A Equalização de Assinaturas Keystroke Oriundas de Teclados Mobile e Físicos Como Instrumento de Viabilização de Interoperabilidade

The Equalization of Keystroke Signatures from Mobile and Physical Keyboards as an Enabler of Interoperability

Ana Beatriz da C. Figueiredo[1], **Rodolfo B. de B. Garcia**[1]

[1]Departamento de Computação (DCOMP) - Universidade Federal de Sergipe (UFS)
Av. Marechal Rondon, – Jardim Rosa Elze – CEP 49100-000
Sao Cristóvão – SE – Brazil

anabeatrizcfig@gmail.com, rodolfo.botto@dcomp.ufs.br

Abstract. *Keystroke authentication needs to be repeatedly expressed. This makes the interoperability between signatures originating from physical and virtual keyboards relevant to this technique. However, this correspondence depends on equalization techniques. Thus, this paper aims to explore the influence of these strategies on keystroke dynamics. To this end, databases were equalized in order to infer accuracy using the KNN algorithm. In addition, a database was created consisting of the same volunteers signing on both types of keyboards. It was observed that the equalization process yielded equal constants for physical and virtual keyboards and that combining data from different types of keyboards did not decrease the algorithm's accuracy. However, the use of equalization in the latter did not improve its performance. Finally, interoperability proved to be feasible; however, new equalization or analysis techniques are necessary.*

Resumo. *O keystroke necessita ser recorrentemente expresso. Isto torna a interoperabilidade entre assinaturas oriundas de teclados físicos e virtuais relevantes para essa técnica. Mas, essa correspondência depende de técnicas de equalização.Assim, esse artigo objetiva explorar a influência destas estratégias no keystroke. Para tal, equalizou-se bancos de dados a fim de inferir a acurácia a partir do algoritmo KNN. Ademais, foi confeccionado um banco de dados composto pelos mesmos voluntários assinando nos dois tipos de teclado. Notou-se a obtenção de constantes do processo de equalização iguais no teclado físico e virtual e que a mescla dos dados de diferentes tipos de teclado não diminuiu a acurácia do algorítmo. Mas, o uso da equalização neste último não potencializou o seu desempenho. Por fim, a interoperabilidade demonstra-se viável porém novas técnicas de equalização ou de análises fazem-se necessárias.*

1. Introdução

Acarretada pela frequente necessidade de autenticações em sistemas que armazenam informações sensíveis, como *netbanking* e redes sociais, a utilização de senhas biométricas vem ascendendo. No Brasil, 89% dos consumidores acham mais fácil usá-las

em detrimento das senhas tradicionais, como mostrado em uma pesquisa realizada pela VISA e pela AYTM Market Research em 2022 [Idwall 2025]. Essa utilização é fomentada pelo fato de que os dispositivos móveis, que são os meios mais usados para acessar esses sistemas, apresentam uma grande quantidade de sensores integrados de fábrica capazes de promover altos níveis de segurança biométrica, como o de impressão digital e o de reconhecimento facial.

Por outro lado, sabe-se que os teclados convencionais como os de notebooks e desktops, também empregados para o acesso a diversos sistemas, tendem a não possuir esses recursos de sensoriamento de forma nativa. Inclusive, para que esses dispositivos tenham o mesmo conjunto de sensores de um *smartphone*, visando aumentar a segurança dos seus acessos, é necessário um investimento financeiro elevado. Diante disso, emerge uma alternativa de autenticação biométrica baseada na forma como o usuário digita uma senha: o *keystroke*. Esta técnica apresenta a vantagem de não requerer hardware específico, além de aspectos como a praticidade e a unicidade, como fomenta [Amin and Di Iorio 2025] e [Teh et al. 2013]. Com isso, o *keystroke* passa a ser uma alternativa simples e barata que pode ser facilmente integrado aos sistemas de segurança tradicionais do tipo login e senha.

É válido ressaltar ainda que a segurança proporcionada pela técnica do *keystroke* pressupõe que o padrão temporal de digitação do usuário seja recorrentemente digitado, funcionando como um processo de treinamento do ritmo de digitação. Caso contrário, a técnica pode ter sua confiabilidade comprometida. O impasse relatado em [Figueiredo and Garcia 2025], é que os desktops apresentam taxas de acesso cada vez menos expressivas ao passo que a utilização de dispositivos móveis cresce significativamente. Em termos práticos, o *keystroke* de teclados físicos costuma ser pouco treinado e não confiável, enquanto que em teclados virtuais sua confiabilidade por conta da grande quantidade de autenticações é elevada.

