic samples is very similar and close to that of human gestures.

4.4 Gesture Descriptors Analysis

In addition to the observable similarity between the real and synthetic gestures appearance (patent from the results obtained in the previous study), a more formal experiment was carried out in order to assess these differences.

We used GREAT (Vatavu et al. 2013) to compute geometric, kinematic, and articulation features of both synthetic and human stroke gestures production. Concretely, GREAT computes 12 gesture descriptors on the gesture path, as introduced below. The reader can refer to Vatavu et al. (2013) for a detailed description thereof. The key idea is that we use a well-defined set of measures that describes different production aspects of a set of gestures as a whole, so we can compare the results pertaining human samples with their synthesized counterparts.

GREAT evaluates stroke gestures production relative to a gesture task axis. The gesture task axis is a fixed example gesture, reflective of relative differences between individual executions, that serves as a reference against which the measures of other candidate gestures are computed. Therefore, which reference should be considered? How should gesture points be aligned? First, we chose the k-medoid of each gesture class as task axis, as it is less sensitive to noise and outliers (Figure 6). The k-medoid is the closest user-articulated sample to the median gesture.[5] Second, unistroke gestures were aligned in their chronological order of input; whereas multistroke gestures were aligned using the point-cloud matching procedure (Vatavu et al. 2012), which is invariant to the number of strokes and stroke ordering. Prior to alignment, gestures were resampled to 32 points and centered at the origin $(0, 0)$.

(a) centroid **(b)** k-centroid **(c)** medoid **(d)** k-medoid

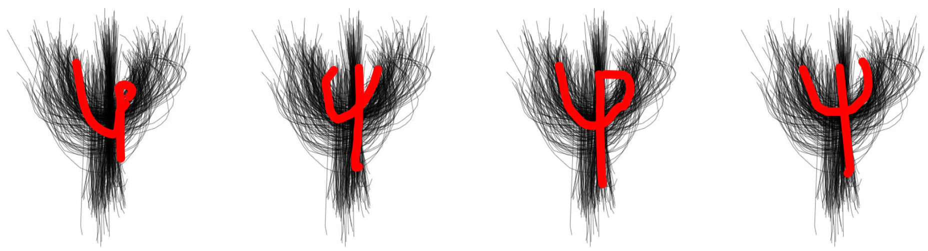

Figure 6: Different task axes (red thick lines) computed for all samples (gray thin lines) of a multistroke gesture from the MMG dataset. Being more robust to noise and outliers, the k-medoid (d) was chosen to compute all relative measures.

In the following we enumerate the 12 gesture descriptors computed by the GREAT toolkit. We should note that in our experiments the task axis is the k-medoid.

Geometric Features or *shape*-related descriptors evaluate the deviation of a candidate gesture from the task axis in terms of shape distance, and capture tendencies of the users to stretch and bend strokes during articulation.

1. Shape Error (px) represents the average absolute deviation of the candidate gesture points from the task axis in terms of the Euclidean distance.
2. Shape Variability (px) computes the standard deviation of the distances between the points of the candidate and the task axis.
3. Length Error (px) measures users' tendencies to stretch gesture strokes with respect to the task axis.
4. Size Error (px^2) measures users' tendencies to stretch gesture strokes in terms of the gesture area size.
5. Bending Error (rad) measures users' tendencies to bend the strokes of the articulated gesture with respect to the gesture task axis.
6. Bending Variability (rad) computes the standard deviation of the differences in turning angle.

[5]We modified GREAT to compute the task axes in Figure 6.

Kinematic Features or *speed*-related descriptors evaluate articulation differences in the time domain, and capture how fluent or smooth the articulated path is in terms of production time and speed.

7. Time Error (ms) measures the difference in articulation time between the candidate and the task axis.
8. Time Variability (ms) represents the standard deviation of the differences between timestamps measured at each individual point on the gesture path.
9. Speed Error (px/ms) measures the difference in the speed profiles of the candidate and the gesture task axis.
10. Speed Variability (px/ms) represents the standard deviation of the local differences between the speed profiles.

Articulation Features or *consistency*-related descriptors measure how consistent users are in producing the individual strokes of gestures.

11. Stroke Count Error reports the difference in the number of strokes between the candidate and the task axis.
12. Stroke Order Error (px) computes the absolute difference between Euclidean distances considering both chronological point-wise alignment or not.

4.4.1 Results

We analyzed the gesture samples both in user-independent and user-dependent scenarios. In user-independent tests, each gesture sample is considered an independent observation. In user-dependent tests, all gesture descriptors are computed first for each user and then are aggregated. We used unpaired two-sample t-tests (two-tailed, Bonferroni corrected) for these comparisons. Figure 7 and Figure 8 summarize the results.

