Investigating the Dexterity of Multi-Finger Input for Mid-Air Text Entry

Srinath Sridhar[1] **Anna Maria Feit**[2] **Christian Theobalt**[1] **Antti Oulasvirta**[2]
[1]Max Planck Institute for Informatics
{ssridhar,theobalt}@mpi-inf.mpg.de
[2]Aalto University
{anna.feit,antti.oulasvirta}@aalto.fi

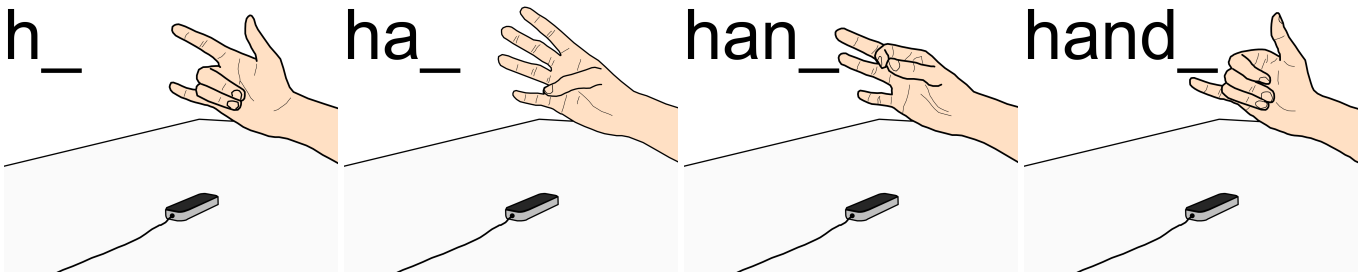

Figure 1. We investigate the dexterity of using multiple fingers for mid-air input. The paper reports performance and individuation characteristics of fingers and deploys them to the design of a mid-air text entry method using multi-objective optimization. Here we show an example of the word 'hand' being typed using one of our automatically obtained designs.

ABSTRACT
This paper investigates an emerging input method enabled by progress in hand tracking: input by free motion of fingers. The method is expressive, potentially fast, and usable across many settings as it does not insist on physical contact or visual feedback. Our goal is to inform the design of high-performance input methods by providing detailed analysis of the performance and anatomical characteristics of finger motion. We conducted an experiment using a commercially available sensor to report on the speed, accuracy, individuation, movement ranges, and individual differences of each finger. Findings show differences of up to 50% in movement times and provide indices quantifying the individuation of single fingers. We apply our findings to text entry by computational optimization of multi-finger gestures in mid-air. To this end, we define a novel objective function that considers performance, anatomical factors, and learnability. First investigations of one optimization case show entry rates of 22 words per minute (WPM). We conclude with a critical discussion of the limitations posed by human factors and performance characteristics of existing markerless hand trackers.

Author Keywords
Mid-air interaction; freehand input; Fitts' law; text entry

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
CHI 2015,, April 18 - 23, 2015, Seoul, Republic of Korea
Copyright is held by the owner/author(s). Publication rights licensed to ACM.
ACM 978-1-4503-3145-6/15/04...$15.00
http://dx.doi.org/10.1145/2702123.2702136

INTRODUCTION
This paper investigates an emerging category of input enabled by progress in computer vision-based hand tracking: *input by free motion of the hand involving any and all fingers*. Until recently, computer vision-based input was limited to gross movements of the arm and a few basic hand poses like pinching [3, 39]. However, recent methods can track full hand articulation using a single camera (*e.g.* [21, 27]). Leveraging the hand's capacity "directly" without intermediary devices like joysticks or buttons has always appealed to HCI researchers. With its many degrees of freedom, and fast and precise movements, the hand is the most dexterous of the extremities [12, 19]. Furthermore, freehand motion could provide an always-on input method, as only a camera is required. The method could alleviate the known input limitations of wearable or mobile devices.

Our goal is to inform the design of *high performance input* using multiple fingers in mid-air. High performance is decisive in activities like text entry, virtual reality, command selection, and gaming. However, previous work has focused on eliciting intuitive multi-finger gestures from users [23, 26]. This leaves out many issues, including performance characteristics of gestures involving single and multiple fingers simultaneously. To push the field forward, designers need to know some key factors affecting performance: How fast can users move their fingers? Can all fingers be moved independently and accurately? What are their movement ranges? How to combine fingers with different properties in one gesture?

Our work focuses on chord-like motions in mid-air as shown in Figure 1. These are easy-to-perform and familiar gestures, and among the few gesture categories that current computer vision sensors can reliably track. In this input gesture, there is no external target like a button (cf. most previous work on mid-air text entry [1, 20, 25, 35]). The involved fingers

are extended or flexed at a single joint to a discriminable end posture. Although this input method *can* be used with visual feedback, it allows for eyes-free input after memorization.

We extensively study the dexterity of single fingers in a target selection task. Users were asked to move a finger quickly and accurately between two *angular targets* (*e.g.* from a neutral resting position to the maximum position "down"). We assess each finger separately to report on three critical factors:

- **Speed and accuracy** of angular motions of fingers measured by Fitts' law models [17].
- **Individuation** of fingers, as measured by the so-called Schieber index [33]. It captures the extent to which non-instructed fingers remain still when a finger is moved.
- **Comfortable motion ranges** of fingers reported by users.

The results afford several insights. First, we report performance characteristics of each finger. The data show differences of up to 50% in movement times. Second, we asked users to move fingers *comfortably* and report on their motion ranges when using computer vision tracking. Third, to our knowledge, this is the first paper to report individuation indices for joints in HCI. For the middle and ring finger, coactivation can be so high that input may be compromised by false activations. In contrast, coactivation of other fingers while moving the thumb is virtually non-existent. We argue that individuation is a critical consideration in multi-finger input in mid-air which lacks physical resistance.

Our second contribution is to propose how to use this data in the design of high-throughput gesture sets. While our study considered only single joints, we attempt to apply our findings in the design of *multi-finger* input. The approach builds on literature in motor learning and assumes that multi-finger performance is limited by the *slowest* joint [13, 32]. Moreover, we exploit the fact that individuation constraints do not apply if co-dependent fingers participate together in a gesture. The benefit of these two assumptions is that the derivation of models to inform hand gestures is significantly less expensive than a study that tried to look at *all* combinations of fingers. Even with only three discretization levels per joint such an approach would have to cover roughly 10^{10} gestures. Finally, we use our findings to construct a proof-of-concept objective function called PALM to optimize text entry in mid-air. PALM considers performance (P), anatomical comfort (A: *i.e.* individuation), learnability (L), and mnemonics (M) to optimize multi-finger gestures. First investigations of a text entry method optimized for one-handed input show entry rates of 22 WPM. However, we note that users' performance was limited by brief training times, individuation constraints, and relatively limited performance of the tracker.

