That Password Doesn’t Sound Right: Interactive Password Strength Sonification

Otto Hans-Martin Lutz
otto.lutz@fokus.fraunhofer.de
Fraunhofer FOKUS
Weizenbaum Institute for the Networked Society

Jacob Leon Kröger
kroeger@tu-berlin.de
Technische Universität Berlin
Weizenbaum Institute for the Networked Society

Manuel Schneiderbauer
Weizenbaum Institute for the Networked Society

Jan Maria Kopankiewicz
Weizenbaum Institute for the Networked Society

Manfred Hauswirth
manfred.hauswirth@tu-berlin.de
Technische Universität Berlin
Weizenbaum Institute for the Networked Society

Thomas Hermann
thermann@techfak.uni-bielefeld.de
Bielefeld University, Faculty of Technology

ABSTRACT

Despite 2-factor authentication and other modern approaches, authentication by password is still the most commonly used method on the Internet. Unfortunately, as analyses show, many users still choose weak and easy-to-guess passwords. To alleviate the significant effects of this problem, systems often employ textual or graphical feedback to make the user aware of this problem, which often falls short on engaging the user and achieving the intended user reaction, i.e., choosing a stronger password.

In this paper, we introduce auditory feedback as a complementary method to remedy this problem, using the advantages of sound as an affective medium. We investigate the conceptual space of creating usable auditory feedback on password strength, including functional and non-functional requirements, influences and design constraints. We present web-based implementations of four sonification designs for evaluating different characteristics of the conceptual space and define a research roadmap for optimization, evaluation and applications.

CCS CONCEPTS

• **Human-centered computing → Auditory feedback**; *Sound-based input / output*; • **Security and privacy → Usability in security and privacy**.

KEYWORDS

Interactive sonification, password security, password strength, sonification, auditory display, auditory feedback, usable security

ACM Reference Format:
Otto Hans-Martin Lutz, Jacob Leon Kröger, Manuel Schneiderbauer, Jan Maria Kopankiewicz, Manfred Hauswirth, and Thomas Hermann. 2020. That Password Doesn’t Sound Right: Interactive Password Strength Sonification. In *Proceedings of the 15th International Audio Mostly Conference (AM’20), September 15–17, 2020, Graz, Austria.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3411109.3412299

1 INTRODUCTION

Passwords are still one of the most common forms of authentication in use today [4]. Security problems caused by choosing weak password or reusing them has significant impact on the economy and society [1]. Creating unique, strong passwords for every user account can be perceived as difficult or annoying. Even though best practices for the creation of good passwords are widely available, they are followed rarely by the general user base and the use of weak passwords persists [24, 33]. The widespread methods of textual or graphical password feedback do not exhaust the possibilities to create an engaging and motivating user experience in the process of choosing a good password [38]. Applying concepts of affective computing may help in reducing user frustration [34].

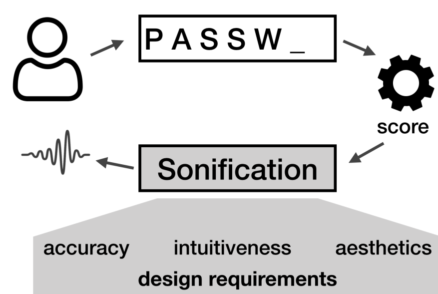

Figure 1: Fundamental interaction with the password sonification system, and selected design requirements

The problem, however, is that the usual feedback of a visual progress bar indicating the level of security is predominantly a pragmatic information, lacking emotional connotations. Colors, such as the culturally trained red-yellow-green traffic light colors do have an affective charge, but still we see opportunities for improvement using auditory (or even multimodal) feedback. It is known that experiencing emotions can nudge and support users in

AM’20, September 15–17, 2020, Graz, Austria

ACM ISBN 978-1-4503-7563-4/20/09…$15.00
https://doi.org/10.1145/3411109.3412299

forming and reinforcing new, positive habits, such as developing an individual password strategy [15]. This approach has been successfully applied by researchers to enhance information security awareness programs [5] and to motivate users to adhere to security policies, including the selection of secure passwords [15]. The use of sound as an affective medium provides a way to arouse or display emotional content [23].

Our goal in this paper is to introduce a new form of password strength feedback through interactive auditory display and to lay out a roadmap for its optimization, its embedding in target contexts, and research required for validation. We can see several partially conflicting requirements on the sonification and interaction design, a selection of which is shown in Figure 1, next to the fundamental interaction principle: While the user is choosing and entering a password, its strength score is calculated at every keystroke and immediately returned in form of either presenting a keystroke-related sound event or updating a continuously playing auditory feedback. As this approach is new, our first aim is to explore the design space, for instance by creating both emotionally neutral designs as well as designs intended to evoke an affective reaction which, in turn, can act as a trigger for cognitive reflection and may encourage playful experimentation. In particular, since positive affect has been shown to increase creativity and broaden the scope of cognition and action [12], we are guided by the hypothesis that charging password feedback with an affective quality would have desirable effects for the task at hand, both in terms of user satisfaction and password quality.