As can be observed in the figures, we did not find statistically significant differences between synthesized and human-generated gestures in most cases. Interestingly, significant differences in Speed Error and Speed Variability in the MMG dataset appeared because mostly half of the timestamps in the human samples are duplicated (M=1.7, SD=0.3), possibly due to being acquired with "higher-than millisecond" precision. Having duplicated timestamps in the original MMG dataset implies that a human gesture may have more than one point at the same time t, which cause point-wise misalignments that influence how these relative measures are computed. In contrast, ΣΛM provides a continuous function of t, therefore only one synthesized point is allowed at a given t, which explains the differences in these speed-related measures.

However, how important are these results? When examining large samples, significance testing can be misleading because even small differences are likely to produce a statistically significant result. Indeed, as shown in Figure 7 and Figure 8, differences between populations are rather small; see e.g., Shape

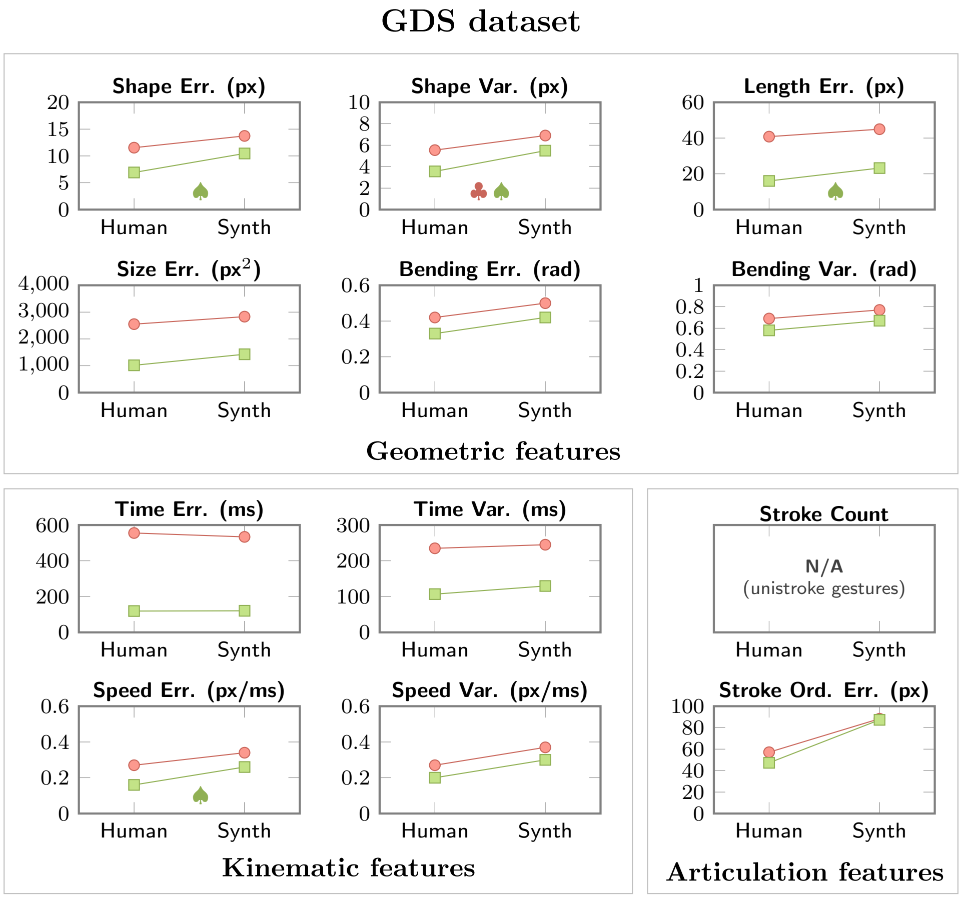

Figure 7: Overall results with human and synthesized gestures over the GDS dataset. Circles (●) denote user-independent tests and squares (■) denote user-dependent tests. 95% confidence intervals are all below 1%, so they are omitted. A statistically significant difference between populations ($p < .05/12$) is denoted with clubs (♣) for user-independent tests and spades (♠) for user-dependent tests. Effects sizes (M=0.13, SD=0.1) suggest low practical significance.

Variability: 5.5 *vs.* 6.9 px (GDS, user-independent) or Speed Error: 1.7 *vs.* 0.6 px/ms (MMG, user-dependent). Furthermore, the observed effect sizes in all cases suggest low practical significance (Cohen's $d < 0.2$, M=0.13, SD=0.1). This was true both for user-dependent and user-independent tests. It should be noted that effect sizes have consequential validity, as they describe the magnitude of the differences.

