A Correlation Method for Handling Infrequent Data in Keystroke Biometric Systems

Steve Kim, Sung-Hyuk Cha, John V. Monaco, and Charles C. Tappert

Department of Computer Science, Pace University
1 Pace Plaza, New York, NY 10038, USA
skim@gmail.com, {scha, john.v.monaco, tappert}@pace.edu

ABSTRACT

Many applications need methods for handling missing or insufficient data. This paper applies a correlation technique to improve the fallback methods previously used to handle the paucity of keystroke data from the infrequently used keys in a keystroke biometric system. The proposed statistical fallback model uses a correlation based fallback table based on the linear correlation between pairs of keys. Two large long-text keystroke databases are used in the study – one to construct the model and the other to evaluate system performance as a function of sample length.

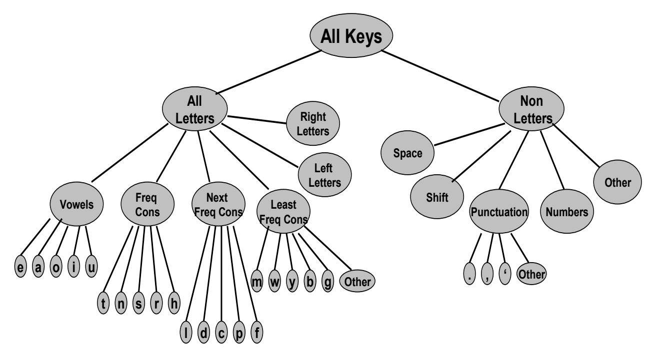
Index Terms— Behavioral biometrics, Correlation, Keystroke, Linear regression

1. INTRODUCTION

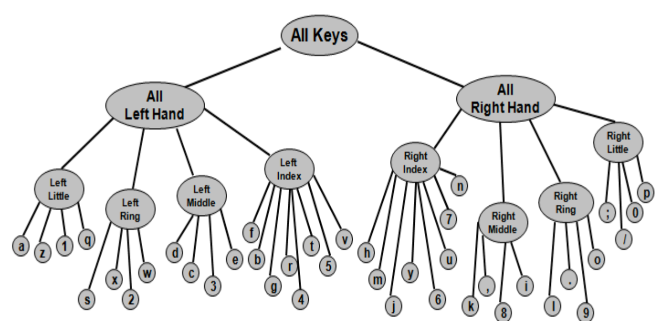
A number of applications need methods for handling missing or insufficient data. In speech and language processing, several methods of handling missing or sparse data are described in [2]. The *N*-gram model of missing or infrequent data is estimated based on the ($N-1$)-gram model of sufficient data recursively in the *back-off* [3] and *deleted interpolation* [4]. Although both models fail if the unigram is missing, this occurs rarely.

In this study, a correlation technique is proposed to handle insufficient keystroke data. The keystroke biometric is a behavioral biometric that has gained contemporary popularity as the keyboard provides a vital input device. Keystroke biometric systems utilize the *keystroke dynamics* as features or measurements [1]. The usual keystroke dynamic measurements are the *dwell* (key press duration) and *flight* (transition between two keys) times. In contrast to password input, which is fixed, long-text input samples can consist of several hundred keystrokes of varying frequency. Therefore, keystroke biometric systems operating on such input can use statistical features such as the mean and standard deviation of the dwell and flight times [6]. These measurements, however, suffer from poor estimates of those keys where the number of samples during an acquisition session is missing or insufficient.

Inspired from the language processing "back-off" models, two hierarchical fallback tree models were evaluated for use in a keystroke biometric system. These hierarchical tree models served two functions. First, they provided fallback to additional data when insufficient keystroke instances were available to compute the statistical features. Second, they provided a granularity of features, where the granularity increases from gross features at the top of the tree to fine features at the bottom. The first hierarchical model, called the '*linguistic*' model, organizes keys based primarily on frequency of use [5, 6]. The second model, called the '*touch-type*' model, groups keys based on the fingers used to strike keys by touch typists [6]. These linguistic and touch-type models are depicted in Figure 1 (a) and (b), respectively.