To the best of our knowledge, this is the first work on password strength feedback using interactive sonification. Our contributions are firstly, to analyze the possible conceptual space of creating usable auditory feedback on password strength, including functional and non-functional requirements, influences and design constraints (section 3), and secondly, to develop and showcase prototypical implementations of four sonification designs, each of which emphasizes different characteristics of the conceptual approach (section 4), and thirdly, to lay out a roadmap for their optimization, application and evaluation (section 5).

2 RELATED WORK

2.1 Password Security and Feedback

A key intention of password policies on websites and other points of authentication is to motivate users to choose strong passwords as they often choose simple, predictable passwords that are susceptible to attacks [8]. Leaked identities and credentials from security breaches of major service providers reveal that the re-use of passwords across different services is prevailing and constitutes an attractive attack vector [22]. When signing up to a new service, the least a new user should do is to not re-use passwords in order to increase the costs for the potential attacker and thus limit his/her vulnerability across sites as each password will require a new cracking attempt [22, 36]. A simple assessment of password strength can be done by applying the so-called LUDS requirements, i.e., counting the numbers of lowercase and uppercase letters, digits and symbols, and similar measures. Although not being very secure or usable, this method is still widely adopted [42]. Over the years, those methods have progressed from decision trees and probabilistic filters to more sophisticated algorithms which recognize and weigh in common passwords, popular names, words, and other common patterns (e.g. zxcvbn by Wheeler [42]) and machine-learning based methods (e.g. by Ur et al. [38]).

Most service providers define requirements regarding password characteristics (e.g. LUDS) and strength [36]. The way these requirements and feedback to the chosen password are presented to the user can impact the security and usability of created passwords [38]. The most common forms of feedback are binary (accepted / not accepted), graphical (stars, bars, colors), and text-based [32, 38]. Binary and graphical feedbacks do not necessarily offer users an explanation of what aspects of the password can be improved. Providing the user with more detailed feedback in the form of redundant information, e.g., bar length and color, and context-aware, textual feedback leads to better passwords [38]. To the best of our knowledge, auditory feedback on password strength has not yet been investigated.

2.2 Sonification and Internet Security

Sonification for process monitoring in security-related fields has been researched widely. There is a comprehensive body of work on using sonification for network traffic monitoring to achieve higher situational awareness in a network operations center, e.g., [3, 7, 39]. A systematic overview is provided in [2]. In this context, users are network security specialists using the auditory modality as a supplementary resource to improve pattern, anomaly and intrusion detection. For non-expert users, hidden online tracking has been sonified to convey privacy and security issues in order to raise awareness for these concerns [21, 25].

Auditory feedback on security-related interactions has been studied in an application for vision-impaired users, in the form of an earcon-based sonification of Internet security threats [37]. Here, warning sounds that convey their intended meanings with little-to-no user training, e.g., casting a fishing reel to warn about a phishing attack, were used to notify users about security threats while browsing on a screen reader.

We did not find any scientific publications on auditory password feedback. However, there is a project on representing passwords through a 'sonic hash' available on GitHub [28]. In this system, a musical representation of the password is generated. In use cases where a password has to be entered twice, e.g., when defining the password for a certain service for the first time, the system facilitates auditory clues for identifying if the two passwords are identical. Sonic hash, however, does not provide any feedback on password strength.

2.3 Affective Sonification

Like graphical visualization, sonification can be seen either as a scientific tool for data analysis and interpretation or as a different way of exploring and understanding data, where hedonic, task-unrelated qualities that engage the senses are important as well [16].

Pragmatic vs. hedonic qualities. Sonification as such is aimed at accurately representing data [9], which is reflected in and assured by the definition and criteria by Hermann [17]: The transformation of data to sound shall be objective, systematic, reproducible and applicable to different data. Mapping of features to synthesis parameters (and ultimately to perceptual qualities) is a frequently applied sonification technique. Most physical quantities are mapped to pitch, whereas the most often used mappings not involving pitch follow natural perceptual associations [9]. Sonification requires active, attentive listening in order to access the abstract information conveyed by the sound [14] and a willingness to learn the auditory symbols. As Vickers et al. state, *'[a] major challenge for sonification designers continues to be that their work is often perceived as annoying, fatiguing, or both'* [39]. In his provocatively named paper "Is Sonification doomed to fail?", Neuhoff makes the case for artistic sonification, as it *'would embrace the more aesthetic aspects of sonic representation, giving listeners a "sense" of the underlying data'* [31]. Using a larger aesthetic variety might capture attention, and motivate curiosity and exploration of the data represented by sound, as well as create a better overall user experience.