Ainda em [Figueiredo and Garcia 2025], as análises de características temporais capturadas durante a digitação em teclados físicos e em teclados virtuais mostraram padrões de comportamentos que possibilitam a viabilidade de interoperabilidade entre os dispositivos. O problema abordado aqui é justamente onde esses padrões se diferenciam, principalmente pela diferença de tempo médio que cada evento temporal gera. Dessa forma, este trabalho tem como objetivo principal utilizar a equalização das características temporais proposta por [Montalvão Filho and Freire 2006], que utiliza como artifício a distribuição log normal, que coincide com as distribuições encontradas em [Figueiredo and Garcia 2025]. É importante mencionar que a equalização foi somente aplicada em intervalos oriundos de teclados físicos. Porém, na presente produção científica, o processo de equalização é realizado tanto em teclados físicos quanto em teclados virtuais, a fim de favorecer a interoperabilidade entre os mesmos e manter o mesmo nível de segurança e confiabilidade das autenticações baseadas em *keystroke*.

Adicionalmente, é necessário contextualizar a situação dos bancos de dados existentes no âmbito da técnica de dinâmica de digitação. Primeiramente, sabe-se que o maior contingente das bases de *keystroke* concentra-se somente no teclado mobile ou no teclado físico, e não em ambos. Além do mais, nessas bases de dados, o universo de participantes é diferente, não apresentando um mesmo contribuinte digitando em dois teclados de tipos distintos. Assim sendo, é considerado relevante e visa-se a criação de um banco de dados

autoral proveniente dos mesmos voluntários para teclados físicos e virtuais.

Em síntese, defronta-se este trabalho com dois contextos principais no cenário da técnica *keystroke*: o prévio indício da positiva influência da equalização na interoperabilidade e a ausência de bancos de dados viáveis. Dessa forma, levanta-se a hipótese de que, através da equalização das características temporais essa interoperabilidade é potencializada. Vislumbra-se, portanto, analisar tal hipótese em bases de dados oriundas de ambos os teclados compostas tanto por usuários diferentes quanto para os mesmos participantes.

Finalmente, o presente artigo científico está organizado a partir daqui da seguitne forma: a Seção 2 traz os trabalhos relacionados a temática. Em seguida, a Seção 3 abrange os principais conceitos que fundamentam a produção. Posteriormente, são apresentados nas Seções 4 e 5, respectivamente, os dois experimentos projetados: o experimento A, que realiza e analisa o impacto da equalização em bases já existentes, e o experiemnto B, que traz o debate a cerca da equalização e análise de base com correspondência de voluntários. Por fim, a Seção 6 apresenta as conclusões deste artigo e propõe trabalhos futuros.

2. Trabalhos Relacionados

Os trabalhos existentes na literatura evidenciam: análises e apresentações do cenário atual do *Keystroke* [Shadman et al. 2025]; constatações de melhorias em identificar usuários a partir da equalização dos intervalos obtidos em teclados físicos [Montalvão et al. 2015] e [Montalvão Filho and Freire 2006]; análises de novas características além das caracteristicas temporais [Kaluarachchi et al. 2023]; comparativos entre modelos de autentificação utilizando diferentes algoritmos [Amin and Di Iorio 2025] e também indícios comprovados da interoperabilidade entre teclados físicos e virtuais e que este fenômeno pode ser corroborado com o uso da equalização [Figueiredo and Garcia 2025].

Além disso, a grande parte das bases de dados existentes não contempla, de forma estruturada, a coleta simultânea em teclados físicos e virtuais com os mesmos participantes. Bases que apresentam essa característica como a [Belman et al. 2019], por exemplo, apresentam informações adicionais que dificultam o estudo da interoperabilidade. Complementarmente, a documentação de bases previamente publicadas para essa área muitas vezes é deficitária no que tange a documentação. Nota-se, em muitos casos, uma fragilidade na divulgação de informações ligadas ao método de coleta dos dados, ao tempo entre as sessões e a delimitação clara das mesmas, além de não elucidar em muitas situações detalhes do controle do ambiente.

Diante desses contextos, esta produção difere-se pela observação específica da influência do processo de equalização para a busca da interoperabilidade entre teclados físicos e virtuais bem como pelo estudo de um banco de dados experimental que contemple os mesmos participantes em ambos os tipos de plataformas de digitação.

3. Fundamentação Teórica

3.1. Keystroke Dynamics

Segundo [Shadman et al. 2025], o *Keystroke Dynamics* é um método de autenticação conveniente e de fácil utilização, no qual uma pessoa é autenticada com base em um modelo extraído a partir do seu padrão de digitação durante um período de registro . Dessa forma,

essa técnica configura-se como um método de biometria comportamental, ou seja, refere-se aos comportamentos e hábitos do usuário.