Next, we split each dataset in terms of input speed and repeated the same analysis. Again, we did not find statistically significant differences in most cases

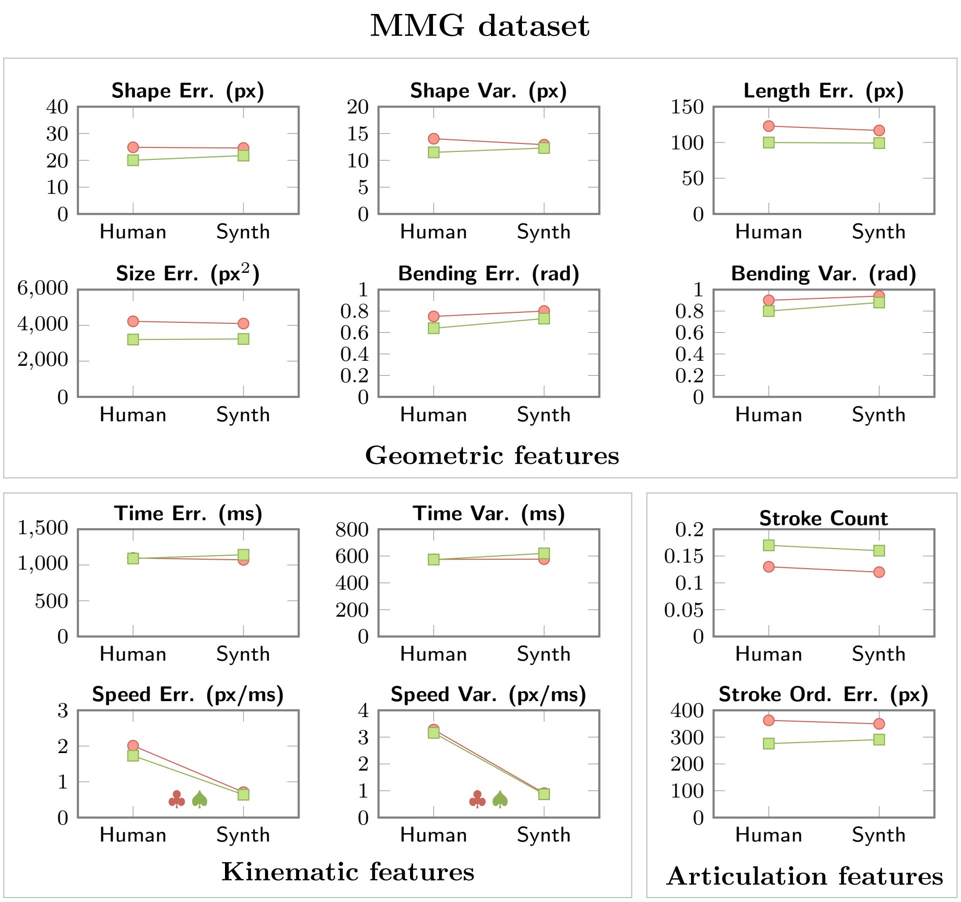

Figure 8: Overall results with human and synthesized gestures over the MMG dataset. Circles (●) denote user-independent tests and squares (■) denote user-dependent tests. 95% confidence intervals are all below 1%, so they are omitted. A statistically significant difference between populations ($p < .05/12$) is denoted with clubs (♣) for user-independent tests and spades (♠) for user-dependent tests. Effects sizes (M=0.13, SD=0.1) suggest low practical significance.

(not discussed for brevity's sake) with similarly low effect sizes. We also analyzed the MMG dataset in terms of input device (stylus vs. finger) and observed the same outcome. Finally, we also conducted an additional experiment to analyze the effect of gesture type, and we confirm that there is no such an effect either. Notice that in this experiment each gesture class is of intrinsic interest so, contrary to the previous experiments, it is not considered a random effect.

Therefore, we conclude that there is no practical difference in gesture production between human and synthesized samples and that they "look and feel" the same. In other words, synthesized gestures are actually reflective of how users produce stroke gestures. This is an important result because recent literature on gesture analysis has shown how different users produce different gesture articulations in various conditions (see Section 2.2). We elaborate more on this discussion in the next section.

5 Discussion and Implications

In this section we connect the main findings of this work with the following areas of interest to the HCI community.

