Exploring the Effect of Music on User Typing and Identification through Keystroke Dynamics

Lukas Mecke
LMU Munich
Munich, Germany
University of the Bundeswehr
Munich, Germany
lukas.mecke@ifi.lmu.de

Assem Mahmoud
German University in Cairo
Cairo, Egypt
assem.saleh@student.guc.edu.eg

Simon Marat
LMU Munich
Munich, Germany
simonmarat@hotmail.de

Florian Alt
LMU Munich
Munich, Germany
University of the Bundeswehr
Munich, Germany
florian.alt@ifi.lmu.de

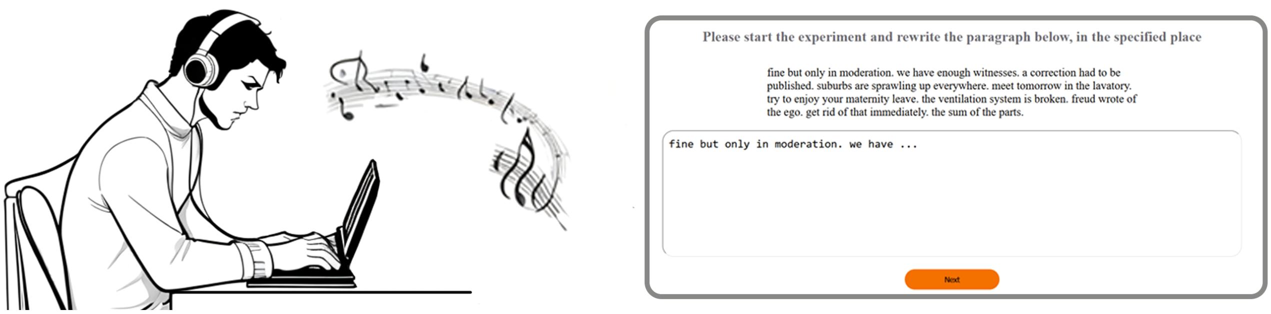

Figure 1: We explore the effect of music on (identification from) typing on a physical keyboard in an online study (N=43) that varied the tempo and loudness of music played during a text reproduction task (right) and analyzed participants' keystrokes.

Abstract

This paper explores the relationship between music and keyboard typing behavior. In particular, we focus on how it affects keystroke-based authentication systems. To this end, we conducted an online experiment (N=43), where participants were asked to replicate paragraphs of text while listening to music at varying tempos and loudness levels across two sessions. Our findings reveal that listening to music leads to more errors and faster typing if the music is fast. Identification through a biometric model was improved when music was played either during its training or testing. This hints at the potential of music for increasing identification performance and a tradeoff between this benefit and user distraction. Overall, our research sheds light on typing behavior and introduces music as a subtle and effective tool to influence user typing behavior in the context of keystroke-based authentication.

CCS Concepts

• **Security and privacy** → **Biometrics**; *Usability in security and privacy*; • **Human-centered computing** → *Text input*; **Empirical studies in HCI**.

Keywords

security, typing, keystroke dynamics, music

ACM Reference Format:
Lukas Mecke, Assem Mahmoud, Simon Marat, and Florian Alt. 2025. Exploring the Effect of Music on User Typing and Identification through Keystroke Dynamics. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26-May 1, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3706598.3713222

1 Introduction

In today's digital age, technology users enter text constantly. Interacting with keyboards forms an important part of our daily activities. This includes activities in our private lives like writing messages to friends and family but it is also a prominent part of many people's day-to-day work routine, including tasks like content generation or programming.

We can also leverage typing for authentication by focusing on the written content or how a user types. Approaches using written content include passwords, PINs, and passphrases and are based on the premise that a shared secret is only known to the legitimate user. As such, they require an additional *explicit* step (i.e., entering the secret) by the user to be authenticated. In contrast, we can also use typing *implicitly* by leveraging the unique way a person enters text. This approach is called keystroke dynamics or typing biometrics. It uses features of typing behavior like the hold time of keys, the flight time between inputs, or the pressure of a key press to find patterns unique for a user [4, 6, 36, 37, 39]. Compared to explicit authentication, this technique can be executed continuously (e.g., running in the background while a user completes a writing task).

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713222

We can use this to provide an additional layer of security after login or to reduce user annoyance by replacing the secondary task of upfront authentication [38].

Factors internal (e.g., emotion) and external to the user (e.g., new keyboard) influence people's typing. The reliability of keystroke dynamics may be affected by both external and internal stimuli altering typing behavior. Related work found internal factors, including the users' emotional state [18], intentional changes in typing [17, 22] or natural changes over time [41] (also called behavioral drift). External factors might be injuries or changes in the input hardware. For users of keystroke dynamics, that can mean false rejects and loss of access to the protected app or device.

One particularly interesting external stimulus that could influence typing is music, both due to its inherent properties like rhythm and tempo as well as the potential emotional response it can provoke. Furthermore, listening to music is common across both writing- and non-writing-focused jobs[1], making it highly relevant in many peoples everyday lives. In previous work, music was shown to increase performance [35] and attention [14] as well as leading to adjustments in working speed for physical activities [28, 35]. Those findings suggest the possibility of an effect on recognition through a keystroke dynamics system (e.g., through changed typing speed). Yet, the specific influence of music on typing patterns and the subsequent implications for biometric recognition remain largely unexplored.

A better understanding of the impact of music on keystroke dynamics has the potential for graspable applications to increase the recognition performance and, thus, the security of this approach. Currently, enrollment in a keystroke dynamics model (i.e., the initial step of learning about a user's unique typing patterns) happens in silence or in uncontrolled conditions. It remains unclear if and how playing music might affect recognition performance later, considering that many users may be listening to music during everyday use. We also hypothesize that playing similar music could elicit similar typing behavior, implying that playing the same song again during authentication might improve performance and thus contribute to security. As a final example, including factors like the tempo and loudness of music played in a model used for recognition might allow for a more dynamic representation of user typing under different conditions and thus could make recognition more robust.

The aim of this paper is to lay the foundation for such applications by gaining a deeper understanding of the effect of music on typing. We investigate the impact of tempo and loudness of music played on measures of typing behavior and compare the performance of a biometric recognition model when trained and tested under different music configurations. We complement this research by investigating user experience and preferences for the different music settings. To this end, we conducted an online study with 43 participants and gave them a text reproduction task which was accompanied by different configurations of tempo and loudness for a fixed music track. We repeated the procedure in two sessions at least three days apart to gather a test set for identification.