Dentro do modelo extraído, têm-se os conceitos das características temporais que formam as assinaturas *keytroke*. Essas características são intervalos capturados durante o registro da assinatura, neste trabalho, representado pela senha. São eles: Holding (H), UpDown (UD) e DownDown (DD). O intervalo H consiste no tempo que uma tecla se mantém pressionada, já o intervalo UD abrange o momento em que uma tecla é liberada até o momento em que a próxima é pressionada e, por fim, o intervalo DD equivale ao tempo entre o pressionamento de duas teclas.

Além disso, foi demonstrado em [Montalvão et al. 2015] o conceito de assinatura estável. Este, é descrito como o comportamento padrão de digitação que vigora após o usuário alcançar a familiarização com o processo de digitação da senha. Considerando a estabilização da digitação, [Figueiredo and Garcia 2025] constatou a partir da análise das características temporais, comportamentos semelhantes de médias e desvios padrões dos intervalos dentro de uma distribuição log-normal porém, com diferentes escalas. Com isso, surgem os primeiros indícios de que é possível correlacionar assinaturas biométricas obtidas de diferentes tipos de teclados. Todavia, faz-se imprescindível a realização de um processo de equalização das características temporais para aproximação das médias e desvios padrões. A semelhança dos comportamentos dos padrões de diferentes teclados é exemplificada na figura 1.

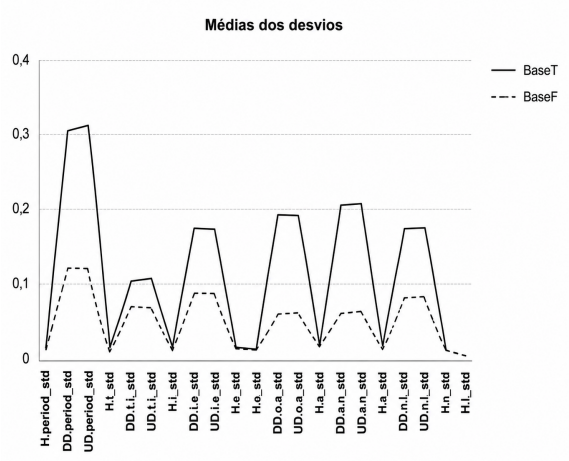


Figura 1. Comportamento do desvio padrão dos intervalos em teclados touchscreen (base T) e físico (base F)

3.2. Equalização das características temporais

Em [Montalvão Filho and Freire 2006], o autor concentra-se nos intervalos DD de 3 bases oriundas de teclados físicos. Para esta análise, é assumido que esses valores são instâncias de uma variável aleatória contínua X. Após a elaboração de um histograma de (X), utilizado como uma aproximação grosseira da Função Densidade de Probabilidade (PDF) real dos dados, observou-se que o cálculo de Y = log (X) resulta em uma distribuição normal. Portando, implica-se que X é aproximadamente log-normal.

Em conformidade com essas constatações, [Montalvão et al. 2015] estabelece in-

ferências a respeito dos intervalos temporais H e UD também obtidos por teclados físicos. O artigo revela que os intervalos Holding (X_H) também seguem uma distribuição aproximadamente log-normal, isto é, $\log(X_H)$ é aproximadamente Gaussiana. Em contrapartida, os intervalos Up Down (X_{UD}) não podem assumir essa distribuição, já que eles podem assumir valores negativos quando a pessoa pressiona uma tecla antes de soltar a anterior, por exemplo. Motivado por esse fato, os dados referentes a eles são desconsiderados nas análises que se sucedem.

Por fim, foi consolidada a seguinte transformação direta de equalização dos intervalos disponibilizada em [Montalvão et al. 2015] tanto para os intervalos DD quanto para os H:

$$g(y; \mu_y, \sigma_y) = \frac{1}{1 + \exp\left(-1.7 \cdot \frac{y-\mu_y}{\sigma_y}\right)} \tag{1}$$

Fonte:[Montalvão et al. 2015]

Sendo y o valor obtido ao aplicar o logaritmo natural no intervalo original, μy a média da variável y e por fim, σy , o desvio padrão de y. Ainda em [Montalvão et al. 2015], os autores estimaram os seguintes parâmetros fixos de $\mu_y^{DD} = -1.61$, $\sigma_y^{DD} = 0.64$, $\mu_y^H = -2.46$ e $\sigma_y^H = 0.33$ a partir de toda a base de dados *tie5Roanl*, a mesma utilizada em [Figueiredo and Garcia 2025], conforme mencionado.