5.1 Kinematic Theory Modeling

The previous instantiation of the Kinematic Theory (the $\Delta\Lambda$M) assumed that the production of a stroke requires the synergetic activation of two neuromuscular systems, one agonist and the other antagonist to the direction of the movement. These synchronous commands propagate in parallel across the two neuromuscular systems, each of which is described by a lognormal impulse response and has its own timing properties. On the contrary, the $\Sigma\Lambda$M does not assume that the two neuromuscular systems are working in precisely opposite directions. The output velocity is thus described by a vectorial summation of the contribution of each neuromuscular system involved in the production of a stroke. This model is actually very general, and is not limited to a single stroke description (O'Reilly and Plamondon 2009, Plamondon and Djioua 2006). Our work builds upon this fundamental notion, however we are the first to use the $\Sigma\Lambda$M to study the production of stroke gestures. This corroborates the prediction of the Kinematic Theory, where it is theorized that every human movement has a lognormal impulse response that results from the limiting behavior of a large number of interdependent neuromuscular networks.

It must be noted that the $\Sigma\Lambda$M is both rotation and scale invariant. On the one hand, the directional parameters (θ_s and θ_e) account for the overall rotation of the handwriting pattern. Djioua and Plamondon (2009) asked several users to write the same word both inside a 3x1.5 cm rectangle and inside the same-sized rectangle but rotated 35 degrees. It was observed that all $\Sigma\Lambda$M parameters remained the same excepting the directional parameters which increased by 35 degrees. On the other hand, the command parameter (D) accounts for the overall size of the handwriting pattern. Djioua and Plamondon (2009) also asked several users to write the same word both inside a 3x1.5 cm rectangle and then inside a double-sized rectangle (6x3 cm). It was observed that the $\Sigma\Lambda$M parameters remained the same excepting the command parameter D, which increased by a factor of 2. In sum, $\Sigma\Lambda$M can deal with any handwriting pattern, regardless of their rotation or scale.

Finally, it should be noted that the Kinematic Theory is not limited to the control of finger or hand movements; it has also been used to analyze wrist, arm, head movements, and eye saccades. In other words, the Kinematic

Theory provides a complete parametric representation space to study motor control behavior. This mathematical demonstration suggests that the asymptotic convergence toward lognormal impulse responses and velocity patterns can be interpreted as reflecting the behavior of subjects who are in total control of their movements. Additionally, from a mathematical point of view, the Kinematic Theory is a theory of convergence toward smoothness. The lognormal function is an optimal descriptor of the velocity profiles: the smoothest velocity being reached when the energy associated with the convergence error toward lognormality is minimized. As such, the Kinematic Theory can be considered as an ultimate minimization theory.

5.2 Artificial Gesture Noise

This work has shown that synthesizing gesture samples using lognormal-based deformations on velocity profiles produces human-like results, as measured by relative measures that compared geometric, kinematic, and articulation aspects of stroke gestures. As discussed in Section 3, previous studies have pointed out a strong connection between the distortion of the ΣΛM parameters and the intra-variability found in complex human movements (Djioua and Plamondon 2009, Martín-Albo et al. 2014). Furthermore, a number of previous works have shown that such distortion of the ΣΛM parameters in turn improves an existing recognizer's accuracy. For example, Almaksour et al. (2011) used a small set of human gestures plus distorted ones using ΣΛM and Martín-Albo et al. (2014) replicated a dataset using the same distortions in a writer adaptation experiment. In both cases, recognizer accuracy improved notably; e.g. in the latter experiment, the inclusion of synthetic samples generated from 50 human samples improved accuracy by 20%. In this regard, Plamondon et al. (2014) showed that augmenting the training set with synthetic gestures from 10 human samples per class reduced the error rate by 50%. Finally, Leiva et al. (2016) showed that synthesizing N samples from just 1 human sample leads to the same recognition accuracy as if they were all human samples. In contrast, our work is the first to quantify the production of synthetic stroke gestures.

As discussed in Section 3, the noise in the peripheral and control parameters reflect real-life situations like performing the same movement under different psychophysiological conditions. Synthesized gestures are thus sample-specific, since there is a model for each user gesture, however they are actually reflective of how users would produce them.

5.3 Human Control Behavior

The concepts concerning internal models of human movements have been well supported by behavioral studies in the field of sensory motor control. Overall, it is assumed that users are "ideal" motion planners who choose movement trajectories to minimize an expected loss (Trommershäuser et al. 2003). Currently, we can find two compelling theories to describe those movements: the Minimization Theory (Flash and Hogan 1985) and the Kinematic Theory (Plamondon 1995a).

Actually, it has been shown that their concepts are linked and describe, with different arguments, a model of velocity profiles (Djioua and Plamondon 2010). However, from a fundamental perspective, it has been shown that lognormal-based models are the most accurate descriptors of human movements and that other models are successive approximations (Djioua and Plamondon 2009). In fact, Plamondon et al. (1993) compared 23 different models to describe human movements and found that the lognormal approach outperfomed all of the other approaches. And even though ΣΛM has been mainly applied to handwriting analysis, our study shows that it can be successfully applied to stroke gestures, too.