Sound and Emotion. Artistic and aesthetic choices can influence the capability of sound and music to arouse or display emotion. Even though a distinction has to be made between an emotion being aroused (e.g., a listener being actually moved by a musical piece, or fear being triggered as a brainstem reflex) and perceived (recognizing what kind of emotion is expressed), these mechanisms often co-occur [23]. Through the display of emotional content, the recipient gets an opportunity to experience the emotional impact of certain actions without having to suffer the real consequences [23]. Similarly, affective components of sound can be used in favor of the sonification objective as well.

In the literature, two types of models for emotions are most prevalent: (1) Categorical models which use qualitative descriptions of certain emotions, for example, Ekman's six universal emotions (happiness, sadness, fear, anger, surprise and disgust [10]); (2) dimensional models with quantitative values along different dimensions, for example, Russel's circumplex model of affect [35], which states a two-dimensional Cartesian space consisting of the axes arousal (vertical) and valence (horizontal). In this paper, we adopt the latter model as it is useful for research on different degrees of emotional characteristics along continuous dimensions.

Emotions function as signals in the brain that some change of priority is needed and help prioritizing conflicting goals [23, p.54]. If, for example, a user has the conflicting goals "create a strong password" and "finish the process quickly," emotions can nudge the user in either of these directions. Here, the interesting question is, how to design the emotional triggers so that they work in favor of the designer's objective, which, in our case, is the stimulation of interest, experimentation and ultimately the creation of stronger passwords.

Compared to music or chord progressions, a single sound by itself has limited, if any, emotional content. But in the context of other information, a task or additional sounds within a soundscape, a single sound can convey not only information but shape an affective reaction, which is demonstrated powerfully by the use of sound effects and music in movie productions and the accompanying sound design research [20]. One way to evoke emotions is via sounds that trigger an association to an emotionally experienced situation, and the International Affective Digitized Sounds (IADS) samples [6] are to a large extend of this type. Their duration of ≈ 6 seconds indicates that a certain time is required for the emotion to be evoked. With the Emosonics project [19], in contrast, no emotionally experienced situations were referenced, but a synthesis model covering various emotional associations was developed and refined by an evolutionary optimization approach. This system to shape the emotional quality of short sound samples was realized using a substantially shorter length of typically 2-3 seconds. In summary, evoking affective associations requires some time and – in our experience – cannot be compressed much further than a few seconds. There exists still no theoretical framework for evoking or displaying a wide variety of emotions in short sounds. Therefore, designing short, singular sounds to convey emotional load continues to be a challenge best met through an iterative approach. A balance has to be found between creating an affective, aesthetically pleasing, engaging sound and keeping the sonification functional, as aesthetically pleasing or rhythmical sounds can distract from listening analytically [14]. As Grond et al. point out: *'A sonification that works is therefore the successful struggle to create a message that points beyond the medium'* [14].

3 REQUIREMENTS AND CONCEPTUAL DIMENSIONS

Following established considerations of sonification design [41], our approach is based on fundamental questions of translating data to sound dimensions: mapping, polarity, and scaling. To meet the special requirements of interactive password strength sonification, we particularly consider their interplay with sonic aesthetics, affective sonification design principles and the influence of mental representations of password strength itself. These, partly conflicting, objectives between functional feedback requirements and sonic user experience are discussed in the following sections.

3.1 Functional password feedback requirements

One of the core components of current password strength feedback systems is a representation of the password quality or its resilience against common attacks. Visual feedback is widely used in the form of filling a bar [38], which displays a fraction of a defined maximum length and/or strength, even though password complexity or strength can theoretically increase infinitely. As the average user can be expected to be familiar with this visualization, a password strength sonification could try to mimic this concept of representation.

When creating a password for a certain service or application, service providers often define a minimum requirement regarding password strength [36]. This "good enough" threshold is an important indicator for the user, hence it needs to be conveyed through the auditory feedback as well. As password re-use is one of the most frequent password security issues [11, 33], a notification if a frequently used one is entered is also beneficial to improve understanding of password strength.