4. Experimento A: Equalização e análise de bases já existentes

4.1. Metodologia

Partindo das observações da Fundamentação Teórica, este experimento objetiva equalizar os dados dos intervalos DD e H nas duas bases utilizadas por [Figueiredo and Garcia 2025] a fim de torná-los mais próximos em ambos os teclados. Para tal, será utilizada a transformação direta de equalização não linear e sem memória desses intervalos, demonstrado na equação 1 da subsessão 3.2 desta produção. Após isto, os resultados serão analisados com o auxílio do algoritmo de classificação KNN, cuja escolha foi devido a sua simplicidade e à possibilidade de comparação com os resultados obtidos por [Figueiredo and Garcia 2025]. Dessa forma, admite-se que quanto mais próximas as amostras obtidas por teclados diferentes, menor será a acurácia do classificador.

As duas bases de dados utilizam a mesma palavra ".tie5Roanl"para modelagem da assinatura biométria por *keystroke*. As bases apresentam a seguinte descrição: A primeira base, apresentada por [Killourhy and Maxion 2009] é formada por 31 características temporais obtidas por 51 participantes, 400 amostras para cada participante, divididas em 8 sessões com 50 assinaturas cada, todas obtidas a partir de teclados físicos. A segunda base, usada em [Antal and Szabó 2015], é composta por 41 características temporais em 51 amostras de cada um dos 42 participantes (37 de tablets e 5 de smartphones), divididas em duas sessões que não foram claramente definidas no trabalho de origem. Ambas as bases originais, apresentam diferentes conjuntos de participantes, isto é, as amostras não foram obtidas pelas mesmas pessoas. Além disso, como observado, as formas de obtenção das assinaturas varia entre tipos diferentes de teclados. Por isso, neste experimento foi mantido somente os eventos em comum entre as bases, totalizando 20 características temporais.

Assim sendo, μ e σ , ou seja, as constantes referentes à média das médias e à média dos desvios padrões, respectivamente, foram calculadas tanto para os intervalos H quanto para o intervalo DD de ambas as bases. Detendo dessas constantes e também do valor do $\log_e$ dos intervalos, a equalização foi realizada. Como resultado, foram geradas 6 novas bases, sendo elas compostas pelos intervalos equalizados DD, H e ambos juntos tanto do teclado físico quanto do virtual. Ademais, foram consideradas na base dos intervalos oriundos do teclado físico apenas as assinaturas estáveis, equivalente a segunda metade de cada sessão. Já para a base oriunda do teclado virtual, foram analisadas todas as suas 51 amostras de cada participante, já que a divisão de sessões não foi declarada no trabalho de origem.

Neste experimento, o KNN foi executado 10 vezes e para cada iteração foi aplicado 80% das amostras para a etapa de treino e 20% para a etapa de teste. Destaca-se ainda que o número de vizinhos estabelecido foi 3. Esse processo foi repetido para as bases que continham apenas intervalos DD, apenas intervalos H e em conjunto os intervalos DD e H. Finalizando esse experimento foram calculadas as médias da acurácia das 10 execuções para os três casos mencionados anteriormente.

4.2. Resultados e discussões

Em primeiro lugar, é consolidado o cálculo das constantes μ (média das médias dos intervalos dos participantes) e σ (desvio padrão da média dos intervalos dos participantes). Esses dados estão elucidados na Tabela 1.

Tabela 1. Média e desvio padrão dos intervalos por base

Teclado	**Intervalo**	**Média (μ)**	**Desvio padrão (σ)**
Físico	H	-2.448047	0.322984
Físico	DD	-1.875258	0.533235
Virtual	H	-2.321705	0.332111
Virtual	DD	-0.932035	0.519733

É possível observar, a partir deste experimento, a proximidade das constantes μ e σ para os intervalos H no teclado virtual e no teclado físico. Demonstrando assim a tendência de pressionar as teclas pelo mesmo intervalo de tempo, independente do tipo de teclado. Já para os intervalos DD, é notada a necessidade da equalização pela diferença de médias obtidas pelos dois tipos de teclados, apesar dos desvios próximos.

Somado a isso, os valores para as constantes μ e σ nesse experimento para o teclado virtual foram extremamente próximos aos valores elucidados por [Montalvão et al. 2015] para o teclado físico citados na subseção 3.2, alcançando até a valores iguais como $\sigma_H = 0.33$. Além disso, a baixa diferença do desvio padrão para os intervalos DD no teclado físico pode ser dada pela quantidade de características temporais pós redução do universo de características para as bases deste trabalho e, como conclusão, mostra que esse tipo de filtragem não compromete a análise, ou seja, há a possibilidade de analisarmos palavras digitadas de formas diferentes entre teclados físicos e virtuais.