To summarize, this paper informs the design of high-performance input methods in mid-air by

1. providing ready-to-use models and look-up tables on performance, individuation and movement ranges of fingers, and
2. showing the applicability of the results by proposing an extension to multi-joint gestures and exploring its use in the multi-objective optimization of mid-air text entry methods.

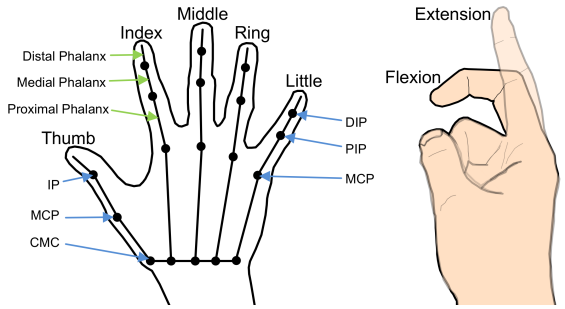

Figure 2. Left: Aspects of human hand anatomy with bones (green) and joints (blue). Right: We focus on *flexion-extension* of the five fingers.

BACKGROUND: CHARACTERISTICS OF FINGER MOTION

Our investigation of multi-finger input is informed by hand anatomy, the degrees of freedom of its joints, the performance of finger motion, and the limitations posed by dependencies on finger movement.

The Kinematic Skeleton

The skeleton of the human hand has 27 bones, the interfaces of which form the wrist and finger joints [12, 32] (Figure 2 shows a simplified skeleton). Together, this results in more than 25 degrees of freedom (DOFs) for the hand. In this paper, we focus on a subset of these DOFs. Finger movement (*flexion–extension* and *abduction–adduction*) is controlled by extrinsic muscles in the arm and intrinsic muscles in the hand. All joints except the Metacarpophalangeal (MCP—Index, Middle, Ring, and Little), and carpometacarpal (CMC—Thumb) have one DOF each. As humans we can describe hand gestures with terms like *thumbs up* or *v sign*. However, a formal representation is needed for study and use in computer vision-based input. We use a *kinematic skeleton* [24] to parametrize gestures. The hand skeleton configuration $\mathbf{\Theta}$ can be specified by angles of the joints connecting the bones, *i.e.* $\mathbf{\Theta} = [\theta_1, \theta_2, \ldots, \theta_i]^{\mathrm{T}}, \theta \in \mathcal{R}$.

Movement Performance

Finger movement performance can be quantified by movement time MT which is the time it takes for an end-effector to reach a target from a given distance. Fitts' law has been highly successful for predicting MT with traditional input devices [17]. It estimates the upper bound of pointing performance achievable after practice. Given a target of width W and distance D, Fitts' law states that the MT to reach the target is given by $MT = a + b \log_2(D/W + 1)$. Fitts' law has also been used previously to quantify performance differences in fingers, wrist, and forearm [4, 7, 15, 18, 28]. However, in this work, we use *angular* motions at joints instead of translation [14]. Considering the angular target width α_W and distance β_D, we get:

$$mt_\theta = a_\theta + b_\theta \log_2 \left(\frac{\alpha_D}{\beta_W} + 1 \right). \quad (1)$$

To acquire the a and b parameters, we conduct an experiment that employs a unidimensional pointing task. We address speed–accuracy trade-off in this task by using *effective* width $\bar{W}$ and distance $\bar{D}$.

Inter-Finger Dependencies

Movements of the hand act over multiple joints which makes coactivation of non-contributing joints common [12]. For example, many people cannot move their ring finger without coactivated movement of the little finger. More generally, coactivation is known to be larger among the metacarpophalangeal and the proximal interphalangeal joints [12, 33]. Hand gestures should minimize the extent of *unintended coactivation* of non-instructed fingers. Coactivations can be hard to inhibit and can cause recognition errors.

Schieber [33] proposed an *index of individuation* that indicates how independently an instructed finger can be moved from all others. The index was modeled for monkeys and humans [10]. A fully independent finger does not involve coactivation of other fingers during its activation, or vice versa. The individuation index is widely known in neuroscience, but largely disregarded in HCI. In order to compute it for every finger, the position of the non-instructed digit is plotted as a function of the instructed digit's position. The resulting trajectories are typically linear and the slope of a line fitted to these data points serves as a measure for the *relative coactivation*: the extent to which a non-instructed finger moves relative to the instructed finger. Given the coactivation C_{ij} of finger i during the movement of finger j, the individuation index of j is

$$I_j = 1 - [(\sum_{i=1}^{n} \mid C_{ij} \mid -1)/(n-1)], \qquad (2)$$

where $n = 5$ is the number of fingers. $I_j = 1$ indicates perfectly individuated movement, and $I_j = 0$ if all non-instructed fingers move simultaneously with j. The original study of individuation was reported for fingers, but it can be extended to multiple *joints* used in multi-finger input.

EXPERIMENT: FINGER DEXTERITY

Our experimental method is based on the reciprocal selection task used in Fitts' law studies [17]. As shown in Figure 3, users move a finger between two targets. Instead of extrinsic targets (*e.g.* buttons), the target here is a joint angle. Visual feedback is provided on a monitor with high refresh rate. In contrast to most Fitts' law studies, we track not only the endpoints of movements but the full motion of the hand. This allows us to quantify three aspects of the dexterity of finger motion: performance (speed and accuracy), individuation (unwanted motion of non-instructed fingers), and comfortable motion ranges. In addition, the data allow us to look at the range of individual differences.

We chose to focus on six joints spanning seven degrees of freedom (see Figure 3). This selection is motivated by the capabilities of present-day trackers and our pursuit of studying joints that could be a "class" of input motions. We conducted a pilot study of the Leap Motion sensor[1] and learned that individuated motions of interphalangeal joints are not well tracked, except for the thumb. Therefore, we decided to focus on the flexion/extension of the MCP joints of the fingers and the CMC joint of the thumb, which intuitively correspond

[1] `https://www.leapmotion.com/`

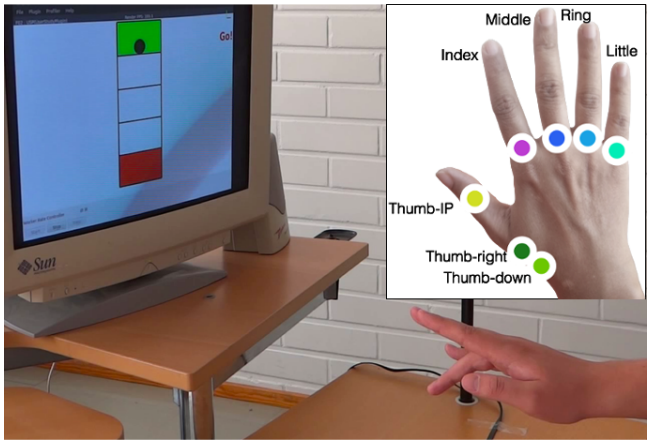

Figure 3. The experiment investigates the dexterity of six joints that can be reliably tracked with the Leap Motion sensor. The user is asked to move a finger between two target angles indicated on a display. Full hand motion was tracked. The color coding for joints is used in the Results section. Note that the CMC joint of the thumb is a special case, as it can be independently moved in two directions.

to "up" and "down" movements when the hand is in a neutral pose. Moreover, we included the IP joint of the thumb which was the only interphalangeal joint that could be moved *and* tracked well. Figure 3 also shows our naming convention and color coding used in the rest of the paper. For the thumb we use Thumb-Down and Thumb-Right to denote "up-down" and "left-right" movement of the CMC joint.