(a) Linguistic Hierarchy Model



(b) Touch-type Hierarchy Model
Figure 1: Two previous fallback models

An early analysis of the fallback aspect of the two models found the linguistic model to be slightly better than the touch-type model [6], although different features were used in the comparison since the models served two different functions. Compared to a default model of simply falling back to the top node of the hierarchy tree, the linguistic model reduced the error rate by 26% and 53%, respectively on two datasets, showing the utility of the hierarchy tree for the fallback function. These datasets contained samples of 500 or more keystrokes. Fallback never occurred more than one level up from the leaf nodes and most of the one-level-up nodes were never used (vowel, frequent consonant, all letters, non-letters) because their leaf nodes were sufficiently frequent to not require fallback.

In this study, the two functions of the hierarchy tree – *feature granularity* and *fallback* – are separated. The hierarchy trees are retained for feature granularity and a sounder statistical model, called the *correlation-based fallback table* model, is proposed for fallback.

Two large independent long-text keystroke databases are used in this study. The first database is used to construct the correlation-based fallback table model. The second is used to evaluate system performance as a function of sample length. The new correlation fallback model should show improvement over the earlier models, especially as the data becomes sparser (fewer keystrokes per sample).

The concept of correlation plays important roles in many aspects of pattern recognition [7]; it can be modeled as an ultimate goal to optimize while it can be a serious problem to mitigate. In keystroke biometric, the problematic side of correlation between feature variables was addressed in [8,9] to justify their choice of *Mahalanobis* distance over *Euclidean* distance. Here we focus on the useful side of correlation for estimating better feature values of sparse keystroke dwell data.

The rest of the paper is organized as follows. In section 2, correlations between key dwell values are studied and the three correlation parameters are discovered from a large representative keystroke database. Section 3 presents the algorithmic use of correlation between key dwells for the keystroke biometric. Finally, section 4 concludes this work.

2. CORRELATIONS OF KEY DWELLS

The i^{th} keystroke of a user is denoted as A_i and consists of three tuples: key value, pressed time, and released time: $A_i = (k_i, p_i, r_i)$. The *dwell* is the duration of a key pressed and a set of samples of dwell of a certain key x is defined in (1).

Assuming that samples in S_x follow the normal distribution, the mean μ_x and standard deviation σ_x of S_x are often used as dwell features to represent a keystroke biometric sample [5,6]. According to the fundamental theorem in probability called '*law of large numbers*' [10], the larger number of dwell information for a certain key, the closer the mean feature to the user's habitual expected value. However, the estimated μ_x and σ_x may be unreliable if the size of S_x, $|S_x|$ is too small.

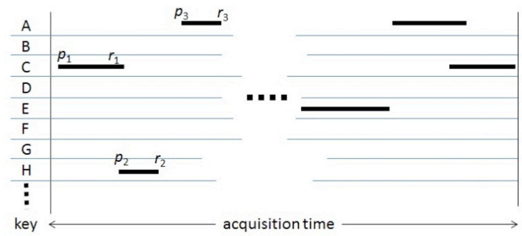


Figure 2: Acquisition of Keystroke Durations.

$$S_x = \{ r_i - p_i \mid (r_i, p_i, k_i) \in A \wedge k_i = x\} \qquad (1)$$

Hence, linguistic [5] and physiological [6] hierarchical fallback models were used to mitigate this data insufficiency problem. If a certain key x occurs infrequently, i.e., the set size $|S_x|$ is less than t, the user defined threshold, the *fallback* procedure as defined recursively in (2) was used to increase the number of samples; $S_x = fallb(x)$.

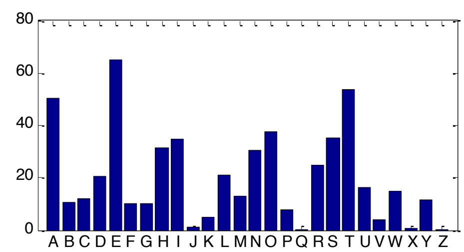


Figure 3: The average frequency of alphabet keys.

$$fallb(x) = \begin{cases} \{r_i - p_i \mid k_i \in leaf(x)\} & \text{if } |S_x| > t \\ fallb(parent(x)) & \text{otherwise} \end{cases} \qquad (2)$$

Poor correlations between parent and children nodes in the previous hierarchical models and the most correlated keys to each key are revealed from a keystroke database. The database contains 1666 sessions from 43 users with an average of 38 sessions each, where each session length has a maximum of 500 keystrokes; and the session length is 495 on average. The data was collected from university students who took 4 online exams over a semester. Each session is approximately one question, with a total of 10 questions in each exam. Not all of the students completed the exam successfully, resulting in some missing sessions. Average frequency of each alphabet key in the database is given in Figure 3, which is astoundingly similar to that in common English published in [11]. This similarity justifies the representativeness of the keystroke database.