We found that participants made more typing errors when listening to music. Faster music increased typing speed by reducing flight time. Hold time remained stable regardless of music exposure. Loud fast music was perceived as distracting and participants generally preferred quiet music for typing. However, training and testing an identification model under different music conditions yielded improved results. We complement our results by discussing practical implications for (adaptive) biometric systems and recommendations for applying our findings in future user interfaces.

Contribution Statement. The contribution in this paper is twofold: We 1) provide experimental evidence of the effect of music on user typing behavior and recognition through a biometric model based on an online study (N=43). We 2) derive practical implications and trade-offs and summarize them in opportunities for practitioners and future research.

2 Related Work

Here, we give some background on authentication, biometrics, and keystroke dynamics in particular before introducing related work on possibilities to influence typing behavior and the specific influence of music on typing.

2.1 Authentication and Keystroke Dynamics

Authentication is essential in today's world to protect sensitive data, memories, and belongings both in the personal and professional environment. According to O'Gorman [30], authentication can be categorized into three approaches: knowledge-based authentication using a secret like a PIN or a password, token-based authentication characterized by the possession of an object like a smartcard, and biometric authentication levering unique characteristics in users' physiology or behavior. The advantage of biometrics is that they cannot be lost or forgotten, and authentication can often be done in the background without active user involvement.

In this work, we focus on keystroke dynamics, a form of behavioral biometrics that leverages typing patterns for authentication. This approach is well explored in related work and was applied both to typing on physical keyboards [26, 27] and mobile phones with and without physical keys [3, 5, 6, 10, 15, 16, 21, 39, 42]. Two surveys by Teh et al. [36, 37] give a good overview of the field.

The most common features used for keystroke dynamics are key hold times (time from pressing to releasing a key) and key flight time (time between releasing a key and pressing the next) [36]. In addition to those temporal features, soft keyboards (like on a mobile phone) enable the use of spatial features like touch area or touch to key offset which has been shown to increase biometric performance [5, 6].

In addition to the features used, keystroke dynamic systems can be classified by their input into either fixed-text or open-text systems [32]. Fixed-text keystroke dynamics recognize users based on their behavior when entering a given sequence and thus are most suited as an additional layer of security for passwords. In contrast, open-text systems are independent of content. They thus are suited for continuous authentication, for example, to ensure that a certain person indeed writes a text or that an unlocked device was not compromised after the initial login [36]. However, this can come with increased implementation effort and decreased recognition performance due to the more challenging underlying problem [32].

[1] https://cloudcovermusic.com/research/music-at-work-research, last accessed March 11, 2025

2.2 Influencing Keyboard Interactions

There are many approaches for influencing typing on a keyboard suggested in related work, including the use of visual and auditory cues [9] or haptic feedback like vibration [19] to communicate (un)desired actions like confirming data transmission on an unsecured website. Other work employed techniques to change the tactile sensation of pressing a key through ultrasonic sounds [11, 12, 25] or changes to the key's perceived structure and stiffness [12, 24].

While the presented approaches were not designed to influence typing behavior they show a broad variety of options to do so. In contrast, Hoffmann et al. [13] could actively show a reduction in typing errors by increasing the resistance of keys that would lead to erroneous input. Similarly, Mecke et al. [23] could show that key press timings could be manipulated through controllable electromagnets under the keyboard exerting force on a permanent magnet attached to the user's finger.

Despite the broad range of approaches to influence typing on physical keyboards and their potential beyond this use case, all share a common drawback: they require (extensive and expensive) setups, rendering them impractical in most daily scenarios. Here we see the potential of music to achieve similar effects without a dedicated setup as it is already a part of people's daily lives.

2.3 Impact of Music on (Typing) Performance

Music can influence humans in many different ways. According to Chamorro-Premuzic and Furnham [8], there are three main uses of music: it is leveraged to evoke an *emotional* response, *cognitively* consumed with a focus on structural and technical aspects or listened to for enjoyment in the *background* during other tasks.

Based on those uses, Sanseverino et al. [33] found emotional music to increase job satisfaction and performance, while listening in the background negatively impacted those factors. Lee et al. [18] also investigated emotional audio cues and found high emotional arousal linked to shorter key hold times. Huang and Shih [14] investigated the effect of background music on worker performance and found it to decrease their attention. Their results suggest a personal and emotional component in this effect with stronger negative impacts if workers liked the presented music.

Beyond its uses, music can also be described by its features like tempo, volume, or content of vocals. Bramwell-Dicks et al. [2] found that music containing vocals can decrease typing performance due to its distracting nature, particularly at high loudness levels. Several papers found faster music speeds to lead participants to adjust their performance and increase their speed in physical (non-typing) tasks [28, 35]. However, this effect also held true for decreasing the speed of music, hinting at a strong connection between music speed and task execution [35].

2.4 Summary

There are numerous studies on the impact of music on human performance, with some of them also focusing on typing. While some effects on typing features like speed were shown, to the best of our knowledge, no related work investigated the effects of music on typing behavior in the context of its impact on recognition through a keystroke dynamic model. In this paper, we close this gap.

3 Study Design & Modeling

With our study, we provide a better understanding of the effect of music on typing text and, as a consequence, on identification through a biometric model based on keystroke dynamics. At the same time, we also investigate potential effects on the user beyond the way they are typing, like the errors made and their perception during the task. From those goals, we derived the following three research questions that we aim to answer in this paper:

- **RQ1** How do the tempo and loudness of music affect hold time, flight time, and error rate during typing?
- **RQ2** How do the tempo and loudness of music impact recognition through a biometric model?
- **RQ3** How do the tempo and loudness of music impact the user experience during typing?

To answer our research questions, we designed a text copying task that participants had to complete while listening to music. We chose a copying task rather than a text production task to remove thinking time as an impact factor on typing. Note that music is very complex and has numerous features that could be investigated concerning typing. In this paper, we focus on the tempo and loudness of music for two reasons: both are unambiguously quantifiable and can be continuously manipulated (compared to, e.g., genre or mood).

3.1 Variables

We followed a within-subject repeated measures design with two independent variables: We varied the TEMPO of music played while typing on two levels: *slow* (60 bpm) and *fast* (160 bpm) following the values used by Nittono et al. [28]. As a second factor, we also varied LOUDNESS of music with two levels. Participants set both volumes with the cues of choosing a volume for background music (*quiet* condition) and one for active music listening (*loud* condition). We chose those prompts to reflect the cognitive and background use of music introduced by Chamorro-Premuzic and Furnham [8]. We did not capture slider settings (beyond the study situation), as they may map to very different absolute volumes depending on participants' hardware. We always started with a baseline condition without music and randomized the order of all other conditions. We repeated this procedure in two SESSIONS at least three days apart to provide realistic test data for our identification analysis.