Hence, an auditory feedback for password strength should indicate the following password characteristics:

- password strength score in the form of a share or percentage
- indication when the password is "good enough"
- indication when the password is a frequently used one

3.2 Intuitiveness / self-explanatory capabilities

One requirement for new or uncommon ways of "displaying" information is intuitiveness / being self-explanatory. Instead of taking time to learn and understand a new feedback mechanism, an instant understanding of the mapping is useful [41, p. 30]. But, even in our visually-dominated world, we have to realize that intuitive understanding is based on a continuous process of familiarization with certain types of visualization: At the first introduction of a line graph, the author devoted several pages of text on how to read and understand this graph [13, p. 24].

Using auditory icons, sounds or soundscapes adapted from real environments can provide a real-world metaphor, a self-explanatory shortcut to information, but in turn might require prior knowledge or experiences, and may depend on cultural background [14]. For example, how does 'security' sound? Is it desirable to create a sound with direct reference to a particular aspect of security or is it advisable to add a layer of abstraction or generalization which does not require the correct mental reference to this one aspect? With regard to these considerations, we argue that self-explanatory capabilities should not be an outstanding priority in the sonification design. It seems reasonable to require the user to go through a short learning and familiarization phase. At the same time, providing ecologically valid mappings and supporting real-world metaphors can greatly assist an intuitive understanding of the sonification.

3.3 Aesthetics and affect

As pointed out above, for the user, the sonic aesthetics are not explicitly linked to the particular metaphor or affective quality intended by the designer. How does a 'secure', 'good' or 'bad' password sound? To what extent should the sounds contain complex aesthetic qualities and evoke or display emotions? How much arousal is beneficial to the objective, how much will distract the user?

Employing an affective display design can lead to increased acceptance and usefulness of the sonification. The emotional level adds an intuitive, subtle, possibly even unconscious component to the auditory display, therefore it increases its self-explanatory capabilities. Information can be coded in a redundant way through affective design. Complex aesthetic qualities could facilitate a sound design both motivates interest and curiosity right with the initial interactions, and will be perceived as pleasant and non-fatiguing in long-term use.

Emotionally neutral sounds, e.g., mapping to basic sound parameters such as tempo or pitch, are better in their pragmatic quality, in the clarity of mapping and scaling. They are less prone to biases of personal aesthetic preferences and are less influenced by prior experiences and cultural background. From a researcher's perspective, they are easier to study as there are less confounding factors in the design.

Comparing these two types of design paradigms along the dimensions of user acceptance and performance can lead to a deeper understanding of the influence of aesthetics and affect in auditory displays, not only in the context of password security but also regarding sonification design in general.

3.4 Perspective, metaphors and polarity of mappings

Depending on the mental depiction or metaphor, different conclusions for an appropriate sound design can be drawn. The perspective or point of view of the sonification may either be *object-centered*, representing properties of the password, or *user-centered*, from a point of user experience and desirable user state (e.g., pleasant sound). For example, one could think of a strong password as a mathematical complexity measure, which leads to an auditory display that would semantically hint at concepts such as 'complexity', 'hard to crack', 'scrambled', or 'distorted'. Following this reasoning, a weak password would therefore be represented as something simplistic or easy-to-predict. From the perspective of the user experience, the user should be rewarded for a strong password, leading to an auditory representation of 'being safe', or expressing 'calmness', or pleasantness'. Accordingly, a weak password would be represented with unpleasant or even annoying sounds. Consequential, the user would be nudged to act in a way that the sound becomes more pleasant, i.e., to choose a stronger password.

These contrasting examples illustrate that there is no single definitive solution or overall clear approach to represent password strength as auditory feedback. We expect to find varied understandings regarding an appropriate depiction of a strong password, depending on different perspectives and mental representations.

The assumptions underlying these types of sound designs are also supported by research on mapping and scaling of fundamental sonic parameters. Even with simple parameters (e.g., pitch), users could be divided in two groups with opposing perceptions on which polarity of mapping seems logical or natural to them [40]. In our case, this means a stronger password might correspond to higher pitch in one group and a lower pitch in the other.

3.5 Accuracy and scaling

Creating a sonification that is as powerful as a visualization is challenging: *'In almost all dimensions but time, the precision with which we can perceptualize data is greater in vision' than in the auditory domain [31].* Applied to password sonification, this means that to create an accurate and precise representation of the password strength score, not only the current score has to be disclosed, but additionally a frame of reference is needed. In a typical visual representation, the password bar shows both the current value (as length of the color-filled proportion of the bar) and the minimum and maximum values (the limits of the bar). In contrast, when playing a sound corresponding to the actual score, the reference minimum and maximum values are not necessarily memorized or available at the same time. Hence, to achieve a comparable auditory display in this regard, both the current value, the minimum and the maximum of the scale have to be presented to the user, for example, simultaneously or consecutively.