Dando prosseguimento, é possível demonstrar agora os valores das acurácias médias das 10 execuções do KNN para as bases com intervalos H e DD, apenas DD e apenas H, baseados nos quantitativos pós equalização através da Tabela 2:

Tabela 2. Média das acurácias por tipo de intervalo pós equalização

Intervalos	Média da Acurácia (%)
H,DD	76.31
DD	78.42
H	69.47

Em comparação com os resultados descritos em [Figueiredo and Garcia 2025], onde foi usado o mesmo protocolo experimental sem utilização da equalização nos intervalos, a acurácia média para a base contendo os intervalos DD e H foi de 94.19% para 76.31%, para a base contendo apenas os intervalos DD foi de 95.24% para 78.42% e, por fim, para a base contendo apenas os inervalos H foi de 74.11% para 69.47%. É visível que o classificador teve mais dificuldade em reconhecer de onde as amostras vieram, com grandes quedas nas duas primeiras bases e com uma leve queda na base contendo apenas os intervalos H. Logo, pode-se afirmar que os dados de ambos ficaram mais semelhantes e equiparáveis já que se alcançou-se uma menor capacidade de distinção entre eles. Devido a isso, nesta etapa, uma acurácia menor é algo positivo, pois aumenta o indício de uma possível interoperabilidade entre assinaturas keystroke, independente do tipo de teclado usado para obter as características temporais.

5. Experimento B: Confecção, equalização e análise de base com correspondência de voluntários

5.1. Metodologia

O Experimento B visa desenvolver uma metodologia própria de captura de assinaturas temporais que funcione em teclados físicos e virtuais. O grande propósito é de obter o mesmo conjunto de participantes digitando a mesma palavra nos dois tipos de teclados. Finalmente, objetiva-se realizar um comparativo das assinaturas neste novo database autoral utilizando também a técnica de equalização dos intervalos DD e H e as 10 repetições do KNN, como realizado no Experimento A.

Portanto, o Experimento B será sistematizado em quatro subdivisões: aquisição das assinaturas *keystroke*, etapa de implementação, confecção do dataset, e equalização e classificação dos intervalos do novo dataset.

5.1.1. Aquisição das assinaturas Keystroke

Faz-se necessário compreender o método utilizado pelos teclados virtuais e pelos teclados físicos para registrar os eventos temporais de digitação. Ao se tratar de teclados físicos, sabe-se que o sistema de eventos do navegador, chamado DOM Events, já oferece os eventos nativos de KeyDown e KeyUp. Dessa forma, quando uma tecla é pressionada o navegador traz as informações como o valor da tecla (key) e o tipo de evento (event) no KeyboardEvent. Nota-se, portanto, que para notebooks e desktops é possível obter o registro preciso do momento que cada letra é pressionada e liberada.

Em contrapartida, a forma como a digitação é analisada em teclados virtuais é diferente. Em [Kantor 2021] é exposto que, ao utilizar teclados virtuais, a norma do

W3C (World Wide Web Consortium) estabelece que o KeyboardEvent deve apresentar *e.keyCode = 229* e *e.key = "Unidentified"*. Dessa forma, não é possível obter a captação individual de cada letra. Devido a esse impasse, pesquisadores utilizam as modificações no input para possibilitar o cálculo entre uma letra e outra. Conclui-se, assim, que não há a exposição da delimitação precisa do momento do KeyUp e do KeyDown nesses teclados.

Dando prosseguimento, no que se refere a esquematização do fluxo de funcionamento e regras de coleta dos intervalos DD e H, foi determinado um quantitativo de 30 voluntários que realizaram a digitação da palavra ***"password"***. Essa atividade foi dividida em duas sessões com intervalos de um dia entre elas e com a palavra escolhida digitada 25 vezes em cada sessão. Em acréscimo, é primordial uma padronização e determinação rigorosa das instruções de coleta passadas para os participantes. A escolha dessa palavra ocorreu devido ao fato dela não apresentar caracteres especiais e, consequentemente, a sequência rítmica ser a mesma no celular e no computador.

5.1.2. Etapa de Implementação

Para a elaboração dos algoritmos foi utilizado o registro do tempo de keydown e keyup das teclas que compõem a palavra digitada. Para tal intuito, utilizou-se os valores ofertados no próprio console do navegador. Desse modo, é possível alcançar o Holding através da subtração do timestamp do keyup e do keydown de uma mesma tecla e o DownDown a partir da subtração do timestamp do keydown da tecla atual e do timestamp do keyup da tecla anterior.