As dependent variables, we measured users' *typing behavior* in the form of timestamped key events. We captured *participants' perception* of each typing task regarding the music they heard through Likert ratings and open questions.

3.2 Procedure

Our study procedure is illustrated in Figure 2. We invited participants to two identical sessions. At the start of each session, they were informed about the study and their rights and could consent to the procedure before providing demographic data. In the next step, participants were asked to calibrate their audio by moving a slider to select a loudness level for loud and quiet music (see Section 3.1). To ensure consistent audio during the study, we asked participants to wear headphones and not change their audio levels after this point. After the calibration, the main part started with the baseline text reproduction task without music. This was followed

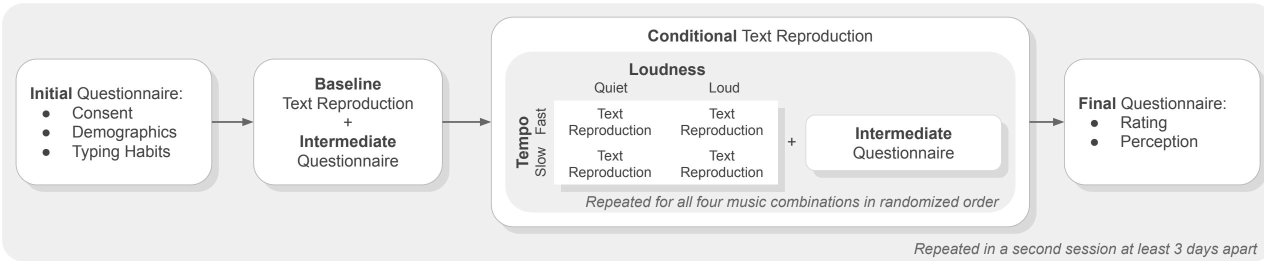

Figure 2: Our online experiment was structured as three questionnaires around a main text reproduction task. The experiment started with an *initial* questionnaire on demographics and typing habits and concluded with a *final* questionnaire comparing the different music conditions. In the main part of the study, participants first executed a baseline text reproduction task with no music. Afterward, they executed the text reproduction task while listening to music at different TEMPO and LOUDNESS at two levels in a repeated measures design in randomized order. Each task was followed by an *intermediate* questionnaire on participants' perception of the given music condition. The procedure was repeated in a second session at least three days later.

by four tasks with the combinations of loud/quiet and fast/slow music in random order. After each task, we asked for participants' perceptions. The study concluded with general questions about participants' music listening preferences.

3.3 Apparatus

To facilitate our study design, we built a website to present both the questionnaires and the text reproduction tasks. We decided on a custom solution to have complete control over the music playback and logging of key events for the keystroke dynamics analysis. The paragraphs of text used in the tasks were constructed from 8–10 sentences (about 290 characters per task) from MacKenzie and Soukoreff's phrase sets [20] each. We chose them as they are short and easy to remember, making them interchangeable. This also counteracts biases in typing like hesitations due to thinking breaks or having to enter challenging words. We constructed five paragraphs that participants received in a random order (independent of the conditions). To ensure the text was actually entered, we required a Levenhstein distance below 5 for participants to be able to continue with the experiment.

3.4 Stimulus

Instead of using different songs for the different tempos, we opted for a single piano piece[2] without vocals. Our goal with this choice was to keep the music as neutral as possible to remove biases such as an emotional connection [14, 18, 35] or an influence on typing performance due to vocals [2] as much as possible. However, we do acknowledge that music is very complex, and thus the exact effect on any single participant is arguably impossible to predict. More work on different pieces of music and/or more standardized sounds (like [40]) would be needed to better understand this area. We both increased and decreased the natural song's tempo to reach our target of 160 bpm and 60 bpm, respectively. We chose those tempos to align with the experiment by Nittono et al. [28]. We used pitch correction to remove artifacts of this manipulation and verified the sound quality through internal testing.

[2]The full piece used in our study is available at: https://pixabay.com/music/beats-romantic-piano-background-music-for-short-video-vlog-blog-1-minute-193855/, last accessed March 11, 2025

Table 1: Demographics of the participants of our online study.

	N = 43		
Gender	20	(47%)	Female
	21	(49%)	Male
	2	(4%)	Prefer not to say
Age	36	(13)	Mean (SD)
	19-65	(31)	Range (Median)

3.5 Recruitment and Participants

We recruited a sample (balanced for gender) of 100 US-based participants through the platform Prolific[3]. We had to exclude a large portion of those participants for various reasons: based on the collected keystroke data, we found that some participants copied (parts of) the text (using drag and drop, thus not generating keystrokes), used special (control) keys that lead to corrupted keystroke data, or tampered with the volume during the study. A large portion did not show up for the second session of our study, even though this was clearly communicated as essential for our experiment. After this pre-processing step, we were left with a sample of 43 participants (female: 20, male: 21, prefer not to say: 2). Participants were between 19 and 65 years old with a mean age of 36 (see Table 1).

Most participants were current undergraduate students (25) or had acquired an academic degree (8). Ten participants were following a non-academic profession. Participants estimated their time using keyboards at 25 hours per week (median = 20, SD = 22). In response to our initial Likert statements, they slightly agreed to be generally fast typists and sometimes make errors while typing. Still, they remained neutral about always listening to music when using the keyboard and finding music helpful for concentration.

Participants were compensated with £2.25 for an estimated completion time of 15 minutes (i.e., at £9 per hour) in each session. Our institute's ethics committee approved our study (EK-MIS-2023-192).

[3]https://www.prolific.co/, last accessed March 11, 2025

4 Results

Following our research questions, we analyzed our data concerning three main aspects: 1) effects of music on typing features, 2) effects of music on recognition between sessions, and 3) user perception of the conditions and their performance thereunder. If not stated otherwise, we tested for significance using repeated measures ANOVA with Greenhouse-Geisser correction where necessary and Bonferroni-corrected post-hoc tests. Due to group sizes above 5 [29] and ANOVA's robustness against non-normal distributed data [1, 31] we assumed normality and omitted respective tests. Our report focuses on significant results at an alpha level of $p < .05$.

4.1 Effects of Music on *Typing Metrics* (RQ1)

To answer our first research question, we first analyze how the tempo and loudness of music impact typing metrics. As typing metrics, we chose flight and hold times (commonly used in biometric models) and error rates as a measure of distraction. In our study, we captured keystroke data through key events. From those, we calculate hold time as the time between pressing and releasing a key. Flight time is the time between releasing a key and pressing the next. Note that this value can be negative when multiple keys are pressed at once. Finally, we calculate the error rate as the number of presses of the 'backspace' key. We excluded keystrokes where one of the measures deviated more than three standard deviations from the mean to remove outliers in the dataset (e.g., due to long breaks between key presses).