The transfer (or scaling) function between password score and sonification parameters has a major influence on the perceptual resolution. Using a linear transfer function provides the same resolution along the whole value range, yet only if the parameter is perceived linearly. A nonlinear function can enhance a certain range of values by changing the slope and therefore make certain values more discernible. This can be useful for emphasizing a range where little differences in score already have a big effect on the time needed to crack this password. An exponential score-to-parameter transfer function would in turn reflect the exponential relationship between password strength (measured in bits) and cracking time.

These considerations on the transfer function already highlight an important factor influencing design considerations on accuracy, precision and scaling: The task the user is to perform. The basic question is: Given the task of supporting the creation of strong passwords, how important is it for the user to perceive the exact value of the password score? The degree of accuracy that the user needs from the auditory feedback varies with task focus: Depending on task context, getting a good estimation of password quality and clearly perceiving that a certain acceptable quality threshold is exceeded may be more important for the user than recognizing the precise score. Some authors have sacrificed accuracy in favor of user experience: Hermann et al., for example, reduced the detail level of their sonification in their second design iteration to reduce repetitiveness and increase pleasantness [18].

3.6 Cold start problem

Initially, all passwords have low scores while the first few characters are being typed by the user. The first feedback sounds provide a reference to the low score. As this happens at every beginning of the interaction, a conflict arises: While starting to type in the password, immediately a 'low score' feedback is given. The user could regard or mistake that as an 'auditive punishment' and get distracted or confused. This effect needs to be avoided or mitigated to create a good user experience. Possible solutions to this cold start problem can be to choose a neutral sound until a certain score is reached or to mute the feedback for the first few characters (e.g., gradually fading in the volume). However, this also increases the complexity of the sonification and obfuscates the absolute reference to the minimal score.

3.7 Discussion of conceptual dimensions

The dimensions, objectives and constraints mentioned above are contradictory in part: For example, optimizing for scaling accuracy and precision requires clearly discernible value steps and a reference to the minimum and maximum of the scale, which impedes intuitive, self-explanatory understanding of the sonification. Profound prior training and very careful listening are necessary to interpret the audio signal correctly, which diminishes the chances for a good user experience. A design which tries to avoid the cold start problem may increase user experience, but may decrease accuracy, precision and intuitive understanding as an additional layer of complexity is introduced to the design. Thus, we will follow a design strategy in designing sonifications which balance these partly conflicting requirements.

As a result, the following overarching research questions emerge: (i) Which parameter selection and combination is best understandable intuitively? (ii) Which is the best design for continued usage? (iii) What design do users prefer? We designed and implemented prototypes which provide different combinations of design choices to gain first insights into these questions and allow us iterative refinement of the designs for future research.

4 IMPLEMENTATION

Our general approach is to calculate a password score with each character entered or removed by the user, to check if the resulting character string matches one of the top 100.000 most common passwords according to the ranking list by Miessner [29], and then to refine the auditory representation with regard to the current score. As the use case of online password selection is browser-based, we chose Web technology to implement the auditory feedback.

For password strength estimation, we use the *zxcvbn algorithm* by Wheeler [42] as it provides a well-usable password strength estimation API[1] in JavaScript and a significantly improved estimation accuracy compared to simple approaches, e.g., LUDS. For sound generation, we use *Tone.js*, a cross-platform, cross-device WebAudio framework for creating interactive sound and music applications in the browser [27]. The source code of our implementation is provided on GitHub and Zenodo: http://doi.org/10.5281/zenodo.3994278

4.1 Sonification Design

We implemented four different parameter-mapping sonification designs, using continuous for three and discrete mapping in one of the designs, i.e. on each keystroke a sound is played or modified according to the updated score. We use linear transfer functions unless stated otherwise.

For all sonification designs, there are three features additionally to indicating the password score: Signaling if the password is 'good enough' (passing a threshold value of > 7 out of 10), notifying if it is one of the most frequently used ones, and optionally a gradual volume fade-in over the first five characters to avoid the cold start problem. If the password is 'good enough', this is indicated by adding reverberation for all but one sonification designs based on distinct notes. For the *Noise to Harmonic* design based on moving, i.e., drone-like sounds, reverberation is already an integral part of the sound design. Hence, passing the 'good enough' threshold is indicated by a key change of the harmonic chord from minor to major.

Additionally to the score estimation provided by the *zxcvbn algorithm*, we check against the top 100.000 most common passwords. If one of these is used, the score is set back to zero and the sound of breaking glass is played to indicate a very high dictionary attack risk. This notification is encoded both by the sonic feedback going back to its form at a score of zero and the breaking glass sound. Figure 2 depicts all input parameters of the sonifications.