Participants

The study was conducted with 13 participants (8 male and 5 female) at two different locations. All participants were right-handed and had an age ranging from 22 to 32 (mean 27). Due to technical issues, one of the participants completed only 4 of the 7 joint conditions. The experiment took 1.5–2 hours per participant. Participants from one location were compensated with cinema vouchers. The trials were carried out under controlled lighting conditions with no distractions.

Experimental Design

The experiment followed a 7×4 within-subjects design with 7 DOFs and 4 index of difficulty (ID) conditions. To minimize order effects, the DOFs and ID conditions were randomized for each participant. Pre-trial practice was employed and breaks were provided after the trial for each joint.

Task, Materials, and Procedure

The task was a unidimensional target selection task. Participants had to move a pointer up and down between two targets on a screen and were instructed to move as fast and accurately as possible without moving non-instructed fingers too much. Control occurred by angular motions of joints that were linearly mapped to a pointer on the display. A trial would start from a comfortable neutral pose. The target region turned green when the pointer reached it and the user had to change direction to select the previous target again. In each condition, users had to perform 50 repetitions. Auditory feedback was given in the form of a low-frequency click. Throughout, participants placed their hand in a horizontal position over the sensor with their arm resting on a support.

Because of anatomical differences, we determined the movement range of each user experimentally, and used it to determine concrete target widths and distances for each user. Therefore, we first recorded the user-specific angular limits of each joint at the beginning of each task. We asked the participants to flex and extend the joint without moving the other fingers too much. The corresponding movement range was then uniformly divided into 2, 3, 4, and 5 bins. This gave us the same four unique IDs for every user: 1, 1.6, 2 and 2.3. Over all discretization levels there were 10 different target pairs for each joint, resulting in $7 \times 10 = 70$ conditions.

Apparatus

The joint angles were tracked using the Leap Motion by transforming its output to a kinematic skeleton. The software for tracking and display of the task ran on a fast desktop computer (3.1 GHz Intel i7 at one place, 3.1 GHz Intel i5 at the other). We showed visual feedback on high refresh rate monitors (112 Hz CRT and 120 Hz LCD respectively) and the Leap Motion was capable of tracking at up to 100 Hz.

Analysis

Performance: The design and evaluation of the Fitts' law task was done according to [36]. Movements with a movement time or distance beyond 3 SD of the median were excluded. Accuracy was adjusted to allow an error rate of 5%, a rate common in high-performance tasks such as text entry. Based on the remaining movements, we determined the *effective* target width $\bar{W}_{5\%}$ and distance $\bar{D}_{5\%}$ which was used to compute the effective index of difficulty (ID_e) of each task: $ID_e = log_2(\frac{\bar{W}_{5\%}}{\bar{D}_{5\%}} + 1)$. This indicates the actual difficulty of the performed task and captures the speed–accuracy trade-off. To account for individual differences, we cluster the effective IDs into 5 equally sized bins and compute the average movement time within each bin. For this purpose, we excluded data points with an effective ID of 3 SD beyond the median. Least-squares linear regression was then used to determine the slope and intercept of the Fitts' law model.

Individuation: We followed the protocol described in [33] to determine individuation indices. We first plotted, separately for each user, the normalized angle of every *non-instructed* joint as a function of the normalized angle of an instructed joint. The resulting 500 trajectories were then averaged by taking the median. Outliers beyond 3 SD of the median were excluded. The slopes of the resulting data were determined by least-squares linear regression. While linear movement trajectories were the norm, there were a few outliers where a linear relationship could not be determined. We observed two reasons: (1) Problems in tracking the joint angle (Figure 5 (b)) and (2) drifting of fingers, a phenomenon in which the non-instructed joint gradually changes its angle due to fatigue, inattention, or corrective behavior (Figure 5 (c)). To account for this, we excluded models with a fit of $R^2 < 0.5$. As suggested by Schieber, we averaged the *absolute* value for each slope, to generalize the relative individuation over all participants. These values were then used to compute the individuation index. In the next section, we report findings for performance, individuation, and movement ranges.

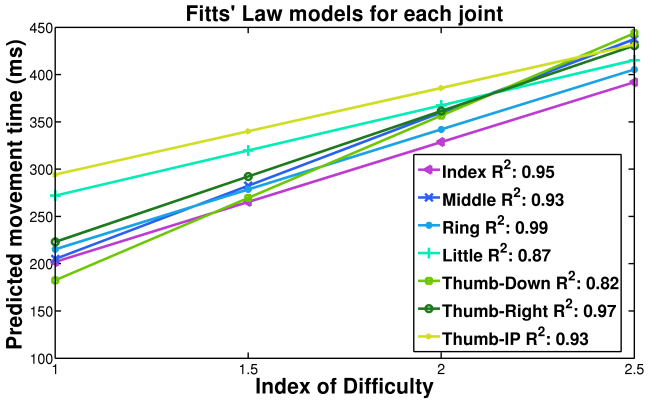

Figure 4. Performance models for each joint as given by Fitts' law. Overall, Index is the fastest, while Thumb and Little finger are the slowest.

Joint	Intercept *a*	Slope *b*	R^2
Index	75.140	126.77	0.95
Middle	49.940	155.03	0.93
Ring	88.450	126.79	0.99
Little	176.52	95.510	0.87
Thumb-Down	8.1900	174.26	0.82
Thumb-Right	84.590	138.44	0.97
Thumb-IP	202.73	91.590	0.93

Table 1. Fitts' Law models for each joint, given by intercept and slope.

RESULTS

Performance: Fitts' Law Models

Fitts' law models and fitness scores for the joints are given in Table 1. The R^2 values range from high (0.82) to excellent (0.99). One-way repeated measures ANOVA showed statistically significant differences among the joints for MT: $F(6, 60) = 3.3$, $p < 0.05$. Overall, Index had the highest performance, while Thumb-IP was the worst.

More subtle differences can be observed by looking at the cross-over points of the slopes in Figure 4. The Index finger was the fastest for most part of the ID range. However, for small IDs, corresponding to large neighboring targets, Thumb-Down outperformed Index. We also observe that for small IDs, MTs are spread for the different fingers (difference of 112 ms, $ID = 1$) while they become more condensed for larger IDs (51 ms, $ID = 2.5$). In other words, there is more variation for "easy" movements.