4.1.1 Presence of Music. We tested for the effect of the *presence of any music* by introducing a dummy variable combining the data of all conditions where music was played. We then conducted a repeated measures ANOVA with the presence of music and the session as factors. We did not find an effect of the presence of music on either hold time ($p = .199$) or flight time ($p = .401$). However, we found a significant increase in errors ($F(1,42) = 4.402$, $p = .042$), with an additional 1.549 errors being made when exposed to music. In addition, we observed that flight times in the second session were significantly shorter by 7.208 ms ($F(1,42) = 10.771$, $p = .002$).

4.1.2 Configuration of Music. Next, we tested for the *effect of music tempo and loudness* on the collected measures of typing behavior. Figure 3 gives an overview of the results. We conducted repeated measures ANOVA tests with tempo, loudness, and session as factors. We did not find any effects of loudness ($p = .183$), tempo ($p = .055$), or session ($p = .407$) on *hold time*. When analyzing *flight time*, we found a significant effect of both session ($F(1,42) = 5.499$, $p = .024$) and tempo ($F(1,42) = 5.181$, $p = .028$). Flight times were shorter by 4.901 ms in the second session and by 3.803 ms when exposed to fast music. We did not find an effect of music loudness ($p = .987$). Finally, we did not find an effect of either loudness ($p = .550$), tempo ($p = .323$), or session ($p = .524$) on *error rates*.

4.1.3 Summary. Participants made more typing errors when exposed to music. Listening to faster music leads to increased typing speed in the form of reduced flight time. Contrary to our expectation, we found that participants were typing faster in the second session, possibly due to familiarity with the task. We found no effects on hold time (neither through testing nor visual inspection), hinting at this measure being stable when exposed to music.

4.2 Effects of Music on *Identification* (RQ2)

There are a plethora of different models that can be used for identification through keystrokes. We decided on a random forest classifier in this scenario as a well-established and explainable model that is outlier tolerant and requires minimal pre-processing or assumptions about the data [34]. We chose to train our model on keystrokes from all participants collectively (identification/classification approach) rather than treating each participant in isolation (verification/anomaly detection approach). This method potentially offers greater discriminative power by enabling the model to learn distinguishing features across users, instead of focusing solely on individual patterns. [7]. However, it is reliant on the availability of such data from other users and thus not suitable in all cases.

For our analysis, we removed all keystrokes for letters not present in all paragraphs of text. We did so to avoid bias to the model through additional or missing keys, for example, due to participants mistyping or letters not being present in a paragraph.

Based on this dataset of 135,114 keystrokes, we built random forest classifiers with default parameters and 100 estimators. We used the keystrokes provided in the first session and evaluated the models on data collected in the second session. Note that we do not optimize our models as this analysis does not aim to achieve competitive recognition performance but rather to uncover effects on the performance induced by our study conditions. While we expect trends found in our analysis to transfer to more optimized as well as fundamentally different models (e.g., optimized for temporal data), this needs to be confirmed in future work.

To test the effects of the music played on identification, we trained several random forest classifiers. We used the hold time, flight time, and the key pressed as features. We trained separate classifiers for 1) each combination of tempo and loudness, 2) for the baseline of no music, 3) for all conditions with music playing (i.e., all conditions except the baseline), and 4) for all conditions at once. Models were trained on the first and tested on the second session.

For a more stable estimate of identity, we make predictions based on the whole paragraph of text rather than single keystrokes. We take the most predicted class based on all keystrokes within a paragraph as the prediction for that paragraph. To account for randomness in training, we report the mean F1 score (i.e. the harmonic mean of precision and recall) over 10 random forest executions throughout this section. Table 2 gives an overview of the results.

4.2.1 Presence of Music. Analogously to the previous analysis we first tested for the effect of the *presence of any music*. We conducted a repeated measures ANOVA on the F1 scores with the training (first session) and testing (second session) configuration as factors and three levels each: the baseline (i.e. no music), music (i.e. all conditions except the baseline) as well as all data for comparison. Results show a significant effect of the *training configuration* ($F(2,18) = 111.757$, $p < .001$) with Bonferroni corrected post-hoc tests confirming F1 scores being significantly higher by 0.109 when training with the music configuration ($p < .001$) and 0.094 when training on all data ($p < .001$) compared to the baseline respectively. Similarly, we observed an effect of the *testing configuration* ($F(2,18) = 662.670$, $p < .001$) on the F1 score. F1 scores were significantly higher by 0.175 in the music configuration ($p < .001$) compared to the baseline. When testing with all data, this difference was 0.190 ($p < .001$).

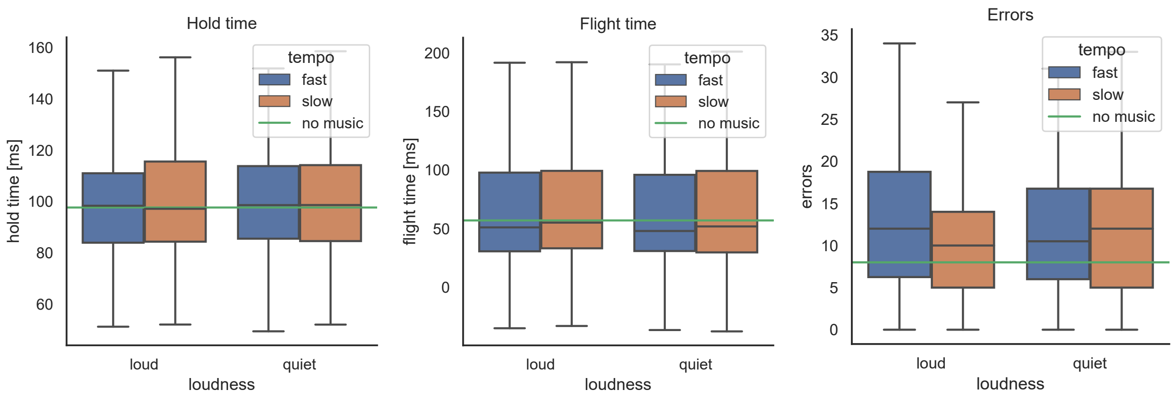

Figure 3: Flight time, hold time, and error rate in relation to tempo and loudness of music played. The median baseline values (no music) are indicated by the horizontal green lines. All values are averaged over both sessions.

Table 2: Overview of the results of the random forest identification analysis depending on the TEMPO and LOUDNESS of music played. Rows denote the configuration under which the model was trained in the first session, and columns show the configuration against which it was tested on data from the second session. Each cell represents the mean of 10 models with the standard deviation given in brackets. The *music* condition includes all configurations except for the baseline. The *all* condition includes all data from the respective sessions. Column-wise maximum and minimum values (i.e., best/worst training configurations for each test case) are marked in bold.