The database is represented as a table R as given in Figure 4 where each row represents a session and column represents the average key dwell and its frequency value, which is parenthesized. Prior to the correlation analysis, the preprocessing consists of extracting a sufficient co-exist table of two keys with a user defined threshold, t_1, is defined in (3) and illustrated in Figure 4.

$$D_{x,y} = \{(\mu_{i,x}, \mu_{i,y}) \mid s_{i,x} > t_1 \wedge s_{i,y} > t_1\} \qquad (3)$$

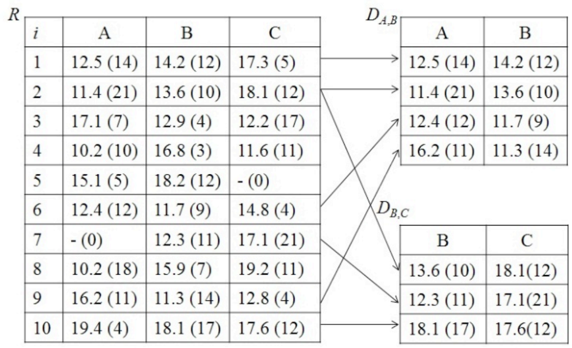
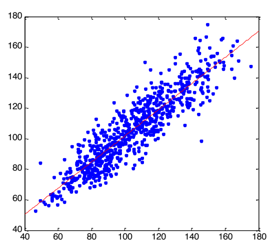
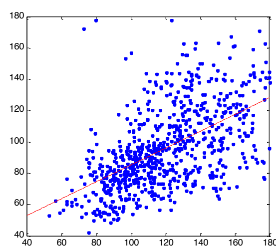


Figure 4: Extracting sufficient co-exist table with $t_1 = 7$.

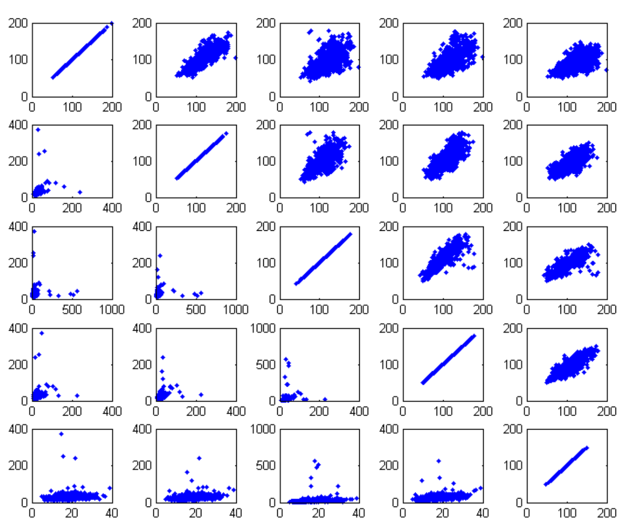
The table $D_{x,y}$, not R, is used to derive the correlation between x and y. Let $n_{x,y} = |D_{x,y}|$ be the size of instances and the *Pearson product-moment correlation coefficient*, $\rho_{x,y}$ is defined in (4). A value of $\rho_{x,y}$ closer to 1 indicates better correlation between two keys.



(a) Good correlation (0.908) b/w D and E



(b) Bad correlation (0.561) b/w A and I



(c) Correlations among vowels

Figure 5: Correlation Plots b/w pairs of key dwells

$$\rho_{x,y} = \frac{\sum_{i=1}^{n_{x,y}} (\mu_{i,x} - \overline{\mu_x})(\mu_{i,y} - \overline{\mu_y})}{\sqrt{\sum_{i=1}^{n_{x,y}} (\mu_{i,x} - \overline{\mu_x})^2} \sqrt{\sum_{i=1}^{n_{x,y}} (\mu_{i,y} - \overline{\mu_y})^2}} \qquad (4)$$

Figures 5 (a) and (b) are examples of good and bad correlated cases, respectively. Figure 5 (c) shows correlations of all pairs of vowels {a, e, i, o, u}. Upper right and lower left triangles contain the mean and standard deviation value plots for each pair of vowels. Even after outlier remover, the standard deviation distribution did not form a linear correlation.