On the other hand, experimental studies have consistently found that the kinematics of arm movements are highly stereotypical under a large variety of experimental conditions (Engelbrecht 2001). Such a generalization has lead to postulate the underlying existence of a lognormality principle that guides human beings throughout their life, from the early steps of their motor learning processes to increasing departure from the ideal lognormal behavior, as the control of the fine motricity begins to decline with age and illness (Plamondon et al. 2013). From a more pragmatic perspective, the Kinematic Theory considers that the velocity is the main control variable used by the central nervous system to generate a handwritten trajectory. From a physical point of view, the velocity is the sole information that can be partly recovered by visual inspection from an image (see e.g. Figure 3).

As discussed by Tu et al. (2012), the study on how humans control their motor behavior has historically centered on the debate between the centrists and the peripheralists among motor control theorists (Schmidt 1988). The centrists tend to view motor control behavior as an inside-out process, driven by "motor programs" from human internal representations. In contrast, peripheralists tend to emphasize motor control behavior as regulated by outside-in feedback from the environment. Thus, a centralist would suggest that there is little difference between synthetic and human gestures since their production are both driven from internal representations, as is indeed proposed in the effector independence theory concerning handwriting (Wright 1990). A peripheralist, however, would argue that the different feel and interaction with the touchscreen surface afforded by the pen vs. the bare finger would impact how a gesture is produced. Our study therefore sheds light to this discussion and puts forward the fact that the Kinematic Theory, via ΣΛM, produces actual human-like stroke gestures. Admittedly, ΣΛM successfully explains the variability observed in repeated generation of movements and how do they change over time. This is a desired property to efficiently generate multiple samples computationally.

Following this centralist *vs.* peripheralist debate, in this context one can generate numerous human like gestures using the same command parameters and modifying the peripheral parameters. However, this does not exclude the peripheral model, since human-like gestures can also be generated by keeping the same peripheral parameters and varying the command parameters. However, as shown by Djioua and Plamondon (2009), the sensitivity to fluctuations in the command parameters is higher in this latter case. One interesting feature of the Kinematic Theory is that it allows to investigate this debate further with a new mindset, since the parameters t_0, D, θ_s, and θ_e refer to central commands while the parameters μ and σ describe the peripheral reaction to these commands. Under specifically controlled experimental conditions, these questions could be studied and validated. For a given protocol affecting the central control, one should observe changes in the command parameters, leaving the peripheral parameters intact. Then, under another set of conditions, one could observe changes in the peripheral parameters, the command parameters remaining almost unaffected. Finally, under normal life conditions, most of the time it is expected that a mixture of central and peripheral parameters will be exploited to maximize some efficiency criteria. Overall, the capability of reconstructing any gesture in a human-like fashion provides a new window to address this issue and give researchers new tools to investigate these fundamental questions.

5.4 3D Gestures

In this article we have conducted different analyses on stroke gestures, which are in essence 2D trajectories. Thus, in future work we would like to extend this work to 3D gestures, derived from e.g. accelerometers, Kinects, Wii controllers, and similar devices. Actually, this kind of data can be reconstructed using lognormals provided that the torsion is taken into account, which just requires introducing an additional ΣΛM parameter (Leduc and Plamondon 2001). We have discussed interesting approaches to 3D gestures generation in Section 2.1. However, if we look at other research fields, we can also find interesting approaches to the generation of multidimensional trajectories. These approaches, of course, can serve as an inspiration to improve the synthesis of 3D gestures.

Concretely, in the field of robotics, Schaal et al. proposed a generic modeling approach to generate multidimensional systems using nonlinear differential equations to capture an observed behavior in an attractor landscape (Schaal et al. 2007; 2004). Their work assumed that a complex movement is composed of simpler patterns, called primitives, that can be used as building blocks for generating complex movements. There is a certain parallelism between this work and ours, and even between said work and the equilibrium models in the neuromotor control literature; see Section 5.3.

5.5 Practical Applications

Our work is primarily fundamental research, although it has a number of practical implications and applications. First, having a reliable way to synthesize gesture sets without having to expressly collect them from human subjects allow for unprecedented savings in terms of time and cost, since recruiting participants to a lab, collecting human samples, and data labeling are all time-consuming and expensive tasks. Also, being able to generate synthetic gestures that hold the same statistical characteristics as human-generated gestures enables a deeper understanding of synthetic gestures' production, which can inform the design of gesture interaction by e.g. automatically augmenting current gesture sets with more human-like samples and, in consequence, building more accurate gesture recognizers. Indeed, previous work (Leiva et al. 2016) has shown that increasing the number of gestures used in an application with synthetic samples do increase the recognizer's accuracy. More generally, improved accuracy has been confirmed when training with a dataset that is extended with synthetic data (Fischer et al. 2014, Galbally et al. 2012, Martín-Albo et al. 2014, Plamondon et al. 2014); e.g., from words or signatures collected from various writers sitting in front of a digitizer tablet, to sentences written on a whiteboard using full arm movements while standing up.