Significant individual differences could be observed. Differences in MT for the same joint were as large as 418 ms. The top performance was 91 ms for $ID = 1$, while the worst user performed at a speed of 509 ms per movement $ID = 1$.

Individuation: Schieber Indices

Table 2 provides an overview of the findings. We report aggregate indices per finger and by finger-pair coactivation.

Individuation Index: The individuation index for each finger can be found in the second column of Table 2. The values range from 1 for perfect individuation to 0 for perfect coactivation. Thumb-IP was found to be the most individuated joint, while Thumb-Down seemed to be the one with the highest coactivation. The individuation indices of the MCP joints showed only marginal differences.

Instructed Joint	Index of Individuation	Relative Coactivation						
		Index	Middle	Ring	Little	Thumb-Down	Thumb-Right	Thumb-IP
Index	**0.819**	1	0.24	0.20	0.19	0.29	0.11	0.06
Middle	**0.817**	0.16	1	0.41	0.14	0.20	0.11	0.07
Ring	**0.808**	0.16	0.20	1	0.36	0.15	0.22	0.06
Little	**0.806**	0.18	0.35	0.29	1	0.14	0.12	0.08
Thumb-Down	**0.792**	0.12	0.12	0.10	0.08	1	0.69	0.14
Thumb-Right	**0.853**	0.07	0.09	0.10	0.09	0.27	1	0.26
Thumb-IP	**0.889**	0.11	0.13	0.11	0.09	0.12	0.12	1

Table 2. Individuation index and relative coactivation describe the involuntary motion of joints. The individuation index is an aggregate that describes the independence of a finger when averaged over all other fingers (1 = perfect individuation). Relative coactivation denotes the movement of an non-instructed joint when the instructed joint (each row) is moving. A value of 1 denotes that the two joints always move together.

Joint	Min° (SD)	Max° (SD)	Range (SD)
Index	48.39 (12.25)	−21.19 (8.70)	69.58 (11.81)
Middle	37.58 (11.95)	−18.69 (8.02)	56.27 (12.54)
Ring	44.66 (8.320)	−12.24 (7.70)	58.90 (11.46)
Little	39.47 (15.78)	−20.81 (8.64)	60.28 (14.89)
Thumb-Down	27.31 (1.680)	−6.280 (6.54)	33.58 (7.130)
Thumb-Right	22.18 (10.53)	−11.99 (8.43)	31.32 (12.59)
Thumb-IP	62.97 (12.94)	−27.41 (4.37)	90.38 (13.93)

Table 3. Angular limits and movement range of each joint. The table shows values averaged over all users together with standard deviations.

Relative Coactivation: While the individuation index provides an elegant way to summarize the independence of each finger, greater insight is provided by the *relative coactivation* of joints, which denotes the movement of an non-instructed finger when the instructed finger is moved. In Table 2, we present the relative coactivation averaged over all users. It ranges from 0 to 1, where 1 is perfect coactivation, *i.e.* the non-instructed finger moves exactly along with the instructed finger. Note that the value range is the opposite to the individuation index, where 1 is better. We observe that Thumb-Down is closely correlated with Thumb-Right, explaining why it has the lowest individuation index. This indicates that the two DOFs of the thumb's CMC joint cannot be reliably distinguished and should be combined when implementing thumb movements for gestural input. Particularly high values were also observed for the movement of Ring during instructed movement of Middle, and the other way around (Figure 6). Thumb-IP shows low values throughout all joints which explains the good individuation index.

Comfortable Movement Ranges

The average angular limits and movement range for each joint are given in Table 3. The values represent joint limits that are comfortable for the user in this setting and reachable without moving the other joints too much. One-way repeated measures ANOVA (subjects with missing data excluded) showed statistically significant differences between movement ranges: $F(6, 60) = 39.19$, $p < 0.0001$. We observe that the CMC joint of the thumb has the smallest movement range in both movement directions (34°and 31°). The range of the MCP joints is twice that, and Index has the largest range (70°). Thumb-IP has overall the largest movement range with an average of 90°.

Observations on Individual Differences

Large differences among users were observed. Some users were able to keep their non-instructed finger nearly static (slope close to 0), while others moved them to a large extent along with the instructed joint (slope $= 0.4$). Figure 7 shows the coactivation of Index relative to Middle. Movement strategies vary too, resulting in a positive slope (moving along with the instructed joint) or even a negative slope (moving opposite to the instructed joint). If a joint could not be kept static, users either moved it along with the instructed joint or opposite to it. Attempts at "counteracting" movement like this were also observed in the original work by Schieber [33]. It may represent a strategy for preventing non-instructed fingers from moving along instructed digits. This suggests that these strategies are applied unconsciously.

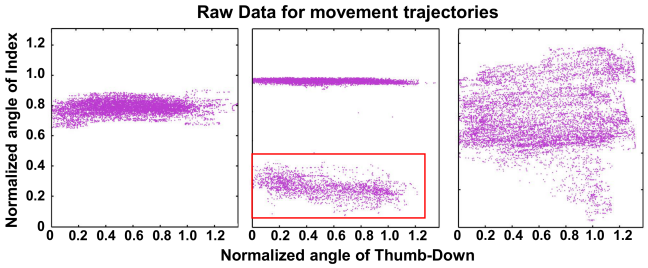

Figure 5. Raw data for movement of Index relative to instructed movement of Thumb-Down. Left (a): Example of high individuation, Middle (b): Tracking errors (red box), and Right (c): "drifting finger".

We also observed what we denote as the *drifting finger effect*: the position of non-instructed fingers may change gradually over time for some users, as they "forget" to keep the finger still. For some users, this poses no problem, they are able to produce the exact same movement over and over (Figure 5 (a)). We show raw data of this "drifting finger" problem in Figure 5 (c). Due to user-specific differences like this, the linear model of Schieber does not always fit to a user's motion. On average, an R^2 of 0.77 (SD 0.14) was found, ranging from 0.5 to excellent fits of 0.99. As discussed above, we excluded the data where no sufficient linear relationship could be found. On average, this amounted to excluding data from 4 users per joint-joint condition.

Finally, despite our efforts to ensure the ergonomics of the posture and to provide enough breaks, some users complained about fatigue, especially with their wrist or arm getting tired. This suggests that these motions are tiring even if they do not require the use of large forces.