Table 1 shows the best four keys that correlate to each key. . It should be noted that even though the key x correlates best with a key y, x is not necessarily the best correlating key for y. In all, it can be observed that majorities of keys correlate highly with other keys. Keys 'E' and 'S' have the highest correlation coefficient value and are one of the only two keys which are symmetrically correlated (the other being 'N' and 'O'). Letters like {'Q', 'X', 'Z'} are not frequently used and thus do not correlate well with others

Table 1. Correlation Coefficient-based Fallback Table.

	1st Choice		2nd Choice		3rd Choice		4th Choice	
A	S	0.847	T	0.818	E	0.801	R	0.760
B	T	0.555	H	0.527	D	0.520	S	0.518
C	T	0.797	S	0.772	E	0.763	R	0.754
D	E	0.773	T	0.759	S	0.750	C	0.715
E	S	0.876	T	0.835	R	0.826	A	0.801
F	T	0.740	E	0.708	R	0.692	S	0.690
G	T	0.707	E	0.649	R	0.641	S	0.629
H	N	0.803	I	0.760	U	0.748	T	0.744
I	N	0.809	O	0.804	T	0.771	H	0.760
J	U	0.407	I	0.400	O	0.386	P	0.374
K	O	0.577	L	0.558	I	0.546	N	0.544
L	O	0.775	T	0.720	S	0.719	I	0.716
M	N	0.789	O	0.729	U	0.724	I	0.723
N	O	0.833	I	0.809	H	0.803	U	0.790
O	N	0.833	I	0.804	U	0.777	L	0.775
P	H	0.600	I	0.599	O	0.595	U	0.587
Q	E	0.596	S	0.594	T	0.592	A	0.574
R	T	0.849	E	0.826	S	0.796	A	0.760
S	E	0.876	T	0.860	A	0.847	R	0.796
T	S	0.860	R	0.849	E	0.835	A	0.818
U	N	0.790	O	0.777	I	0.759	H	0.748
V	E	0.581	T	0.562	S	0.559	R	0.541
W	E	0.775	S	0.771	T	0.742	A	0.710
X	E	0.538	T	0.529	R	0.510	S	0.504
Y	T	0.595	E	0.569	S	0.569	R	0.551
Z	E	0.460	A	0.445	T	0.443	S	0.433

Figure 6 shows a plot of key frequency and the first choice correlation coefficient. The correlation coefficient decreases for infrequently used keys. This suggests a limiting factor in the effectiveness of the regression model, since correlations with other keys are low and it is the infrequently used keys that usually must be accounted for small sample sizes.

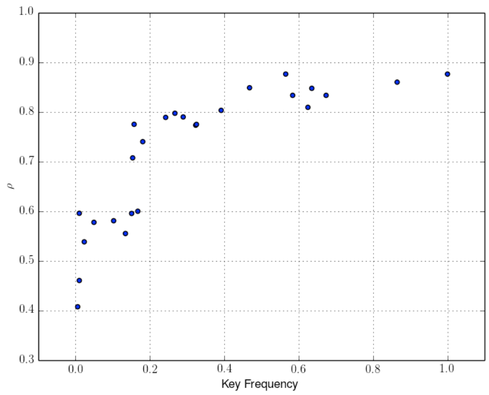


Figure 6: Max Correlation vs. Key Frequency.

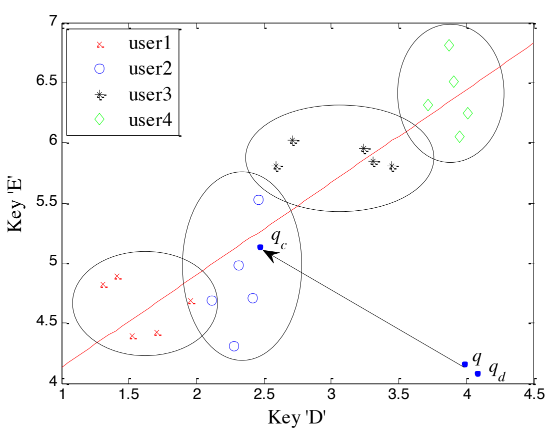
There are two other important correlation parameters that can be useful in the later fallback model. They are the slope, $\alpha_{x,y}$ (6) and the intercept, $\beta_{x,y}$ (7) of simple linear regression line (5) that fits two correlating keystroke dwell variables.