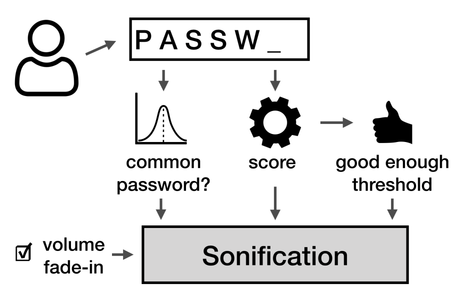

Figure 2: Input parameters of the sonification

[1] Application Programming Interface, an interface to a software component

The sonification designs can be divided in classes: Originating from a real-world metaphor vs. being synthetically designed like earcons, and holding simple vs. affective/complex sonic aesthetics. A schematic representation of the sonification designs is given in Figure 3, which depicts the pitch in relation to password strength. Videos with example interactions are provided at [26]. We created implementations of all possible combinations of these classes. Their sound design choices are detailed in the following paragraphs

Two Tones. "Simple sonic aesthetics and no real-world metaphor." This design consists of two short tones played after each key press. The first tone represents the current score, the second tone the maximum score. Thereby, both the current value and a target reference are conveyed on every key stroke. We chose an interval of a perfect fifth between the minimum (0) and maximum (10) score, and mapping linearly to pitch from A_4 to E_5, i.e. ranging from $440Hz$ to $660Hz$.

Inverse Parking Sensor. "Simple sonic aesthetics with real-world metaphor." This design resembles the sound of a parking sensor auditory display found in cars, which maps the spatial distance to an object to the rate of beeps. Here, the password score is mapped to the time interval between beeps. The transfer function between score and time interval between beeps was explicitly chosen not to be linear, but to have perceptual similarity to a parking sensor, with $\Delta t = 0.00005 * score^5$. The minimum time interval is limited to $0.01s$ for perceptual reasons. The beeps have a frequency of $440Hz$. We call it *inverse* parking sensor as the sonic experience when typing in a password is inverse to parking the car: When typing, the user starts in the "danger zone," when the password is weak. The soundscape the user experiences first in this case corresponds to the sound he or she experiences last when parking the car.

Noise to Harmonic. "Complex sonic aesthetics and no real-world metaphor." This design implements a paradigm in which the user is rewarded with a pleasant sound for choosing a strong password. A score of 0 is represented by the sound of pure white noise. A score of 10 produces a harmonic, evolving pleasant drone sound in B minor, slightly fluctuating over time in order to provide a long-lasting pleasant sonic experience. The score is mapped to an equal-power crossfader position between the noise and harmonic synthesizers. When the 'good enough' threshold is passed, the key changes to B major.

Encrypted Transmission. "Complex sonic aesthetics and real-world metaphor." This design is based on the sound of a radio transmission which gets chopped up or 'encrypted' with increasing password strength. We used a recording of a Space Shuttle final approach radio transmission from NASA (public domain, available at Wikimedia Commons [30]), in which we deleted the silent parts, so there is always some chatter going on. This signal is routed through distortion and bitcrusher effects, which are ramped up in relation to the password strength. At a score of 0, the radio transmission is clearly audible, at a score of 10 it is highly distorted. Additionally, we use the transmission's power envelope to modulate the amplitude of a noise generator, which results in a noise signal which keeps the rhythm of the original radio transmission, but without any understandable syllables. This noise signal is put into the mix with increasing password strength. The resulting auditory percept is a radio transmission, which gets increasingly distorted and noisy with password strength, but is always identifiable as originating from speech.

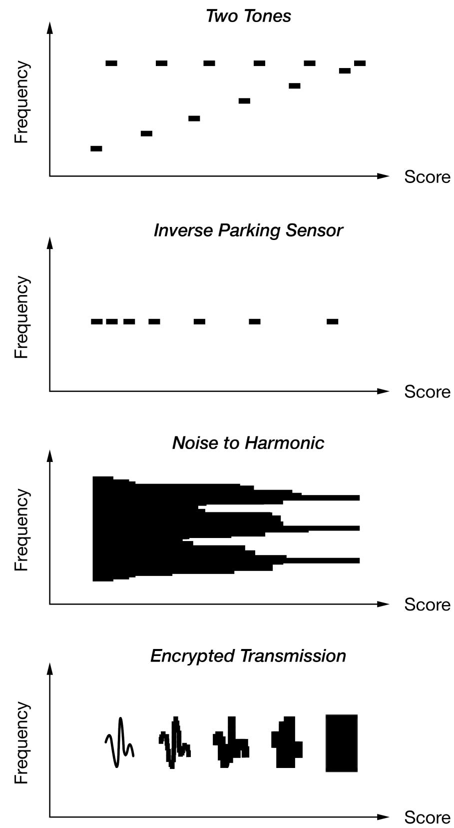

Figure 3: Sound design of the four sonification prototypes

5 RESEARCH AND EVALUATION ROADMAP

Both our considerations above and experience with the implemented sonifications give rise to a multitude of research questions. We finally identified three categories and propose a research and evaluation roadmap to iteratively proceed from general applicability and usability to fundamental questions of sonification design. Due to the COVID-19 lockdown regulations, we were not able to run user studies in our lab. The roadmap below provides the foundation for continuing this research.