TESTING CONFIGURATION

TRAINING CONFIGURATION: **loudness**	**tempo**	loud fast	loud slow	quiet fast	quiet slow	music	baseline	all
loud	fast	0.747 *(.041)*	0.697 *(.036)*	**0.610** *(.024)*	0.745 *(.027)*	0.862 *(.026)*	0.651 *(.032)*	0.889 *(.027)*
	slow	0.717 *(.025)*	0.732 *(.027)*	0.759 *(.046)*	0.737 *(.034)*	0.848 *(.027)*	0.682 *(.039)*	0.856 *(.023)*
quiet	fast	0.747 *(.049)*	0.697 *(.037)*	0.697 *(.033)*	0.656 *(.045)*	0.853 *(.021)*	0.736 *(.063)*	0.863 *(.031)*
	slow	**0.673** *(.029)*	0.728 *(.034)*	0.728 *(.039)*	**0.643** *(.032)*	0.862 *(.021)*	**0.575** *(.035)*	0.843 *(.021)*
music		**0.773** *(.041)*	0.786 *(.025)*	0.797 *(.021)*	0.796 *(.035)*	**0.943** *(.010)*	**0.741** *(.028)*	**0.943** *(.010)*
baseline		0.736 *(.023)*	**0.622** *(.049)*	0.728 *(.035)*	0.661 *(.030)*	**0.798** *(.033)*	0.664 *(.032)*	**0.836** *(.029)*
all		0.748 *(.039)*	**0.806** *(.026)*	**0.800** *(.016)*	**0.819** *(.029)*	0.921 *(.028)*	0.731 *(.027)*	0.927 *(.026)*

4.2.2 Configuration of Music.

As a second step, we tested for differences induced by the configuration of music, i.e. the different levels of loudness and tempo tested in the study. We conducted a repeated measures ANOVA on the F1 score with the training loudness and tempo as well as the testing loudness and tempo as factors. Each factor had two levels (loud/quiet or fast/slow respectively). We found significant effects of training loudness ($F(1,9) = 10.446$, $p = .010$) and tempo ($F(1,9) = 10.977$, $p = .009$) as well as testing loudness ($F(1,9) = 17.384$, $p = .002$). We found no effect on the testing tempo ($p = .442$). Post-hoc tests revealed a significantly increased F1 score by 0.022 when training with loud music ($p = .010$) and by 0.015 when training with slow music ($p = .009$) respectively. The F1 score was positively influenced by 0.020 when testing with loud music ($p = .002$).

4.2.3 Identical Testing Configurations.

Beyond the effects of particular music conditions, we hypothesized that training and testing under similar conditions would increase recognition performance. Here, we test this assumption. We conducted a repeated measures ANOVA on the F1 score with similarity of tempo and similarity of loudness as factors. Each factor had two levels (identical or different conditions). Results show no effect of the similarity of tempo ($p = .487$). However, we found a significant effect of the similarity of loudness ($F(1,9) = 7.441$, $p = .017$) with post-hoc tests showing a decrease of 0.010 in F1 score when training and testing using identical loudness. As we were explicitly interested in completely identical or completely different conditions we investigated interaction effects as well but found no significant effect ($p = .120$).

4.2.4 Best and Worst Configurations.

As a final step, we visually inspected the identification results for interesting patterns. Minimum and maximum F1 scores are marked in bold in Table 2 for reference. The single *best* identification performance was achieved with a mean F1 score of 0.943 when training and testing on the music condition (i.e. on all data but the baseline). More generally, training on either all data or the music condition yielded the best results. The combination of training on quiet and slow music and testing on the baseline yielded the single *worst* performance with a mean F1 score of 0.575. Generally, the weakest results were produced when training on this condition or the baseline.

4.2.5 *Summary.*

Regarding identification performance, we observed very strong positive effects of the presence of music both during training and testing. The combination of slow and quiet music was similar to our baseline of no music and both yielded overall the weakest results. However, we found a positive effect of training with slow music in general. Both training and testing with loud music impacted performance positively. Against our expectation, we did not find positive effects of training and testing under similar conditions. On the contrary, having differing loudness levels when training and testing yielded slightly better results.

4.3 User Perception (RQ3)

We assessed participant perception during the study through an initial questionnaire on demographics and habits, intermediate questionnaires on the specific configuration they had just experienced, and a final questionnaire on their overall experience and preferences. Here we give an overview of the results from those questionnaires.

4.3.1 *Ratings of Conditions.*

After each typed paragraph, participants had to rate three 5-point Likert statements about their *experience with the specific condition.* For this analysis, we focus on answers from the first session to capture participants' first impressions. Figure 4 shows the detailed answers. The statement 'This track helped me *concentrate* more.' was overall rated neutrally, while participants disagreed with the statements 'I would *prefer* to type with this type of music in the background in the future.' and 'I felt *relaxed* while typing with this music track.'.

We tested for the effects of the conditions on those ratings using the Kruskal-Wallis test. We did not find an impact on the ability to concentrate ($p = .068$) but observed significant effects on the reported preference ratings ($\chi^2(4) = 11.136$, $p = .025$) and participants' feeling of relaxation ($\chi^2(4) = 37.899$, $p < .001$). Bonferroni corrected post-hoc Dunn tests revealed that participants felt significantly more *relaxed* in the baseline condition compared to both fast loud ($Z = 5.116$, $p < .001$) and fast quiet music ($Z = 3.128$, $p = .018$). Both fast loud music ($Z = -5.106$, $p < .001$) and fast quiet music ($Z = -3.105$, $p < .019$) were perceived as significantly less relaxing than slow quiet music. Participants *preferred* the baseline condition over the combination of fast and loud music ($Z = 2.992$, $p < .028$)

4.3.2 *Comparative Ratings.*

In the final questionnaire, we asked participants to *compare the conditions* they had experienced and report on their overall perception. Participants rated high loudness (53%) as the *most difficult condition to type in* followed by high tempo (28%). Slow tempo (9%), low loudness (7%), and no music at all (2%) were only rated as the most difficult by a few participants. Analogously, participants ranked loud fast music as the *most distracting configuration.* Fast conditions were ranked as more distracting and within those groups, loud music was ranked more distracting than quiet music. The baseline was the least distracting.

In response to general 5-point Likert statements about their *perception of the music* participants rated both increased music tempo (median = 2) and loudness (median = 2) as detrimental to their concentration and were neutral about music improving their typing experience (median = 3).