$$f_{x,y}(x) = \alpha_{x,y}x + \beta_{x,y} \tag{5}$$

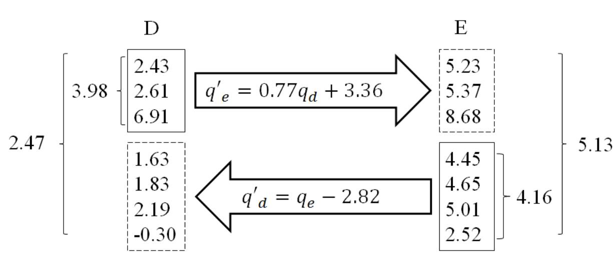
$$\alpha_{x,y} = \rho_{x,y}\frac{\sigma_y}{\sigma_x} \tag{6}$$

$$\beta_{x,y} = \mu_y - \alpha_{x,y}\mu_x \tag{7}$$

These parameter values are given in Table 3 at the end of the article.



(a) 2-Feature space of keys D and E



(b) Shifting from q (3.98, 4.16) to qc (2.47, 5.13)
Figure 7: Proposed Scenario.

3. CORRELATION BASED FALLBACK TABLE

In the previous section, it was claimed and discovered empirically that most key dwell mean values having high correlation with other keys. This section focuses on how to utilize these discovered correlation parameters to better estimate the features.

Figure 7 illustrates the essence of our claims. Suppose that there are four users and each user provided five mean feature values as plotted with its linear regression line in Figure 7 (a). Consider a query session, q, which claims to be a user 2. Only three and four samples appeared for the keys 'D' and 'E', respectively. The computed mean values q = (3.98, 4.16) are poor estimates.

In this infrequent case, the most correlating key dwell values may be summed to compute the new mean value. As illustrated in Figure 7 (b), the linear regression line can be utilized to convert the value. This new mean value augmented with linearly transformed highly correlating key dwell values is denoted as q_c = (2.47, 5.13) and we claim that this is a much better estimate as depicted in Figure 7 (a).

If the linear regression line is not used but the other key dwell values are directly augmented, this will also result in a poor estimate which is denoted as q_d = (4.08, 4.08). Previous hierarchical fallback models used direct values rather than linear transformed values.

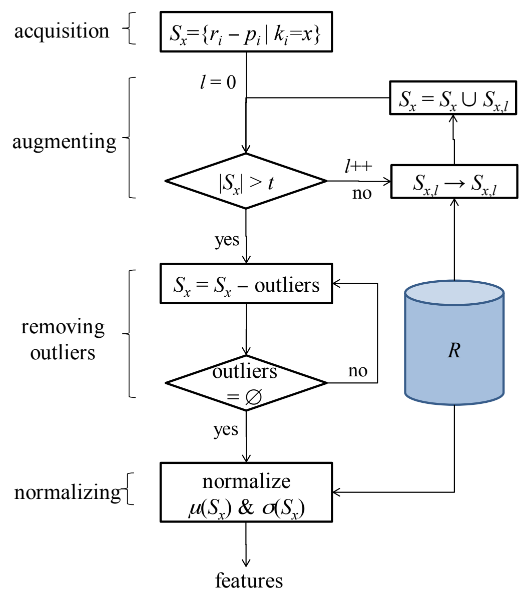


Figure 8. Flow chart of proposed correlation-based fallback table model

$$S_{x,l} = \{\alpha_{x,l}(r_i - p_i) + \beta_{x,l} \mid (r_i, p_i, k_i) \in A \wedge k_i = k_{x,l}\} \tag{8}$$

Figure 8 shows the flow chart of the proposed model that utilizes the correlation information R discovered in the previous section. The rows of R are alphabet keys containing the sorted other keys in descending order of correlation coefficient where each key has four tuples, $(k_{x,l}, \alpha_{x,l}, \beta_{x,l}, \gamma_{x,l})$ denoting the l^{th} rank key, slope, intersect, and correlation coefficient for the key x, respectively. Let $S_{x,l}$ defined in (8) denote the set of linearly transformed values, for example in Figure 7 (b), $S_{\text{'E'},1}$ = {5.23, 5.37, 8.68} is transformed from $S_{\text{'E'}}$ = {2.43, 2.61, 6.91}.