5.6 Open Fields

From a more general perspective, this article should be seen as one corner stone of a broad series of potential applications of the Kinematic Theory in the development of Personal Digital Bodyguards (Plamondon 2015) for e-security, e-health and e-learning. Indeed, by actively monitoring the user's handwriting activity, the Kinematic Theory proposes a new signal representation space and automatic segmentation capabilities that can be used for writer style characterization, automatic database generation, the development of new on-line recognizers and verifiers, as well as interactive tools to help children to learn handwriting. Furthermore, regarding biomedical signal processing, the Kinematic Theory proposes a new set of parameters to characterize the human motor control system. Therefore, the theory can be used for example to design psychomotor evaluation tests for the detection of fine motor control problems (e.g. Parkinson or Alzheimer diseases), for the development of prevention and rehabilitation tests and tools, or to study the effects of medication, alcohol, drugs, or weight loss. Finally, in a longer time perspective, we should mention that the Kinematic Theory proposes a new set of functions for 2D and 3D smooth curve modeling, which can be used for example in anthropomorphic arm design, exoskeletons and prosthetics, as well as in the modeling of humanoid movements in virtual reality environments.

5.7 Postscript

Finally, we should point out that the goal of this work is not to encourage the substitution of human gestures by synthetic ones, but rather to provide an automated way to lower the need of recruting a large number of users and subsequent data labeling. The use of synthetic gestures should be understood as a valid alternative in order to obtain a fast and reliable estimate of the actual performance of stroke gestures under controllable and repeatable conditions.

6 Conclusion

Our evaluations provide evidence against the implied alternate hypothesis of a difference between human and synthesized stroke gestures. Indeed, we have empirically demonstrated that the Kinematic Theory produces stroke gestures that "look and feel" the same as human-generated gestures. To support this assertion, we have conducted an online survey and found that it is very difficult for the users to tell human and synthetic gestures apart. We also have used relative measures to compare geometric, kinematic, and articulation aspects from thousands of human and synthetic gestures and found no practical differences between both populations under different conditions: device type (pen vs. finger), gesture type (unistrokes vs. multistrokes), and execution speed (slow vs. medium vs. fast). It is thus reliable to generate synthetic datasets this way, since the overall performance and behavior of gesture samples will be consistently similar to that of actual users.

Our findings have demonstrated the importance of our study: until now the "human likeness" of synthesized gestures has been measured indirectly, intermediated by classification/recognition accuracy performance. Our work is the first to quantify the production process itself; i.e., how synthetic gestures are articulated. This is important because recent literature on gesture analysis has shown how different users produce different gesture articulations in various conditions. Yet, finally researchers and practitioners can be confident that synthesized gestures using ΣΛM are actually reflective of how users produce stroke gestures.

In sum, this work provides the HCI community with a reliable way to synthesize gesture sets without having to expressly collect them from a large pool human subjects. However, we do not to encourage the substitution of human gestures by synthetic ones, but rather to provide an automated way to lower the need of recruiting a large number of users and subsequent data labeling. It is our hope that this work will be useful to anyone interested in improving gesture interaction by automatically augmenting current gesture sets with more samples that ultimately will help to building more accurate gesture recognizers.

Funding

This work was part of the International Excellence Campus program of VLC/-CAMPUS. This work has been partially supported by the Spanish MINECO under FPU scholarship AP2010-0575 and by NSERC-Canada grant RGPIN-2015-06409.

Acknowledgments

We thank the IwC referees for providing feedback to improve the manuscript. We also thank the participants who took part in the *Human or Machine?* online test: Muchachada Jaraíz, Colegio Nazaret, and Trabuquete's Team, among others.

End Notes

We make our sythesized datasets publicly available at https://luis.leiva.name/g3/ so that others can build upon our work. Our software can be accessed either through a web application or as a RESTful API at https://g3.prhlt.upv.es. Due to legal restrictions, the core software for gesture synthesis cannot be made publicly available as a standalone software. The interested people must sign an agreement with the École Polytechnique de Montréal through a collaborative project to get a license.

References

Almaksour, A., Anquetil, E., Plamondon, R., O'Reilly, C., 2011. Synthetic handwritten gesture generation using Sigma-Lognormal model for evolving handwriting classifiers. In: Proceedings of Biennial Conf. of the Intl. Graphonomics Society (IGS).