De modo concomitante, a elaboração das tabelas relacionais do banco de dados foram sendo concretizadas. Nele, foram armazenadas as seguintes informações: o id do voluntário, o id da linha, a senha digitada, os intervalos Holdings de cada letra e os intervalos DownDown de uma letra para outra, o tipo de teclado utilizado e a sessão correspondente. Subsequentemente, evidenciou-se a necessidade de uma função de validação que impedisse a ocorrência de senhas digitadas incorretamente ou tentativas de escritas corrigidas com o uso do *backspace*. Essa necessidade manifestou-se devido ao fato da sequência analisada precisar ser composta exatamente pelas letras que compõe a palavra escolhida, ou seja, tentativas de correção podem alterar o padrão dinâmico de digitação. Após testes, foi observado que o computador normalmente está por default com letra minúscula acionada enquanto a maioria dos celulares já começam com letra maiúscula e alterá-lo para minúsculo alteraria o ritmo de digitação. Em vista disso, a fim de contornar esta problemática, foi utilizada a estratégia do teclado virtual aceitar a primeira letra da senha tanto maiúscula quanto minúscula. Isso ocorreu para que o voluntário não precisasse alterar a forma como a tecla já vinha no seu celular e pressionar uma tecla a mais alterando assim o padrão rítmico. Entretanto, foi frisado aos participantes da pesquisa que digitassem diretamente da forma que a primeira letra já estava, ou seja, a primeira tecla a ser pressionada precisaria ser o "p"sendo ela maiúscula ou minúscula. Essa decisão também foi tomada a fim de não solicitar aos voluntários que alterassem as configurações dos seus aparelhos.

Ainda no contexto do mobile, devido ao que foi elucidado na subseção 5.1.1, a informação disponível referente ao pressionamento da tecla, era somente a mudança no valor do input e o momento em que essa alteração ocorreu. Em contrapartida, a

informação de qual tecla foi pressionada não era ofertada. É imprescindível salientar que essa limitação é decorrente do funcionamento padrões do navegador em dispositivos móveis e não a implementação desenvolvida no presente artigo. Entretanto, ainda era realizável identificar o intervalo de tempo entre uma mudança e outra que aconteceu no input. Concluiu-se, portanto, que utilizando teclados *mobile* de modo convencional e com o hardware cotidiano não seria possível calcular o Holding, apenas os intervalos de uma tecla para a seguinte, ou seja, o intervalo DownDown.

Diante disso, determinou-se calcular no computador ambos os intervalos mediante a viabilidade e apenas o intervalo DD no teclado do celular, já que se configurava como a única possibilidade para o mesmo. Adicionalmente, uma nova forma de validação da senha digitada precisava ser implementada. Para isso, foi utilizado a própria função que é acionada quando ocorre mudança no campo de entrada e a que é chamada quando a senha é submetida. A lógica utilizada foi comparar o tamanho dos inputs para verificar se alguma letra foi apagada ou se mais de uma letra havia sido inserida por vez, evitando que o voluntário pudesse submeter apenas colando a palavra, usando sugestões do teclado ou a ferramenta de autocompletar.

5.1.3. Confecção de um banco de dados experimental

Tendo em vista o contexto e os déficits supracitados referentes aos bancos de dados de *keystroke* já existentes, opta-se pelo desenvolvimento de um dataset preliminar próprio contendo os intervalos das assinaturas *keystroke* de teclados *touchscreen* e tradicionais oriundos das mesmas pessoas com o auxílio da página implementada nas etapas antecedentes. Dessa forma, todo o processo de confecção do banco foi realizado tendo um foco específico na questão da interoperabilidade, trazendo ainda explicitamente os intervalos H e DD para os teclados físicos e DD para os teclados virtuais. Dessa forma, com um banco personalizado foi possível chegar ainda a um maior conhecimento e controle das variáveis envolvidas no processo de coleta como a quantidade de sessões, o tempo determinado entre elas e a quantidade de suas respectivas assinaturas, a seleção da palavra chave, as instruções que foram passadas para o usuário. Dessa maneira, gera-se consequentemente uma análise com maior eficácia, organização e confiabilidade.

Para tal, inicialmente, foi realizada a seleção dos participantes e o estabelecimento de um ID correspondente a ele. Foi solicitado que os participantes cumprissem as orientações estabelecidas na etapa inicial mencionadas no tópico 4.2.1 e além disso, que a posição na qual elas digitariam fosse sempre sentada e relativamente semelhante para todas as assinaturas.