The proposed correlation based fallback table model is defined in (9) as a recursive function, *cft*. It is called initially, $S_x = cft(S_x, 0)$ with the user defined threshold for the minimum number of observations.

$$cft(S_x, l) = \begin{cases} S_x & \text{if } |S_x| > t \\ cft(S_x \cup S_{x,l}, l+1) & \text{otherwise} \end{cases} \tag{9}$$

It should be noted that (9) is imperfect. It can be infinite when the size is never greater than t or undefined if there are no more keys available. Yet it gives the succinct definition of the proposed system.

If the alphabet 'A' appears sufficiently enough, i.e., above the user defined threshold t, we just use the feature set. If not, use the linear regression function in the first choice, i.e., 'E' for 'A'. If $|S_{\text{'A'}} \cup S_{\text{'A'},1}|$ are sufficient, use them for the feature value for $S_{\text{'A'}}$.

The system was tested on a population of 30 users with a total of 400 sessions. Each user recorded between 10 and 25 sessions and each session contained between 500 and 1000 keystrokes. The samples were recorded over a semester and the population consisted mainly of university students. For each session, users were instructed to respond freely to essay-type questions.

Four different fallback models were evaluated with the same feature set. The features consist of the mean and standard deviation of each letter key, for a total of 52 features. Using only the 26 letter keys, a linguistic, physiological, regression, and default one-level model were used. The linguist and physiological models are subsets of the trees in Figure 1 containing only the relevant letter nodes, and the default model falls back to all letter keys when there are insufficient samples. A threshold value of $t=5$ was empirically chosen so that the number of infrequently used keys decrease as the number of keystrokes increase.

Each session was truncated at various intervals from 50 to 500 keystrokes to get the EER as a function of input length for each model. Figure 9 shows the equal error rate (EER) for each type of fallback model as a function of input length. Table 2 contains the EER values of each model at each input length.

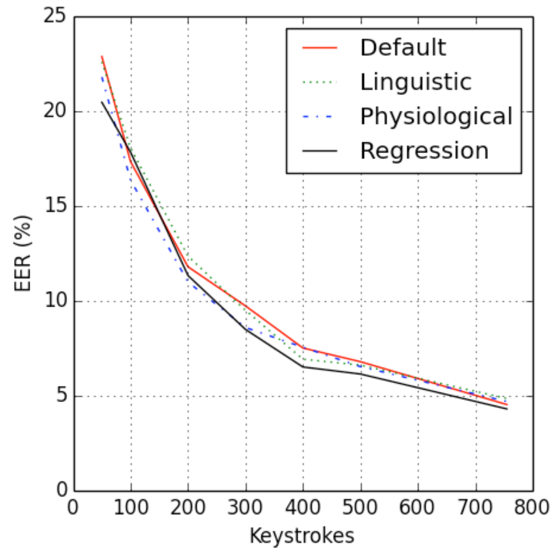


Figure 9. EER of each fallback model as a function of $|S_x|$

Table 2. Fallback Model EER Table.

$\|S_x\|$	Default	Linguistic	Physiologic	Regression
50	22.88	22.60	21.80	20.47
100	17.34	18.06	16.37	17.84
200	11.80	12.36	11.00	11.34
300	9.74	9.51	8.60	8.50
400	7.52	6.93	7.56	6.52
500	6.80	6.62	6.54	6.15
Max	4.54	4.86	4.70	4.31

4. CONCLUSIONS

The linear regression model offers modest improvements over the linguistic and physiological model with the exception of two different input sizes. This may be a result of the regression model having a higher number of fallbacks for the shorter input lengths than the hierarchical models, which rarely fall back beyond the first level. Features for infrequently occurring keys are more likely to "run out of data" quicker in the regression model, since each level contains only a single key. Calculating an accurate and useful linear regression model requires large amounts of data since its purpose is to effectively handle low occurring keystrokes. In general, the keys with the lowest frequency also have the weakest correlation with other keys, which limit the effectiveness of any fallback model. With modest improvements in accuracy, the novelty of this fallback method lies in the potential applicability to datasets where domains specific information is not available. Previously, extending a linguistic or touch-type fallback hierarchy to other languages and locales would require a hierarchy for each application. Using a correlation-based fallback only requires additional data to create the fallback table and linear regression functions.