APPLICATION TO TEXT ENTRY

The results of the study offer a nuanced picture of the two characteristics of finger motions. The performance and independence of fingers differ and are inter-connected in sub-

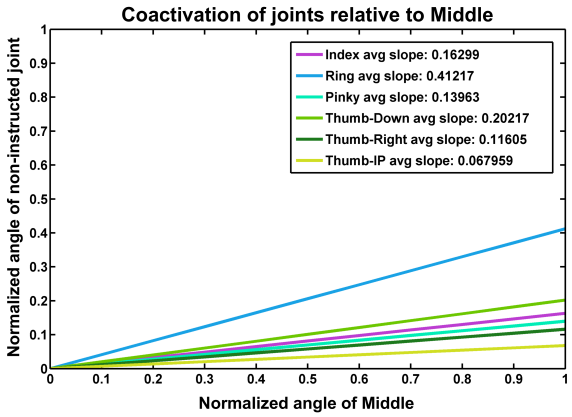

Figure 6. Average coactivation of all joints relative to the instructed movement of the middle finger. The slopes are the average of the absolute values over all users.

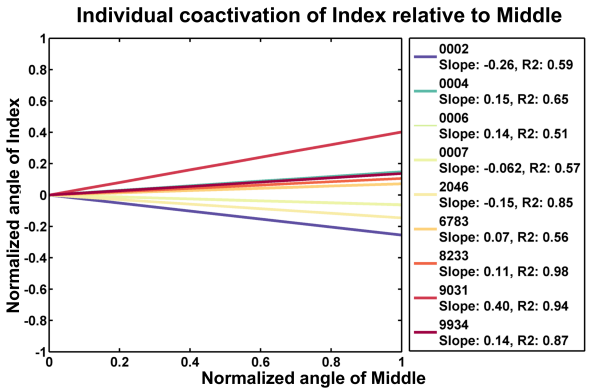

Figure 7. Differences among users (denoted by four digit user ID) in the movement of the index finger relative to the middle finger. A positive slope indicates that it follows the instructed joint, negative slope that it moves in the opposite direction.

tle ways. In this section, we present a proof-of-concept that shows how to use the results to design multi-finger gestures for a high-performance input task. We chose to focus on text entry by mapping *static mid-air hand postures* to letters. We use the terms 'gesture' and 'posture' interchangeably in this section to denote static postures. Mid-air input is a promising input modality for emerging devices like smartwatches and heads-up displays [20]. In contrast to previous mid-air text entry methods which used *extrinsic* key targets or handwriting gestures [1, 20, 25, 35], we focus on chord-like gestures controlled by angular motions. Although more complex than single finger input, it has been shown that a large number of chords can be memorized [34] and used for text entry (*e.g.* [9, 16]), as well as on multitouch displays [2].

Since the space of possible posture-letter mappings is (exponentially) large, we follow an optimization approach (*e.g.* [8, 40]). We outline a novel objective function called PALM that optimizes mappings for four objectives. In addition to performance and individuation constraints, it considers learnability and mnemonics. The outcomes can be used to enter text with any hand tracker and gesture recognizer. Our approach has four main steps, which serve as a roadmap for designing tasks other than text entry: (1) Discretizing Joint Angles, (2) Generalizing to Multi-Joint Gestures, (3) Formulating an Objective Function, and (4) Optimization.

Step 1: Discretizing Joint Angles

We first need to select the number of discretization levels of angular motion that each joint can afford. This is determined by the robustness of the hand tracker and by performance data we obtained. Our estimate for angular discretization when using the Leap Motion is between 2 and 5 levels per joint angle. For each joint, an integer from 0–k is used to represent the current joint angle, where k is the highest level. Thus, the posture of the hand can be compactly represented using a string of numbers which we call a *bin address*. For instance, the posture corresponding to the letter 'h' in Figure 1 can be denoted by the string [0,0,1,1,0] (using 5 joints). We also define a neutral pose for the hand, which is a comfortable position, and calibrate such that it corresponds to the bin address [0,0,0,0,0].

Step 2: Generalizing to Multi-Joint Gestures

Since the findings from our study are for single joints, we make two assumptions to generalize to multi-joint gestures. First, to estimate movement time (MT) for gestures involving multiple joints, we assume that it is bounded by the performance of the slowest contributing joint. We base this on evidence that movement of arm joints are timed so that all joints reach their final positions simultaneously [13, 32]. Thus, we estimate the time for a multi-joint gesture as the maximum over each of the MTs of all joints involved. Formally, we define time for moving from one posture to another as,

$$MT = \max\{mt_{\theta_i}\}, \theta_i \in \Theta, \tag{3}$$

where mt_θ corresponds to the movement time of one joint as given in Equation (1).

Second, to estimate individuation constraints of a multi-finger gesture, we extend the individuation index of Schieber to take into account the fact that coactivation between fingers is not an issue when those fingers are used in the same gesture. The middle finger, for example, has a poor individuation index, which is mainly dominated by the relative coactivation of the ring finger. A gesture involving both fingers can therefore be performed with higher individuation than a gesture involving only one of the fingers. To this end, we define the coactivation C_{iG} of a joint i relative to a gesture (or posture) G as the maximal coactivation of i relative to any joint j involved in the gesture: $C_{iG} = \max_{j \in G} C_{ij}$. Then, following the original Equation (2), we compute the individuation index for any multi-joint gesture as

$$I_G = 1 - [(\sum_{i=1}^{n} |C_{iG}| - |G|)/n - |G|], \tag{4}$$

where $|G|$ denotes the number of actively involved joints, and n is the total number of joints.

Step 3: Objective Function Formulation

Our design task is to maximize the *usability* U of a letter assignment, *i.e.* the mapping of each character in a character set to a unique posture (gesture) of the hand. To characterize U, we formulate a multi-term objective function for mid-air text entry called PALM which addresses

four factors affecting mid-air text entry with multiple fingers: Performance, Anatomical comfort (individuation), Learnability, and Mnemonics. In addition to performance and individuation, we formalize learnability and mnemonics based on existing literature.

Usability U is thus defined as a weighted sum of four normalized (*i.e.* $\in [0,1]$) terms[2]. Formally, we write our usability objective as

$$U = w_p \hat{P} + w_a \hat{A} + w_l \hat{L} + w_m \hat{M}, \quad (5)$$

where the positive weights w_p, w_a, w_l, and w_m set by a designer sum up to 1. The remaining terms in the objective function are described below in turn.

Performance Term (P)

Our performance score P is measured in words per minute (WPM). Following previous work on keyboard optimization [8, 40], we use Fitts' law models to predict the time $mt_{k\ell}$ to articulate a joint from letter k to letter ℓ by computing the movement time as described in Equation (3).

We then compute WPM with 5 % error rate as:

$$P = 60/(\sum_k \sum_\ell f_{k\ell} mt_{k\ell}) \times 5, \quad (6)$$

where $f_{k\ell}$ is the frequency of bigram $k\ell$.

Anatomical Comfort Term (A)

For each gesture, we use Equation (4) to estimate how well it individuates. An index of 1 corresponds to perfect individuation where none of the non-instructed joints moves along with the joints involved in the gesture, a value of 0 would mean that all fingers move to the same extent, even if they are not part of the gesture. Thus, $\hat{A}$ takes the value of the individuation index.