5.1 Research Questions

The overarching research question is how to find the best trade-off between sonification design choices and parameters, both with respect to different partly conflicting objectives: Understandability of the mapping, resulting password strength, providing motivation for cognitive reflection and playful experimentation, sounds to be self-explanatory, and suitability for ongoing usage. In addition, it is open what the users' preferences will be regarding the designs.

We clustered our research questions in three categories: (i) questions regarding the auditory representation, (ii) questions concerning the sonified information itself, and (iii) user interaction related questions. In the following, we outline how these questions can lead to empirical studies to elucidate them.

Auditory representation. The appropriate auditory representation of password quality depends on the mental representation and perspective of the recipient. Empirical research has to show the prevalence of the perspectives *object-centered* vs. *user-centered* in larger sample sizes. The same applies to the users' ideas how a secure, weak, hard to crack, etc. password should sound.

Regarding scaling, the question arises if changing the transfer functions from linear to nonlinear is beneficial for emphasizing a range in light of the above objectives.

To inspire cognitive reflection and playful experimentation, we see the need to investigate to what extent the sounds should (i) contain complex aesthetic qualities and (ii) evoke or express emotions, i.e. what amount of arousal hits the sweetspot between being beneficial for motivation and distracting the user? What aesthetic qualities do best support the objective?

Sonified information. Regarding the sonified information, we plan to examine if users understand the change in sound parameters when reaching the 'good enough' score. To an average, not audio-oriented user, our design choices (reverberation or chord change) might not be as apparent as to sound-affine users or even sonification researchers. We expect that other additional information in the sonification design, e.g., the notification if the password is common, are easier to understand for average users, as the sound event (breaking glass) and setting the score to zero encode this additional information in two complementary ways. The optional volume fade-in to mitigate the cold-start problem (avoiding initial punishment) needs to be studied as well, as it adds a layer of complexity to the sound design and might either support good user experience or irritate the users.

User interaction. Concerning interactivity, all designs couple auditory changes tightly to user actions, using either added sound events, or updating parameters of a continuous auditory stream on every keystroke. Yet, both types have opposing advantages (e.g. silence in absence of interaction) and disadvantages (e.g. lack of a persistant reference), therefore the actual user preferences regarding usability and subjective experience are particularly relevant to be studied as a function of this feedback type (discrete vs. continuous).

5.2 Evaluation roadmap

First and foremost, we plan to investigate if users understand the implemented sonification designs with only a short introduction, explanation and training. The research questions regarding the sonified information itself are fundamental to any further investigation and hence need to be addressed first. If this basic understanding is given, we will turn to more detailed questions, for example, regarding polarities of mapping and perception of scaling and transfer functions. Next to these functional aspects, the influence of aesthetic choices regarding performance, preference as well as user experience and long-term usability should be investigated. Designs resembling a real-world metaphor should be compared with alternative sounds in order to gain more insight on the capability of sounds representing semantic meaning. Finally, our fundamental motivation for using sound, the possibility to add an affective component to the feedback to elicit a certain emotional response, needs to be evaluated in respect to the resulting password strength and the degree of motivation for cognitive reflection on password security and playful experimentation with the interface. A comparison between the modalities —textual, visual, auditory, and combined multimodal feedback— will shed light on the question what contribution auditory feedback can make to enhance understanding and the creation of better passwords. These final questions bridge the gap towards real-world application of interactive password strength sonifications.

5.3 Applications

In the application domain, both an easy-to use implementation in the form of a JavaScript library as well as guidelines for developers are beneficial to the dissemination of scientific knowledge regarding password authentication systems. The library can be built on already existing software components such as password strength estimation and visualization modules. Equally important though, is to provide a suitable handout for developers that summarizes the most important findings and offers concrete guidelines for a reasonable implementation of password authentication systems.

For such real-world applications, we must consider that context-aware textual feedback was shown to have a major influence on password security, compared to different design choices in visual representations [38]. Therefore, where practical, this feature should be added to any form of visual, auditory or multimodal feedback as well as it provides in-situ guidance on how to improve a password. We are curious if there will be a measurable interplay between this cognitive guidance and affective triggering through auditory feedback.