Após um período de aproximadamente 1 semana abrangendo a contribuição de todos os colaboradores, foi realizada a organização no dataset. Com isso, os dados com menor precisão não foram utilizados. Complementarmente, ocorreu também o tratamento dos dados removendo linhas de testes ou que possuiam inconsistências como a quantidade incompleta de registros, por exemplo.Com isso, no total, foram solicitadas 47 pessoas, porém, foram obtidos **25 conjuntos de amostras válidas**, compostos por um total de 100 registros por participante ($25\ amostras \times 2\ sessões \times 2\ tipos\ de\ teclado$). É importante mencionar que a base de dados construída representa uma amostragem cujo domínio é constituído por usuários digitando em teclados físicos de computadores em navegadores

diversos, ficando essa escolha a critério dos participantes e celulares de sistema operacional Android. Em acréscimo, as pessoas colaboradoras apresentavam variação quanto à idade, sexo e grau de familiaridade com digitação.

5.1.4. Equalização e análise dos intervalos do novo dataset

O protocolo de equalização é semelhante ao do Experimiento A. Inicialmente, os valores dos intervalos foram transformados de milissegundos para segundos a fim de estar compatível com a unidade de medida que acarretou nos valores de μ e σ que serão utilizados para a equalização que são os oriundos do Experimento A. Após equalizar os intervalos desse novo banco, o KNN foi utilizado juntamente com a validação de modelos 10-fold cross validation, cujo desempenho final é obtido pela média dos resultados das 10 execuções. É válido mencionar ainda a utilização da composição estratificada a fim de garantir a mesma quantidade de amostras por pessoa em cada repartição. Por fim, utilizou-se a acurácia como métrica de avaliação e após testes , optou-se pela escolha de 7 vizinhos e pela distância Manhattan. Ademais, o critério de ponderação *weights* foi alterado de ”uniform”para ”distance”.

Por conseguinte, chega-se ao valor das médias da acurácia para o novo banco autoral antes e após a equalização. É de suma importância frizar que apenas os intervalos DD estão sendo observados devido ao fato de estarem presentes em ambos os tipos de teclado. Em complemento, para fins comparativos, foi aplicado novamente esse algorítmo e observada as suas respectivas médias de acurácias agora nos seguintes dados: A base do teclado físico usada no experimento A (base F) composta pelos intervalos H e DD com e sem equalização , somente com os intervalos H equalizados e somente com os intervalos DD também equalizados seguida da mesma análise para a base do teclado touchscreen (Base T) analisada no experimento A.

5.2. Resultados e discussões

Os resultados referentes ao Experimento B estão relatados na sequência:

Tabela 3. Acurácia obtida no banco autoral (PC + Mobile) considerando o intervalo DD com e sem equalização

Intervalo	Equalização	Acurácia (%)
DD	Não	59,80
DD	Sim	57,96

Tabela 4. Acurácia obtida para as bases F (físico) e T (touchscreen), considerando os intervalos H+DD com e sem equalização, e os intervalos individuais com equalização

Base	Intervalos	Equalização	Acurácia (%)
F	H + DD	Não	79,90
F	H + DD	Sim	85,13
T	H + DD	Não	58,50
T	H + DD	Sim	63,16
F	H	Sim	62,59
F	DD	Sim	50,33
T	H	Sim	23,25
T	DD	Sim	46,50

Com isso, pode-se observar duas inferências principais. A primeira é que com a análise somente dos intervalos DownDown para teclados de apenas um tipo e somente desses intervalos para o banco que contém ambos os tipos de plataforma, a acurácia foi extremamente próxima. Inclusive, com os dados compostos pelos dois tipos de teclado apresentou maior desempenho. Com isso, nota-se que o fato de misturar os teclados tradicionais e virtuais não tornou mais difícil identificar a pessoa a qual pertence aquela assinatura.

A segunda observação é que ao analisar o desempenho do KNN antes e depois da equalização dos dados do banco autoral, verificou-se que a acurácia anterior à equalização foi superior. Isso ocorre porque a medida que misturado e aproximado os valores das assinaturas dos dois tipos de teclados com a equalização, o algorítmo passou a confundir uma assinatura no teclado virtual da pessoa A com o do teclado tradicional da pessoa B. Isto é, o grau de semelhança potencializado foi de teclados diferentes para pessoas diferentes e não de pessoas iguais para teclados diferentes como almejava-se.