4.3.3 *Summary.*

Across multiple ratings, we observed that participants perceived loud fast music as the most distracting and found it hardest to type and concentrate while listening to it. Analogously, conditions were rated increasingly relaxing with decreasing loudness and tempo of music played. Participants were neutral about music improving their typing experience and preferred not to listen to the presented conditions while typing in the future.

5 Discussion

In this paper, we conducted an online experiment to explore the impact of music on typing features and identification through a biometric model. Here we summarize and discuss our findings. We reflect on the limitations and challenges of our work and propose research opportunities and application areas based on our findings.

5.1 Impact of Music on Identification and Typing

As the key contribution of our work, we gathered empirical evidence for the impact of music on typing behavior and identification performance through keystroke dynamics. Previous work found increases in music tempo to lead to faster task performance [28, 35]. We confirmed a similar effect for typing behavior with faster music leading to faster typing speeds. Other than Lee et al. [18], we did not observe an effect of the music played in our experiment on key hold times. However, in contrast to their study, music in our experiment was chosen with the goal of being neutral instead of emotional which they found to be connected to the influence on key hold times. Huang and Shih [14] found an effect of background music played on listener attention scores based on them having strong feelings about the music. While their results also suggest a general effect of the presence of music, they did not find statistical evidence for this. Our findings complement their research and support the notion of such a more general effect of music on attention. This is evidenced by increased errors made and decreased reported concentration when listening to music in our study.

With regards to identification, we revealed a strong positive effect of the presence of music on performance, both when it was present during training and during testing. The best performance was achieved when both training and testing on data collected while participants were listening to music. If music was only present during testing, this increased performance as well. Thus, music seems to both *provoke and emphasize unique features in user typing behavior.* That said, slow quiet music performed similarly to having no music present, which hints at the *necessity of music being consumed cognitively (i.e. actively heard)* [8] and not only in the background for the effects to occur. This is in line with our finding of slow music increasing identification when used during training, and loud music increasing identification in general.

Finally, we revealed a tradeoff between the described effects and user preferences, as loud music was rated to be distracting, and to reduce concentration. This was also evidenced by the increased error rates we observed when participants were exposed to music.

5.2 Challenges and Limitations

We only use a single simple piece of music. Our aim with this choice was to reduce external factors as much as possible, but as

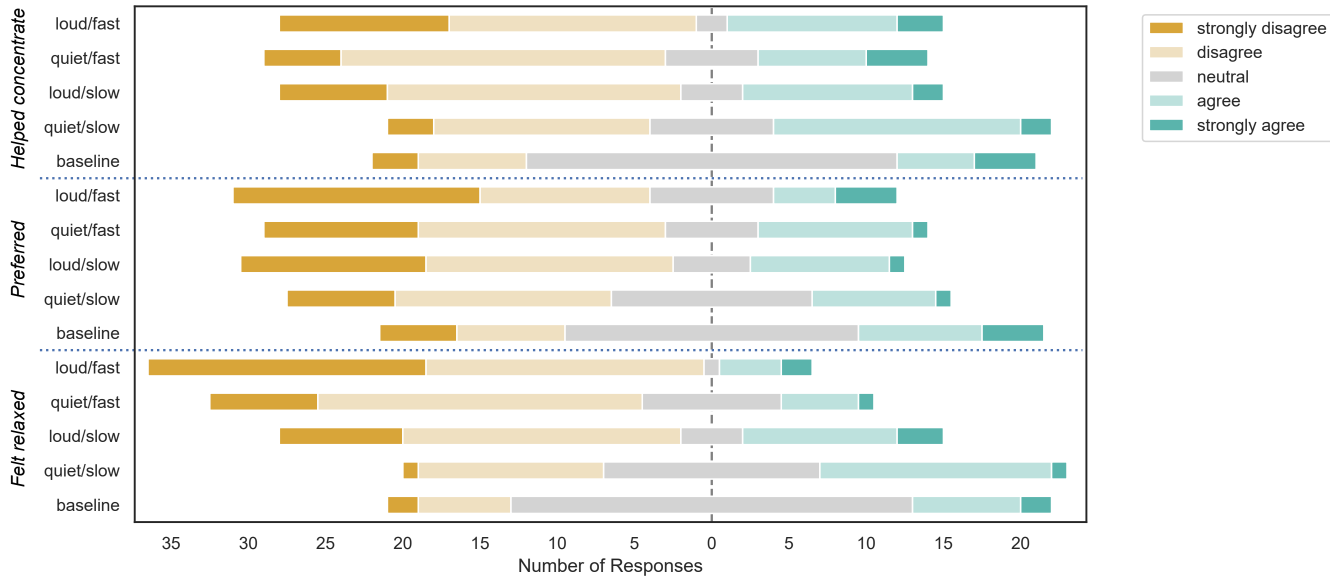

Figure 4: Participants' ratings of Likert statements (full text in Section 4.3.1) given after each condition in the first session.

a drawback, we have to assume that participants did not have an emotional connection to the piece, which was shown impactful in related studies [14, 18, 33]. Investigating a wider range of musical genres, individual differences in musical preferences, and the impact of lyrical versus instrumental music could provide a more comprehensive understanding of the relationship between music, typing performance, and keystroke dynamics. However, each choice will presumably come with other drawbacks, as features of music are hard to disentangle. As of now, we do not know if our results translate to other music. One compromise for a future study could be to allow participants the choice of a song for slow and fast music, respectively, to achieve an emotional connection and get a wide range of different features.

When conducting the statistical tests, we observed an unexpected effect of the session on flight time, implying either a habituation effect or decreased user involvement in the second session. In addition, we opted for letting users choose the appropriate audio levels for their specific hardware setup. This makes this setting very subjective and hard for us to control and further quantify. We had to exclude a few participants as we found evidence of them changing the audio levels. While we initially had considered an online study to reach a broader participant pool, those challenges call for more long-term data collection under more controlled conditions. Future work could repeat our study in a lab environment using the same hardware across all participants. This would also allow for a deeper analysis of the potentially different effects of absolute volume and perceived loudness of the music. Adding a session might help to better understand the effects we observed. However, this solution comes with its own challenges of harder participant recruitment, unfamiliarity with the hardware, and potential observation bias reducing ecological validity.

Finally, we acknowledge, that the *music* and *all* configuration in our identification analysis were combined from other conditions and thus contained more samples. This might have biased results, as, for example, more training data can by itself lead to better model performance. However, we also observed the positive impact of training with music against the baseline in the non-combined conditions, and having more samples in the test dataset should not have an impact on the results. We are thus reasonably confident that our results hold true.

5.3 Opportunities for Security and Privacy Research and Applications

As of now, keystroke dynamic systems are often trained in silence or under unknown conditions. Our results show that music is a factor that should be considered to improve, better understand, and more accurately capture the real-world use of future systems. Here we propose opportunities for security and privacy research and applications of our findings.