Parent nodes in previous fallback models in [2,3] played not only the substitute role for insufficient keys, but also global feature roles. Separating these two functions allows for greater flexibility in choosing features and a particular fallback model.

Correlation attributes (alpha beta gamma) among different key dwells are used to estimate any infrequent key dwell mean value. These attributes within a user may be utilized as distinctive features for verification purpose and it remains as a future work. The correlation between groups of keys and flight times were also not considered in this study and may play a role in a complete regression fallback method.

5. REFERENCES

[1] R.M. Bolle, J.H. Connell, S. Pankanti, N. K. Ratha, and A. W. Senior, Guide to Biometrics: Springer-Verlag, 2004.

[2] D. Jurafsky and J. H. Martin, Speech and Language Processing. New Jersey: Prentice, 2000.

[3] S. M. Katz, Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer, IEEE Transactions on Acoustics, Speech, and Signal Processing, 35(3), 400–401, 1987.

[4] F. Jelinek and R. L. Mercer, Interpolated Estimation of Markov Source Parameters from Sparse Data, in proceedings of Workshop on Pattern Recognition in Practice, North Holland, Amsterdam, pp.381–397, 1980.

[5] M. Villani, C. C. Tappert, G. Ngo, J. Simone, H. St. Fort, S.-H. Cha, Keystroke Biometric Recognition Studies on Long-Text Input under Ideal and Application-Oriented Conditions, in IEEE Computer Society Workshop on Biometrics, June 2006, New York

[6] C.C. Tappert, S. Cha, M. Villani, and R.S. Zack, "A Keystroke Biometric System for Long-Text Input," *Int. J. Info. Security and Privacy (IJISP)*, Vol 4, No 1, 2010, pp 32-60.

[7] B. V. Kumar, A. Mahalanobis, and R. Juday, Correlation Pattern Recognition, Cambridge university press, 2005.

[8] S. Bleha, C. Slivinsky, and B. Hussein, "Computer-Access Security Systems using Keystroke Dynamics", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 12, no. 12, 1990, pp. 1217–1222.

[9] S. Cho, C. Han, D. H. Han, and H. Kim. "Web-based keystroke dynamics identity verification using neural network", Journal of Organizational Computing and Electronic Commerce, 10(4):295–307, 2000.

[10] M. Hazewinkel, Law of large numbers, Encyclopedia of Mathematics, Springer, 2001

[11] H. Beker, and F. Piper, Cipher Systems: The Protection of Communications. Wiley-Interscience. p. 397, 1982.