Learnability Term (L)

Learnability is an important factor to consider for any activity involving rapid and careful articulation of multiple joints. To develop a score for learnability of a gesture, we build on some prevalent theories of motor learning that view learning as a *hierarchical combination of primitives* [22]. According to this view, the brain simplifies multi-dimensional motor control by collapsing it into a few dimensions. Practicing a complex gesture gradually increases hierarchical organization and decreases reliance on feedback. This has two consequences. First, the fewer DOFs a gesture involves, the easier it will be to learn. For instance, gesturing with one finger is easier to learn than a gesture using three fingers. We name the number of involved DOFs u_{dofs}. Second, if the involved digits involve the same *end posture*, it will be easier to learn because the articulations can be represented with a single learning primitive. For example, it is easier to extend all digits by 40° than to extend some by 20° and others by 40°. We denote the number of DOFs for which a target angle is defined in a gesture by u_{targets}. Our learnability score combines these two aspects:

$$L = 1 - \sum_k (0.5\, \hat{u}_{\text{targets, k}} + 0.5\, \hat{u}_{\text{dofs, k}}). \quad (7)$$

[2]Normalized variables are marked with a hat.

Mnemonics Term (M)

Studies of human memory suggest that categorization, chunking, and mnenomics help forming more durable long-term memory traces among otherwise unrelated materials [38]. Our mnemonics score M considers the memorability of a letter assignment *as a whole*. We call a *mnemonic set* a set of similar gestures, such as gestures that all have a neighboring finger. To identify finger mnemonics, we build on a recent study of multi-finger chord gestures that showed a positive effect on learning [38]. We take the mnemonic principles presented there and extend them from three fingers to five. In particular, we include the following mnemonics rules: neighboring fingers (*e.g.* thumb and little finger together), base (*e.g.* thumb or index with other fingers), and single finger.

The M-score considers two aspects: (1) the proportion of gestures belonging to a mnemonic set m_{coverage} and (2) how *few* mnemonic sets are required m_{sets}, which is the inverse of the proportion of all mnemonic sets being in use. We define $M = 0.5\,(m_{\text{coverage}} + m_{\text{sets}})$. M thus rewards designs where a large proportion of gestures belong to a few mnemonics sets. While our learnability score L looks at motor learning "from scratch", this score focuses on the benefit of the set consisting of easily recognizable gestures.

Step 4: Optimization

To optimize the multi-term objective function we use techniques from multi-dimensional Pareto optimization [29]. Instead of searching for a global optimum in a single run, we use a multi-start local search method. Local search starts from a random position in the search space and randomly samples its neighborhood. When search converges, we store the incumbent to a file and restart search. A similar approach was used in a previous paper addressing a multi-objective task [8]. Our implementation reaches reasonable designs in minutes while good ones take about one day on a cluster computer.

DESIGN CASES

This section presents mappings optimized for fast performance, learnability, as well as for different character sets. Apart from this, we present solutions with multiple discretization levels for the joint angles. This demonstrates how the approach can be used across varying design interests. Finally, we present a preliminary evaluation of one of our designs.

Before discussing the designs, we report our experiences regarding the value of optimizing for all four objectives of PALM. To learn if performance and individuation are compatible design goals, we optimized for P, A, and P+A goals separately. The results showed that the benefit of optimizing for only one of the goals is negligible. In other words, performance and individuation may not always be competitive goals for design. The P-only design has fewer multi-joint gestures, whereas both A-only and P+A have more gestures involving neighboring fingers. This encouraged further exploration of the multi-objective design space.

Table 5 lists all outcomes along with two alternative text entry methods: Engelbart's chording keyboard [9] and a fingerspelling method (American Sign Language). Words per minute is predicted considering expert motor performance

Bin Address	Character	Bin Address	Character
0,1,0,0,0	_	1,1,0,0,0	n
1,0,0,0,0	a	1,0,0,1,0	o
0,0,1,0,1	b	0,0,0,1,1	p
1,1,0,1,0	c	0,1,1,1,1	q
0,1,1,1,0	d	0,1,0,1,0	r
0,0,0,1,0	e	0,1,1,0,0	s
1,1,1,1,0	f	0,0,1,0,0	t
0,1,0,0,1	g	0,0,0,0,1	u
0,0,1,1,0	h	1,0,0,1,1	v
1,0,1,0,0	i	1,0,0,0,1	w
0,1,1,0,1	j	0,0,1,1,1	x
1,1,0,0,1	k	0,1,0,1,1	y
1,1,1,0,0	l	1,0,1,0,1	z
1,0,1,1,0	m		

Table 4. FASTTYPE was optimized favoring Performance. The bin addresses describe each gesture, see text for explanation. Observe how commonly occurring letters like 'a' are assigned to easy postures such as flexing the thumb.

only, using Equation (6). Due to space limitations we report the full mapping only for FASTTYPE in Table 4. Please see the supplementary material for the remaining mappings.

Standard Character Sets: NUMPAD is a solution that maps the numbers from 0–9 to postures formed by the 5 joints, one per finger. Each joint angle is discretized into 2 levels. The predicted performance for this mapping is the highest at 113.0 WPM due to the small character set. FASTTYPE is a solution with the letters *a–z* (including space), and 5 joints each with 2 discretization levels. This mapping was optimized for typing speed and uses chord-like movements with a predicted performance of 54.7 WPM. We show this mapping in Table 4. In the table, we use the concept of bin address as explained earlier. The joints are ordered from Thumb to Little. For example, [0,0,0,0,1] would mean flexing Little but keeping the rest in a neutral pose. BALANCETYPE, a variant with balanced weights for the four objective function weights had a predicted performance of 50.1 WPM.

Extended Character Sets: FULLTYPE is optimized to map all letters of the alphabet, numbers, and special characters for a total of 48 characters. The predicted performance was 50.7 WPM with 5 joints and 5 discretization levels per joint. While this mapping has a good predicted performance, we hypothesize that it is hard to perform because of 5 discretization levels for joint angles. Finally, THREETYPE optimizes a full keyboard to the three fingers with the highest individuations: Thumb, Index, Middle. It, too, assumes 5 discretization levels which is presently impossible with our tracker and would require a long time to learn.

We also represented fingerspelling in American Sign Language using our bin address notation. For the represented mapping, our objective function predicts an entry rate of 43.9 WPM which is surprisingly close to the empirically observed rate of 40–45 WPM for experienced practitioners [30].

First Observations on User Performance: FASTTYPE

In order to estimate if the predicted performance is indeed achievable with mid-air text entry, we conducted a preliminary evaluation of FASTTYPE with 10 users. We followed a word-level paradigm previously used by Zhai *et al.* [5]. Here, a randomly sampled word is practiced until performance peaks. The benefit of this is that the upper boundary of entry performance can be estimated even without having to learn the full gesture set.