5.3.1 Enhancing Keystroke Dynamics with Music. Our results show, that it can be largely beneficial for identification performance to add music, even if it was not present during training. Considering this factor during training can further increase performance. However, music can also be distracting and provoke errors. Thus, its use is a tradeoff between usability and security and should be considered based on the security requirements of the specific scenario.

5.3.2 Increasing Robustness to Music. In our study, we did not observe an effect of music on hold time. While this does not imply that there is no effect, it prompts the more general question as to which factors of typing are influenced by which factors of the music played. For the case of hold time as an example, related work implies that the emotional connection may play a role [18]. For flight time, we found music tempo as an influencing factor. Overall, this suggests that certain features of keystroke dynamics may be more resilient to different auditory influences. A better understanding of those connections could have important implications for the development of more robust user identification.

5.3.3 Adapting Biometrics to the Environment. We observed varying identification performance depending on both the training and testing configuration. We see this as an opportunity for varying degrees of adaption to improve identification. When the expected authentication environment is known, our results (compare Table 2) give a starting point for finding the best training configuration. If the current music can be sensed, this could be done in real-time, for example by choosing an appropriate model from a pre-trained ensemble. Finally, it might be an option to actively play or adapt music to improve identification accuracy in high-stakes situations.

5.3.4 Avoiding Recognition. We mainly focused on improving model performance to enhance security. However, it may also be desirable to avoid recognition to protect one's privacy, for example when being tracked online based on typing behavior. As such, the previous point could analogously be adapted to hamper identification.

5.3.5 Personalizing Music. Participants in our study felt distracted by loud music and preferred silence in many cases. However, many people enjoy listening to music they like at a loudness they can adapt to their current task and that is comfortable for them. Thus, taking a broader approach and understanding personal influencing factors could complement our more fundamental analysis of music and its influence on typing and identification.

6 Conclusion

In this paper, we explored the relationship between music and keyboard typing behavior, focusing on how it affects identification through keystroke dynamics. To this end, we conducted an online experiment (N=43), where participants had to type text under varying music configurations. We found, that the presence of music during training and testing positively influences identification performance. Loud music had a positive effect but was perceived as distracting by participants, presenting a tradeoff between security and usability. With our work, we could more generally show that music is a factor that should actively be considered in future keystroke dynamic systems. We hope to inspire further research into the effects of music on typing and its complex interplay with authentication through keystroke dynamics.

Acknowledgments

This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 425869382 and is part of Priority Program SPP2199 Scalable Interaction Paradigms for Pervasive Computing Environments.

References

[1] C Alan Boneau. 1960. The effects of violations of assumptions underlying the t test. *Psychological bulletin* 57, 1 (1960), 49.

[2] Anna Bramwell-Dicks, Helen Petrie, and Alistair Edwards. 2016. Can listening to music make you type better? the effect of music style, vocals and volume on typing performance. In *International Community on Auditory Display*.

[3] A. Buchoux and N. L. Clarke. 2008. Deployment of Keystroke Analysis on a Smartphone. In *Australian Information Security Management Conference*. http://ro.ecu.edu.au/ism/48/

[4] Ulrich Burgbacher and Klaus Hinrichs. 2014. An Implicit Author Verification System for Text Messages Based on Gesture Typing Biometrics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 2951–2954. doi:10.1145/2556288.2557346

[5] Daniel Buschek, Benjamin Bisinger, and Florian Alt. 2018. ResearchIME: A Mobile Keyboard Application for Studying Free Typing Behaviour in the Wild. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. ACM, New York, NY, USA, Article 255, 14 pages. doi:10.1145/3173574.3173829

[6] Daniel Buschek, Alexander De Luca, and Florian Alt. 2015. Improving Accuracy, Applicability and Usability of Keystroke Biometrics on Mobile Touchscreen Devices. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 1393–1402. doi:10.1145/2702123.2702252

[7] Daniel Buschek, Alexander De Luca, and Florian Alt. 2015. Improving accuracy, applicability and usability of keystroke biometrics on mobile touchscreen devices. In *proceedings of the 33rd annual ACM conference on human factors in computing systems*. 1393–1402.

[8] Tomas Chamorro-Premuzic and Adrian Furnham. 2007. Personality and music: Can traits explain how people use music in everyday life? *British journal of psychology* 98, 2 (2007), 175–185.

[9] Alexander De Luca, Bernhard Frauendienst, Max Maurer, and Doris Hausen. 2010. On the Design of a "Moody" Keyboard. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems* (Aarhus, Denmark) *(DIS '10)*. Association for Computing Machinery, New York, NY, USA, 236–239. doi:10.1145/1858171.1858213

[10] Benjamin Draffin, Jiang Zhu, and Joy Zhang. 2014. KeySens: Passive User Authentication through Micro-behavior Modeling of Soft Keyboard Interaction. In *Mobile Computing, Applications, and Services*, Gérard Memmi and Ulf Blanke (Eds.). Springer International Publishing, Cham, 184–201.

[11] David Gueorguiev, Anis Kaci, Michel Amberg, Frédéric Giraud, and Betty Lemaire-Semail. 2018. Travelling ultrasonic wave enhances keyclick sensation. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 302–312.

[12] Chris Harrison and Scott E. Hudson. 2009. Texture Displays: A Passive Approach to Tactile Presentation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 2261–2264. doi:10.1145/1518701.1519047

[13] Alexander Hoffmann, Daniel Spelmezan, and Jan Borchers. 2009. TypeRight: a keyboard with tactile error prevention. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2265–2268.

[14] Rong-Hwa Huang and Yi-Nuo Shih. 2011. Effects of background music on concentration of workers. *Work* 38, 4 (2011), 383–387.

[15] Seong-seob Hwang, Sungzoon Cho, and Sunghoon Park. 2009. Keystroke dynamics-based authentication for mobile devices. *Computers & Security* 28, 1–2 (2009), 85–93. doi:10.1016/j.cose.2008.10.002

[16] Sevasti Karatzouni and Nathan Clarke. 2007. Keystroke Analysis for Thumb-based Keyboards on Mobile Devices. In *New Approaches for Security, Privacy and Trust in Complex Environments*, Hein Venter, Mariki Eloff, Les Labuschagne, Jan Eloff, and Rossouw von Solms (Eds.). Springer US, Boston, MA, 253–263.

[17] Hassan Khan, Urs Hengartner, and Daniel Vogel. 2018. Augmented Reality-based Mimicry Attacks on Behaviour-Based Smartphone Authentication. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 41–53.