Amini, S., Li, Y., 2013. CrowdLearner: Rapidly creating mobile recognizers using crowdsourcing. In: Proceedings of Annual ACM Symp. on User Interface Software and Technology (UIST). pp. 163–172.

Anquetil, E., Miclet, L., Bayoudh, S., et al., 2007. Synthetic on-line handwriting generation by distortions and analogy. In: International Graphonomics Society. pp. 10–13.

Anthony, L., Vatavu, R.-D., Wobbrock, J. O., 2013. Understanding the consistency of users' pen and finger stroke gesture articulation. In: Proceedings of Graphics Interface (GI). pp. 87–94.

Anthony, L., Wobbrock, J. O., 2012. $N-protractor: a fast and accurate multistroke recognizer. In: Proceedings of Graphics Interface (GI). pp. 117–120.

Appert, C., Zhai, S., 2009. Using strokes as command shortcuts: Cognitive benefits and toolkit support. In: Proceedings of SIGCHI Conf. on Human Factors in Computing Systems (CHI). pp. 2289–2298.

Bullock, D., Grossberg, S., 1988. The VITE model: a neural command circuit for generating arm and articulator trajectories. In: Dynamic Patterns in Complex Systems. pp. 305–326.

Caramiaux, B., Montecchio, N., Tanaka, A., Bevilacqua, F., 2014. Adaptive gesture recognition with variation estimation for interactive systems. ACM TiiS 4 (4), 18:1–18:34.

Djioua, M., Plamondon, R., 2008a. An interactive system for the automatic generation of huge handwriting databases from a few specimens. In: Proceedings of Intl. Conf. on Pattern Recognition (ICPR). pp. 1–4.

Djioua, M., Plamondon, R., 2008b. The Kinematic Theory and minimum principles in motor control: A conceptual comparison. Tech. Rep. EPM-RT-2008-03, École Polytechnique de Montréal.

Djioua, M., Plamondon, R., 2009. Studying the variability of handwriting patterns using the Kinematic Theory. Hum. Mov. Sci. 28 (5), 588–601.

Djioua, M., Plamondon, R., 2010. The limit profile of a rapid movement velocity. Human Movement Science 29 (1), 48–61).

Engelbrecht, S. E., 2001. Minimum principles in motor control. Journal of Mathematical Psychology 45, 497–542.

Fischer, A., Plamondon, R., O'Reilly, C., Savaria, Y., 2014. Neuromuscular representation and synthetic generation of handwritten whiteboard notes. In: Proceedings of Intl. Conf. on Frontiers in Handwriting Recognition (ICFHR). pp. 222–227.

Flash, T., Hogan, N., 1985. The coordination of arm movements: an experimentally confirmed mathematical model. J. Neurosci. 5 (7), 1688–1703.

Galbally, J., Plamondon, R., Fierrez, J., Ortega-García, J., 2012. Synthetic on-line signature generation. Part II: Experimental validation. Pattern Recogn. 45 (7), 2622–2632.

Hollerbach, J. M., 1981. An oscillation theory of handwriting. Biol. Cybern. 39 (2), 139–156.

Kane, S. K., Wobbrock, J. O., Ladner, R. E., 2011. Usable gestures for blind people: Understanding preference and performance. In: Proceedings of SIGCHI Conf. on Human Factors in Computing Systems (CHI). pp. 413–422.

Koch, P., Konen, W., Hein, K., 2010. Gesture recognition on few training data using slow feature analysis and parametric bootstrap. In: Proceedings of IEEE Intl. Joint Conf. on Neural Networks (IJCNN). pp. 1–8.

Kohlsdorf, D., Starner, T. E., 2013. MAGIC Summoning: Towards automatic suggesting and testing of gestures with low probability of false positives during use. J. Mach. Learn. Res. 14 (1), 209–242.

Leduc, N., Plamondon, R., 2001. A new approach to study human movements: The three dimensional Delta-Lognormal model. In: Proceedings of Biennial Conf. of the Intl. Graphonomics Society (IGS). pp. 98–102.

Leiva, L. A., Alabau, V., Romero, V., Toselli, A. H., Vidal, E., 2014. Context-aware gestures for mixed-initiative text editing UIs. Interact. Comput. 27 (1).

Leiva, L. A., Martín-Albo, D., Plamondon, R., 2016. Gestures à go go: Authoring synthetic human-like stroke gestures using the kinematic theory of rapid movements. ACM TIST 7 (2).

Lü, H., Fogarty, J. A., Li, Y., 2014. Gesture Script: Recognizing gestures and their structure using rendering scripts and interactively trained parts. In: Proceedings of SIGCHI Conf. on Human Factors in Computing Systems (CHI). pp. 1685–1694.