Method: 10 right-handed participants took part in the experiment (9 male, 1 female; ages from 21 to 39, mean 26). The experiment took 1.5–2 hours and all participants were compensated. We randomly sampled 4–8 character strings from the Enron Email Dataset [37] for the stimulus. Each contained 1–2 frequently entered words and also included the space character. A task consisted of repeatedly entering a word. At the beginning, participants were allowed to practice the word by going through the gestures for all letters and exploring the fastest transitions between each gesture. As soon as they could memorize the mapping of the corresponding letters, the task started. The task was terminated by the experimenter when a performance plateau could be observed.

Prototype: We built a prototype that allowed users to enter text, and recorded performance of typed words. Our gesture recognizer used joint angle data from the Leap Motion, and used a combination of dwell times and signal peak detection to detect when users made a particular posture which was converted to text. A custom-built application displayed information to the user as well as recorded data for analysis. The hardware used was identical to the first experiment.

Result: Overall, the 10 users entered 53 words at an average peak performance of 22.25 WPM (SD 8.9). For analyzing the peak performance of each word, we extracted the top 3 repetitions with an error rate less than 15% (measured by Damerau-Levenshtein distance). Three words had to be excluded due to this restriction. The remaining words were typed with an average error rate of 2.3% (SD 0.04). A one-way ANOVA on WPMs showed a statistically significant difference among users: $F(9, 49) = 7.68$, $p < 0.001$. Average peak performances ranged from 13 WPM to 38.1 WPM. This large performance range clearly shows the influence of individual differences in performance, individuation and anatomical limitations found in our first experiment. While these results serve as a first exploration of PALM, further detailed studies are needed to validate the effectiveness of our model.

DISCUSSION

The results presented in this paper deepen the understanding of multi-finger input in mid-air. The findings show that multi-finger input has potential for high throughput. While it was known previously that differences existed in performance and individuation between fingers, they were not quantified in a setting that is representative of modern computer vision-based input. Our results were obtained by adapting the familiar methodology of Fitts' law studies along with a measurement of individuation adopted from motor control research. This is in contrast to existing work in gesture design that has considered elicitation methods to learn about user preferences, intuitiveness, and social acceptability [23, 26, 31].

In a proof-of-concept, we demonstrated the applicability of our results by computationally optimizing a mid-air text entry method. Based on prior work on motor performance [13, 32], we extended our findings from single fingers to multi-

Mapping	Character Set	Joint Discretization	Weights (PALM)	Objective values (PALM)	Predicted WPM
NUMPAD	0–9	2, 2, 2, 2, 2	0.30, 0.30, 0.05, 0.05	0.27, 0.03, 0.22, 0.22	113.0
FASTTYPE	*a*–*z*	2, 2, 2, 2, 2	0.50, 0.10, 0.10, 0.30	0.53, 0.03, 0.18, 0.50	54.7
BALANCETYPE	*a*–*z*	5, 5, 5, 4, 4	0.25, 0.25, 0.25, 0.25	0.42, 0.02, 0.19, 0.17	50.1
FULLTYPE	0–9, *a*–*z*	5, 5, 5, 5, 5	0.20, 0.20, 0.20, 0.20	0.41, 0.14, 0.19, 0.33	50.7
THREETYPE	*a*–*z*	5, 5, 4	0.40, 0.40, 0.20, 0.00	0.38, 0.01, 0.28, 0.00	65.1
Fingerspelling	*a*–*z*	4, 3, 3, 3, 3	0.25, 0.25, 0.25, 0.25	0.51, 0.02, 0.28, 0.80	43.9
Engelbart's Chord Kbd	*a*–*z*	2, 2, 2, 2, 2	0.25, 0.25, 0.25, 0.25	0.58, 0.03, 0.17, 0.69	49.0

Table 5. An overview of optimized mappings and predicted WPM. The bottom part shows predictions for two existing methods.

joint gestures. The P and A terms of PALM are based on the empirical results, whereas the L and M terms are derived from prior work on human memory and motor learning [22, 38]. While further evaluation is needed to prove the validity of these assumptions, we show how our findings can serve in the search for good solutions among millions of designs.

To analyze the outcomes, we built a prototype and explored the performance for one of the optimized mappings which showed an entry rate of 22 WPM. While the performance predicted by Equation (6) was surprisingly close to the observed performance in fingerspelling, FASTTYPE falls short of the predicted rate of 54.7 WPM. As Equation (6) only predicts expert motor performance, this can be partially attributed to the lack of training and limited tracker performance. However, further evaluation is needed to investigate learning over time and cognitive effort involved in mid-air input.

CONCLUSION AND FUTURE WORK

In this paper, we investigated the dexterity of fingers for mid-air input. The results provide insights into the performance of individual fingers and their coactivation. The findings suggest that mid-air input is a promising input modality, but there are limitations to the capacity of the human hand.

The physiology and cognitive skills of humans pose two critical constraints that future work should consider. First, the learnability of gestures is a pragmatic obstacle for multi-finger input. If a gesture set for text entry is prohibitively time consuming to learn it will affect large-scale adoption. With PALM, we propose a method to optimize for learnability. However, further evaluation is needed to investigate the influence of the L and M term on performance and learnability, and evaluate the involved models. Second, the effect of fatigue in multi-finger input is not fully understood yet. Users in both our studies reported discomfort in their arm and wrist.

The technological challenges of hand tracking without markers pose additional constraints to mid-air input. Despite much progress, markerless hand and finger tracking is still a challenging problem. We restricted our study to 6 joints since even commercial sensors like the Leap Motion could not reliably track certain finger joints. Our evaluation showed that users were limited in their speed by errors in tracking all joint angles under fast motion. We assume that some of these issues arise from assumptions about finger individuation used by the tracker.

A notable omission in our investigation is appropriate feedback for mid-air input. While no visual feedback is needed for our text entry method, it is unknown if proprioception alone suffices to perform fast and accurate mid-air gestures. As an alternative, tactile feedback was shown to improve performance on touch screens [11] and new technologies such as UltraHaptics [6] provide a way to bring non-contact haptic feedback to mid-air input.

This paper has contributed empirically derived models of performance factors involved in mid-air input and a proof-of-concept approach to design. Our optimizer allows finding designs that strike desirable trade-offs in this demanding design landscape. Although our final evaluation of mid-air text entry fell short of the performance predicted by our Fitts' law model, the result is promising and justifies further research. We believe that when the outstanding human and technological issues are solved, this category of input can achieve performance that is currently seen only for physical keyboards.

Acknowledgments: This research was funded by the ERC Starting Grant projects CapReal and COMPUTED (637991), and the Academy of Finland.