Table 3. Linear regression functions for each key

	First choice	Second choice	Third Choice	Fourth Choice
A	$f_{E,A}(x) = 0.96x + 18.20$	$f_{S,A}(x) = 0.90x + 19.71$	$f_{D,A}(x) = 0.85x + 30.26$	$f_{W,A}(x) = 0.86x + 26.19$
B	$f_{G,B}(x) = 0.78x + 17.02$	$f_{M,B}(x) = 0.79x + 13.89$	$f_{O,B}(x) = 0.65x + 22.56$	$f_{N,B}(x) = 0.73x + 17.91$
C	$f_{D,C}(x) = 0.91x + 8.73$	$f_{S,C}(x) = 0.84x + 10.28$	$f_{E,C}(x) = 0.89x + 9.42$	$f_{G,C}(x) = 0.91x + 19.26$
D	$f_{E,D}(x) = 0.96x + 2.71$	$f_{S,D}(x) = 0.89x + 5.51$	$f_{C,D}(x) = 0.80x + 21.19$	$f_{F,D}(x) = 1.01x + 6.01$
E	$f_{D,E}(x) = 0.86x + 16.18$	$f_{S,E}(x) = 0.84x + 11.95$	$f_{R,E}(x) = 0.85x + 16.45$	$f_{W,E}(x) = 0.82x + 16.31$
F	$f_{E,F}(x) = 0.74x + 19.36$	$f_{G,F}(x) = 0.79x + 24.30$	$f_{D,F}(x) = 0.68x + 26.19$	$f_{T,F}(x) = 0.67x + 31.34$
G	$f_{T,G}(x) = 0.75x + 18.40$	$f_{E,G}(x) = 0.79x + 9.81$	$f_{F,G}(x) = 0.88x + 6.94$	$f_{V,G}(x) = 0.77x + 19.49$
H	$f_{N,H}(x) = 0.92x + 4.07$	$f_{O,H}(x) = 0.81x + 11.14$	$f_{U,H}(x) = 1.01x + 1.58$	$f_{I,H}(x) = 0.78x + 18.95$
I	$f_{O,I}(x) = 0.87x + 6.74$	$f_{H,I}(x) = 0.92x + 9.45$	$f_{N,I}(x) = 0.95x + 2.87$	$f_{K,I}(x) = 0.91x + 8.93$
J	$f_{K,J}(x) = 0.47x + 39.99$	$f_{B,J}(x) = 0.63x + 31.69$	$f_{P,J}(x) = 0.53x + 36.09$	$f_{I,J}(x) = 0.39x + 46.33$
K	$f_{J,K}(x) = 1.52x + -33.27$	$f_{M,K}(x) = 0.87x + 10.48$	$f_{L,K}(x) = 0.84x + 13.11$	$f_{I,K}(x) = 0.74x + 23.22$
L	$f_{O,L}(x) = 0.76x + 18.97$	$f_{N,L}(x) = 0.83x + 16.24$	$f_{K,L}(x) = 0.80x + 19.74$	$f_{I,L}(x) = 0.71x + 28.52$
M	$f_{N,M}(x) = 0.83x + 15.18$	$f_{O,M}(x) = 0.71x + 23.58$	$f_{U,M}(x) = 0.88x + 16.13$	$f_{H,M}(x) = 0.76x + 25.10$
N	$f_{H,N}(x) = 0.86x + 17.12$	$f_{M,N}(x) = 0.93x + 7.66$	$f_{O,N}(x) = 0.76x + 19.90$	$f_{I,N}(x) = 0.75x + 25.33$
O	$f_{H,O}(x) = 0.96x + 12.77$	$f_{I,O}(x) = 0.87x + 19.01$	$f_{U,O}(x) = 1.10x + 2.37$	$f_{L,O}(x) = 0.96x + 8.91$
P	$f_{O,P}(x) = 0.77x + 18.86$	$f_{U,P}(x) = 0.92x + 13.95$	$f_{L,P}(x) = 0.79x + 21.20$	$f_{K,P}(x) = 0.83x + 19.46$
R	$f_{E,R}(x) = 0.88x + 11.41$	$f_{T,R}(x) = 0.81x + 24.46$	$f_{G,R}(x) = 0.87x + 23.42$	$f_{F,R}(x) = 0.92x + 14.45$
S	$f_{E,S}(x) = 0.96x + 10.20$	$f_{D,S}(x) = 0.90x + 17.53$	$f_{A,S}(x) = 0.78x + 18.61$	$f_{C,s}(x) = 0.82x + 26.58$
T	$f_{R,T}(x) = 0.90x + 5.21$	$f_{E,T}(x) = 0.91x + 2.79$	$f_{G,T}(x) = 0.96x + 9.74$	$f_{D,T}(x) = 0.82x + 13.69$
U	$f_{H,U}(x) = 0.74x + 21.41$	$f_{O,U}(x) = 0.68x + 21.27$	$f_{N,U}(x) = 0.74x + 18.33$	$f_{M,U}(x) = 0.78x + 15.74$
V	$f_{G,V}(x) = 0.86x + 14.84$	$f_{F,V}(x) = 0.91x + 6.72$	$f_{E,V}(x) = 0.80x + 11.53$	$f_{T,V}(x) = 0.80x + 18.15$
W	$f_{E,W}(x) = 0.89x + 14.72$	$f_{S,W}(x) = 0.81x + 18.68$	$f_{D,W}(x) = 0.80x + 25.74$	$f_{R,W}(x) = 0.83x + 22.07$
Y	$f_{H,Y}(x) = 0.79x + 17.78$	$f_{N,Y}(x) = 0.81x + 12.32$	$f_{I,Y}(x) = 0.72x + 23.10$	$f_{O,Y}(x) = 0.69x + 20.76$