MacKenzie, I. S., Kauppinen, T., Silfverberg, M., 2001. Accuracy measures for evaluating computer pointing devices. In: Proceedings of SIGCHI Conf. on Human Factors in Computing Systems (CHI). pp. 9–16.

Martín-Albo, D., Plamondon, R., Vidal, E., 2014. Training of on-line handwriting text recognizers with synthetic text generated using the kinematic theory of rapid human movements. In: Proceedings of Intl. Conf. on Frontiers in Handwriting Recognition (ICFHR). pp. 543–548.

Martín-Albo, D., Plamondon, R., Vidal, E., 2015. Improving sigma-lognormal parameter extraction. In: Proceedings of Intl. Conf. on Document Analysis and Recognition (ICDAR).

O'Reilly, C., Plamondon, R., 2009. Development of a Sigma-Lognormal representation for on-line signatures. Pattern Recogn. 42 (12), 3324–3337.

Plamondon, R., 1995a. A kinematic theory of rapid human movements. Part I: Movement representation and control. Biol. Cybern. 72 (4), 295–307.

Plamondon, R., 1995b. A kinematic theory of rapid human movements. Part II: Movement time and control. Biol. Cybern. 72 (4), 309–320.

Plamondon, R., 2015. Personal digital bodyguards for e-security, e-health and e-learning. Proceedings of Biennial Conf. of the Intl. Graphonomics Society (IGS), opening Keynote.

Plamondon, R., Alimi, A. M., Yergeau, P., Leclerc, F., 1993. Modelling velocity profiles of rapid movements: a comparative study. Biol. Cybern. 69 (1), 119–128.

Plamondon, R., Djioua, M., 2006. A multi-level representation paradigm for handwriting stroke generation. Hum. Mov. Sci. 25 (4–5), 586–607.

Plamondon, R., O'Reilly, C., Galbally, J., Almaksour, A., Anquetil, E., 2014. Recent developments in the study of rapid human movements with the Kinematic Theory: Applications to handwriting and signature synthesis. Pattern Recogn. Lett. 35, 225–235.

Plamondon, R., O'Reilly, C., Rémi, C., Duval, T., 2013. The lognormal handwriter: learning, performing, and declining. Front. Psychol. 4 (1), 945:1–945:14.

Schaal, S., Mohajerian, P., Ijspeert, A., 2007. Dynamics systems vs. optimal control–a unifying view. Prog. Brain. Res. 165, 425–45.

Schaal, S., Peters, J., Nakanishi, J., Ijspeert, A., 2004. Learning movement primitives. In: Intl. Symposium on Robotics Research (ISRR).

Schmidt, R. A., 1988. Motor control and learning: a behavioral emphasis, 5th Edition. Human Kinetics.

Simard, P., LeCun, Y., 1991. Reverse TDNN: an architecture for trajectory generation. In: NIPS. pp. 579–588.

Smithies, S., Novins, K., Arvo, J., 2001. Equation entry and editing via handwriting and gesture recognition. Behav. Inform. Technol. 20 (1), 53–67.

Thomassen, A. J. W. M., Keuss, P. J. G., van Galen, G., 1983. Motor aspects of handwriting. Acta Psychol. 54 (1–3), 354.

Trommershäuser, J., Maloney, L. T., Landy, M. S., 2003. Statistical decision theory and trade-offs in the control of motor response. Spat. Vis. 16 (3–4), 255–275.

Tu, H., Ren, X., Zhai, S., 2012. A comparative evaluation of finger and pen stroke gestures. In: Proceedings of SIGCHI Conf. on Human Factors in Computing Systems (CHI). pp. 1287–1296.

Vatavu, R.-D., Anthony, L., Wobbrock, J. O., 2012. Gestures as point clouds: a $P recognizer for user interface prototypes. In: Proceedings of Intl. Conf. on Multimodal Interaction (ICMI). pp. 273–280.

Vatavu, R.-D., Anthony, L., Wobbrock, J. O., 2013. Relative accuracy measures for stroke gestures. In: Proceedings of Intl. Conf. on Multimodal Interaction (ICMI). pp. 279–286.

Wobbrock, J. O., Wilson, A. D., Li, Y., 2007. Gestures without libraries, toolkits or training: A $1 recognizer for user interface prototypes. In: Proceedings of Annual ACM Symp. on User Interface Software and Technology (UIST). pp. 159–168.

Wright, C. E., 1990. Generalized motor programs: reexamining claims of effector independence in writing. In: Attention and Performance XIII: Motor representation and control. pp. 294–320.

Zhai, S., Kristensson, P. O., Appert, C., Anderson, T. H., Cao, X., 2012. Foundational issues in touch-surface stroke gesture design — an integrative review. In: Foundations and Trends in Human-Computer Interaction. Vol. 5. pp. 97–205.