REFERENCES

1. Amma, C., Georgi, M., and Schultz, T. Airwriting: Hands-free mobile text input by spotting and continuous recognition of 3D-space handwriting with inertial sensors. In *Proc. ISWC* (2012), 52–59.
2. Bailly, G., Müller, J., and Lecolinet, E. Design and evaluation of finger-count interaction: Combining multitouch gestures and menus. *Int. J. Hum.-Comput. Stud.* (2012), 673–689.
3. Bailly, G., Müller, J., Rohs, M., Wigdor, D., and Kratz, S. ShoeSense: a new perspective on gestural interaction and wearable applications. In *Proc. CHI* (2012), 1239–1248.
4. Balakrishnan, R., and MacKenzie, I. S. Performance differences in the fingers, wrist, and forearm in computer input control. In *Proc. CHI* (1997), 303–310.
5. Bi, X., Smith, B. A., and Zhai, S. Multilingual touchscreen keyboard design and optimization. *Human–Computer Interaction 27*, 4 (2012), 352–382.
6. Carter, T., Seah, S. A., Long, B., Drinkwater, B., and Subramanian, S. UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proc. UIST* (2013), 505–514.
7. Crossan, A., Williamson, J., Brewster, S., and Murray-Smith, R. Wrist rotation for interaction in mobile contexts. In *Proc. MobileHCI* (2008), 435–438.

8. Dunlop, M., and Levine, J. Multidimensional Pareto optimization of touchscreen keyboards for speed, familiarity and improved spell checking. In *Proc. CHI* (2012), 2669–2678.
9. Engelbart, D. C., and English, W. K. A research center for augmenting human intellect. In *Proc. of Fall Joint Computer Conference* (1968), 395–410.
10. Häger-Ross, C., and Schieber, M. H. Quantifying the independence of human finger movements: comparisons of digits, hands, and movement frequencies. *J. Neuroscience 20*, 22 (2000), 8542–8550.
11. Hoggan, E., Brewster, S. A., and Johnston, J. Investigating the effectiveness of tactile feedback for mobile touchscreens. In *Proc. CHI* (2008), 1573–1582.
12. Jones, L. A., and Lederman, S. J. *Human Hand Function*, 1 ed. Oxford University Press, 2006.
13. Kaminski, T., and Gentile, A. Joint control strategies and hand trajectories in multijoint pointing movements. *Journal of Motor Behavior 18*, 3 (1986), 261–278.
14. Kondraske, G. An angular motion Fitt's law for human performance modeling and prediction. In *Proc. IEEE EMBS* (Nov. 1994), 307–308 vol.1.
15. Langolf, G. D., Chaffin, D. B., and Foulke, J. A. An investigation of Fitts' law using a wide range of movement amplitudes. *Journal of Motor Behavior 8*, 2 (1976), 113–128.
16. Lyons, K., Starner, T., and Gane, B. Experimental evaluations of the twiddler one-handed chording mobile keyboard. *Human-Computer Interaction* (2006).
17. MacKenzie, I. S. Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction 7*, 1 (1992), 91–139.
18. Malik, S., Ranjan, A., and Balakrishnan, R. Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In *Proc. UIST* (2005), 43–52.
19. Mao, Z.-H., Lee, H.-N., Sclabassi, R., and Sun, M. Information capacity of the thumb and the index finger in communication. *IEEE Trans. Biomedical Engineering 56*, 5 (May 2009), 1535 –1545.
20. Markussen, A., Jakobsen, M. R., and Hornbaek, K. Vulture: A mid-air word-gesture keyboard. In *Proc. CHI* (2014), 1073–1082.
21. Melax, S., Keselman, L., and Orsten, S. Dynamics based 3D skeletal hand tracking. In *Proc. i3D* (2013), 184–184.
22. Mitra, S., Amazeen, P. G., and Turvey, M. T. Intermediate motor learning as decreasing active (dynamical) degrees of freedom. *Human Movement Science 17*, 1 (1998), 17–65.
23. Morris, M. R., Danielescu, A., Drucker, S., Fisher, D., Lee, B., schraefel, m. c., and Wobbrock, J. O. Reducing legacy bias in gesture elicitation studies. *interactions 21*, 3 (2014), 40–45.
24. Murray, R. M., Li, Z., and Sastry, S. S. *A Mathematical Introduction to Robotic Manipulation*, 1st ed. CRC Press, Inc., 1994.
25. Ni, T., Bowman, D., and North, C. AirStroke: Bringing unistroke text entry to freehand gesture interfaces. In *Proc. CHI* (2011), 2473–2476.
26. Piumsomboon, T., Clark, A., Billinghurst, M., and Cockburn, A. User-defined gestures for augmented reality. In *INTERACT 2013*, no. 8118. Jan. 2013, 282–299.
27. Qian, C., Sun, X., Wei, Y., Tang, X., and Sun, J. Realtime and robust hand tracking from depth. In *Proc. CVPR* (2014).
28. Rahman, M., Gustafson, S., Irani, P., and Subramanian, S. Tilt techniques: investigating the dexterity of wrist-based input. In *Proc. CHI* (2009), 1943–1952.
29. Rao, S. S., and Rao, S. *Engineering optimization: theory and practice*. John Wiley & Sons, 2009.
30. Ricco, S., and Tomasi, C. Fingerspelling recognition through classification of letter-to-letter transitions. In *Proc. ACCV*. 2010, 214–225.
31. Rico, J., and Brewster, S. Usable gestures for mobile interfaces: evaluating social acceptability. In *Proc. CHI* (2010), 887–896.
32. Rosenbaum, D. A. *Human motor control*. Academic Press, 2009.
33. Schieber, M. H. Individuated finger movements of rhesus monkeys: a means of quantifying the independence of the digits. *J Neurophysiology 65*, 6 (1991), 1381–91.
34. Seibel, R. Data entry through chord, parallel entry devices. *Human Factors: The Journal of the Human Factors and Ergonomics Society 6*, 2 (1964), 189–192.
35. Shoemaker, G., Findlater, L., Dawson, J. Q., and Booth, K. S. Mid-air text input techniques for very large wall displays. In *Proc. GI*, Canadian Information Processing Society (2009), 231–238.
36. Soukoreff, R. W., and MacKenzie, I. S. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *Intl. J. of Human-Computer Studies 61*, 6 (Dec. 2004), 751–789.
37. Vertanen, K., and Kristensson, P. O. A versatile dataset for text entry evaluations based on genuine mobile emails. In *Proc. of MobileHCI* (2011), 295–298.
38. Wagner, J., Lecolinet, E., and Selker, T. Multi-finger chords for hand-held tablets: recognizable and memorable. In *Proc. CHI* (2014), 2883–2892.
39. Wilson, A. D. Robust computer vision-based detection of pinching for one and two-handed gesture input. In *Proc. UIST* (2006), 255–258.
40. Zhai, S., Hunter, M., and Smith, B. A. The Metropolis Keyboard - an exploration of quantitative techniques for virtual keyboard design. In *Proc. UIST* (2000), 119–128.