[18] Po-Ming Lee, Wei-Hsuan Tsui, and Tzu-Chien Hsiao. 2015. The influence of emotion on keyboard typing: An experimental study using auditory stimuli. *PloS one* 10, 6 (2015), e0129056.

[19] Zhaoyuan Ma, Darren Edge, Leah Findlater, and Hong Z Tan. 2015. Haptic keyclick feedback improves typing speed and reduces typing errors on a flat keyboard. In *2015 IEEE World Haptics Conference (WHC)*. IEEE, 220–227.

[20] I Scott MacKenzie and R William Soukoreff. 2003. Phrase sets for evaluating text entry techniques. In *CHI'03 extended abstracts on Human factors in computing systems*. ACM, 754–755.

[21] Emanuele Maiorana, Patrizio Campisi, Noelia González-Carballo, and Alessandro Neri. 2011. Keystroke Dynamics Authentication for Mobile Phones. In *Proceedings of the 2011 ACM Symposium on Applied Computing* (TaiChung, Taiwan) *(SAC '11)*. ACM, New York, NY, USA, 21–26. doi:10.1145/1982185.1982190

[22] Lukas Mecke, Daniel Buschek, Mathias Kiermeier, Sarah Prange, and Florian Alt. 2019. Exploring intentional behaviour modifications for password typing on mobile touchscreen devices. In *Fifteenth Symposium on Usable Privacy and Security (SOUPS 2019)*. 303–317.

[23] Lukas Mecke, Ismael Prieto Romero, Sarah Delgado Rodriguez, and Florian Alt. 2023. Exploring the Use of Electromagnets to Influence Key Targeting on Physical Keyboards. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–8.

[24] Viktor Miruchna, Robert Walter, David Lindlbauer, Maren Lehmann, Regine Von Klitzing, and Jörg Müller. 2015. Geltouch: Localized tactile feedback through thin, programmable gel. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 3–10.

[25] Jocelyn Monnoyer, Emmanuelle Diaz, Christophe Bourdin, and Michaël Wiertlewski. 2016. Ultrasonic friction modulation while pressing induces a tactile feedback. In *International Conference on Human Haptic Sensing and Touch Enabled Computer Applications*. Springer, 171–179.

[26] Fabian Monrose, Michael K. Reiter, and Susanne Wetzel. 1999. Password Hardening Based on Keystroke Dynamics. In *Proceedings of the 6th ACM Conference on Computer and Communications Security* (Kent Ridge Digital Labs, Singapore) *(CCS '99)*. ACM, New York, NY, USA, 73–82. doi:10.1145/319709.319720

[27] Fabian Monrose and Aviel Rubin. 1997. Authentication via Keystroke Dynamics. In *Proceedings of the 4th ACM Conference on Computer and Communications Security* (Zurich, Switzerland) *(CCS '97)*. ACM, New York, NY, USA, 48–56. doi:10.1145/266420.266434

[28] Hiroshi Nittono, Akiko Tsuda, Seiki Akai, and Yoshiaki Nakajima. 2000. Tempo of background sound and performance speed. *Perceptual and motor skills* 90, 3_suppl (2000), 1122–1122.

[29] Geoff Norman. 2010. Likert scales, levels of measurement and the "laws" of statistics. *Advances in health sciences education* 15 (2010), 625–632.

[30] Lawrence O'Gorman. 2003. Comparing passwords, tokens, and biometrics for user authentication. *Proc. IEEE* 91, 12 (2003), 2021–2040.

[31] Egon S Pearson. 1931. The analysis of variance in cases of non-normal variation. *Biometrika* (1931), 114–133.

[32] Shimaa S Zeid, Raafat A ElKamar, and Shimaa I Hassan. 2022. Fixed-Text vs. Free-Text Keystroke Dynamics for User Authentication. *Engineering Research Journal (Shoubra)* 51, 1 (2022), 95–104.

[33] Domenico Sanseverino, Andrea Caputo, Claudio Giovanni Cortese, and Chiara Ghislieri. 2022. "Don't Stop the Music," Please: The Relationship between Music Use at Work, Satisfaction, and Performance. *Behavioral Sciences* 13, 1 (2022), 15.

[34] Rashik Shadman, Ahmed Anu Wahab, Michael Manno, Matthew Lukaszewski, Daqing Hou, and Faraz Hussain. 2023. Keystroke dynamics: Concepts, techniques, and applications. *arXiv preprint arXiv:2303.04605* (2023).

[35] Atillal Szabo, Adam Small, and Martin Leigh. 1999. The effects of slow-and fast-rhythm classical music on progressive cycling to voluntary physical exhaustion. *Journal of sports medicine and physical fitness* 39, 3 (1999), 220.

[36] Pin Shen Teh, Andrew Beng Jin Teoh, and Shigang Yue. 2013. A Survey of Keystroke Dynamics Biometrics. *The Scientific World Journal* 2013 (2013). doi:10.1155/2013/408280

[37] Pin Shen Teh, Ning Zhang, Andrew Beng Jin Teoh, and Ke Chen. 2016. A Survey on Touch Dynamics Authentication in Mobile Devices. *Computers & Security* 59, C (2016), 210–235. doi:10.1016/j.cose.2016.03.003

[38] Dirk Weirich and Martina Angela Sasse. 2001. Pretty good persuasion: a first step towards effective password security in the real world. In *Proceedings of the 2001 workshop on New security paradigms*. 137–143.

[39] Hui Xu, Yangfan Zhou, and Michael R. Lyu. 2014. Towards Continuous and Passive Authentication via Touch Biometrics: An Experimental Study on Smartphones. In *Symposium On Usable Privacy and Security (SOUPS 2014)*. USENIX Association, Menlo Park, CA, 187–198. https://www.usenix.org/conference/soups2014/proceedings/presentation/xu

[40] Wanlu Yang, Kai Makita, Takashi Nakao, Noriaki Kanayama, Maro G Machizawa, Takafumi Sasaoka, Ayako Sugata, Ryota Kobayashi, Ryosuke Hiramoto, Shigeto Yamawaki, et al. 2018. Affective auditory stimulus database: An expanded version of the International Affective Digitized Sounds (IADS-E). *Behavior Research Methods* 50 (2018), 1415–1429.

[41] Yafang Yang, Bin Guo, Yunji Liang, and Zhiwen Yu. 2021. Detection of Behavior Aging from Keystroke Dynamics. In *2021 IEEE 27th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 583–590.

[42] Saira Zahid, Muhammad Shahzad, Syed Ali Khayam, and Muddassar Farooq. 2009. Keystroke-Based User Identification on Smart Phones. In *LNCS*, Vol. 5758. 224–243. doi:10.1007/978-3-642-04342-0_12