6. Conclusões

Este trabalho objetivou a aplicação da equalização de intervalos DD e H em assinaturas biométricas obtidas pelo método de *keystroke*, com o intuito de possibilitar a interoperabilidade de ritmos de digitação obtidos por diferentes tipos de teclados (físicos e virtuais). No Experimento A, as médias e desvios padrões dos intervalos H mostram a tendência do usuário de a pressionar as teclas durante o mesmo intervalo de tempo, independente do tipo de teclado. Já os intervalos DD, refletem a necessidade da equalização para possibilitar a interoperabilidade. Por fim, a aplicação da equalização nas amostras de bases reconhecidas na literatura, obtidas por teclados físicos e virtuais, dificultou a identificação da origem de cada amostra por parte do KNN, indiciando mais uma vez a possibilidade de interoperabilidade. No Experimento B, a necessidade de comparar amostras do mesmo conjunto de participantes levou à aquisição de uma base de dados autoral. A acurácia utilizando somente os intervalos DD, que apresentaram desde o início a maior necessidade da equalização, não melhorou após o processo de equalização trazendo resultados desencorajadores referentes ao uso desse tipo em questão de equalização. Esse fato pode indicar que a equalização geral usando todas as amostras das bases de teclado físico e virtual, considerando um único perfil, pode não ser a melhor estratégia.

Visando o aprimoramento e continuidade desta pesquisa, é proposto como trabalhos futuros: (a) a criação de um teclado virtual desenvolvido dentro da própria aplicação, no qual seriam capturados os momentos que cada tecla individualmente é pressionada e liberada possibilitando o cálculo dos intervalos Holdings também nos teclados mobile (b) a utilização de outras estratégias de equalização que se adaptem melhor a bancos compostos por ambos os dispositivos de digitação; (c) realizar os experimentos de cálculo de acurácia no banco autoral utilizando outros classificadores como Random Forest e SVM, por exemplo;(d) ampliar a base autoral mediante a aprovação de um Comitê de Ética em Pesquisa a fim de fortalecer a precissão estatística; (e) Utilizar outros parâmetros de avaliação de desempenho do classificador além da acurácia, como Equal Error Rate (EER), False Acceptance Rate (FAR) e False Rejection Rate (FRR).

Referências

Amin, S. and Di Iorio, C. (2025). A review of several keystroke dynamics methods. *arXiv preprint arXiv:2502.16177*.

Antal, M. and Szabó, L. Z. (2015). Keystroke dynamics – android platform. `https://www.ms.sapientia.ro/~manyi/keystroke.html`. Acesso em: 20 ago. 2025.

Belman, A. K., Wang, L., Iyengar, S. S., Sniatala, P., Wright, R., Dora, R., Baldwin, J., Jin, Z., and Phoha, V. V. (2019). Su-ais bb-mas (syracuse university and assured information security - behavioral biometrics multi-device and multi-activity data from same users) dataset.

Figueiredo, A. B. C. and Garcia, R. B. (2025). Interoperabilidade de assinaturas keystroke para uso em sistemas de segurança baseados em login e senha: Uma análise de características temporais. Brasil. SBC.

Idwall (2025). Autenticação por biometria: o que é e como funciona. `https://blog.idwall.co/autenticacao-por-biometria/`. Acesso em: 23 jul. 2025.

Kaluarachchi, N., Kandanaarachchi, S., Moore, K., and Arakala, A. (2023). Deft: A new distance-based feature set for keystroke dynamics. *arXiv preprint arXiv:2310.04059*.

Kantor, I. (2021). Keyboard: keydown and keyup. `https://id.javascript.info/keyboard-events`. Acessado em: 14 jan. 2026.

Killourhy, K. S. and Maxion, R. A. (2009). Keystroke Dynamics - Benchmark Data Set: Accompaniment to "Comparing Anomaly-Detection Algorithms for Keystroke Dynamics"(DSN-2009). `https://www.cs.cmu.edu/~keystroke/`. Accessed: 2025-07-25.

Montalvão, J., Freire, E. O., Bezerra Jr, M. A., and Garcia, R. (2015). Contributions to empirical analysis of keystroke dynamics in passwords. *Pattern Recognition Letters*, 52:80–86.

Montalvão Filho, J. R. and Freire, E. O. (2006). On the equalization of keystroke timing histograms. *Pattern Recognition Letters*, 27(13):1440–1446.

Shadman, R., Wahab, A. A., Manno, M., Lukaszewski, M., Hou, D., and Hussain, F. (2025). Keystroke dynamics: Concepts, techniques, and applications. *ACM Computing Surveys*, 57(11):1–35.