6 DISCUSSION AND CONCLUSIONS

In the implementation, some limitations with regard to Tone.js, browser and hardware platform have been encountered. The browser support for WebAudio and Tone.js is still limited on browsers other than Mozilla Firefox and Google Chrome. This manifests itself not only in missing support for certain features among browsers, but in differences in the implementations as well. For example, the audible amount of reverberation when using convolution reverb differs among browsers, which has to be accounted for in the sonification script. The proclaimed cross-device support is limited as well, as Tone.js can take a considerable amount of CPU load on less computationally powerful mobile devices. This can lead to artifacts such as crackling or drop-outs, depending on the complexity of the sonification algorithm used.

Regarding the sonification designs, the *encrypted transmission* has a conceptual disadvantage by utilizing real speech: users might try to understand the verbal content of the radio transmission and therefore be distracted from their task of choosing a strong password. Further research will show if the intuitiveness due to the real-world metaphor of this design outweighs this conceptual disadvantage.

For real-world applications, we must consider some general limitations applying to any auditory feedback. Sound output might be muted on many devices. Users are not necessarily used to this modality as a means of conveying more than simple information, e.g., a notification tone. On the other hand, this novelty effect of more complex sonifications might actually be in favor of our objective, stimulating interest and encouraging playful experimentation with passwords through affective sound feedback.

In this paper we have explored the design space and possible features and attributes of password sonification along with with an implementation of four novel sonification designs and condensed our findings into a research roadmap. We expect a fair chance that adding sound as an additional modality to password strength feedback may indeed help to improve password – and thus overall system – security and raise interest and awareness for the topic.

ACKNOWLEDGMENTS

This work has been funded (in part) by the Federal Ministry of Education and Research of Germany (BMBF) under grant no. 16DII111 ("Deutsches Internet-Institut").

REFERENCES

[1] Tahani Almehmadi and Fahad Alsolami. 2019. Password Security in Organizations: User Attitudes and Behaviors Regarding Password Strength. In *16th International Conference on Information Technology-New Generations (ITNG 2019)*, Shahram Latifi (Ed.). Springer International Publishing, Cham, 9–13.

[2] Louise Axon, Sadie Creese, Michael Goldsmith, and Jason R C Nurse. 2016. Reflecting on the Use of Sonification for Network Monitoring. *SECURWARE 2016: The Tenth International Conference on Emerging Security Information, Systems and Technologies* c (2016), 254–261. https://doi.org/10.1016/j.heares.2013.01.009

[3] M Ballora, NA Giacobe, and DL Hall. 2011. Songs of cyberspace: an update on sonifications of network traffic to support situational awareness. *SPIE Defense, ...* 8064 (2011), 1–6. https://doi.org/10.1117/12.883443

[4] Leon Bošnjak and Boštjan Brumen. 2019. Rejecting the death of passwords: Advice for the future. *Computer Science and Information Systems* 16, 1 (2019), 313–332. https://doi.org/10.2298/CSIS180328016B

[5] Debra Box and Dalenca Pottas. 2013. Improving Information Security Behaviour in the Healthcare Context. *Procedia Technology* 9 (2013), 1093–1103. https://doi.org/10.1016/j.protcy.2013.12.122

[6] Margaret M Bradley and Peter J Lang. 2007. *The International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual.* Technical Report. University of Florida, Gainesville, Fl.

[7] Mohamed Debashi and Paul Vickers. 2018. *Sonification of network traffic flow for monitoring and situational awareness.* Vol. 13. 1–31 pages. https://doi.org/10.1371/journal.pone.0195948

[8] Matteo Dell' Amico, Pietro Michiardi, and Yves Roudier. 2010. Password Strength: An Empirical Analysis. In *2010 Proceedings IEEE INFOCOM*. IEEE, 1–9. https://doi.org/10.1109/INFCOM.2010.5461951

[9] Gaël Dubus and Roberto Bresin. 2013. A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE* 8, 12 (2013). https://doi.org/10.1371/journal.pone.0082491

[10] Paul Ekman. 1992. An argument for basic emotions. *Cognition & Emotion* 6, 3-4 (1992), 169–200. https://doi.org/10.1037/0033-295X.99.3.550

[11] Dinei Florêncio, Cormac Herley, and Paul C. Van Oorschot. 2014. Password portfolios and the finite-effort user: Sustainably managing large numbers of accounts. In *Proceedings of the 23rd USENIX Security Symposium*. 575–590.

[12] BL Fredrickson. 1998. What good are positive emotions? *Review of general psychology* 2, 3 (1998), 300–319.

[13] Michael Friendly. 2008. A Brief History of Data Visualization. In *Handbook of Data Visualization*, Chun-houh Chen, Wolfgang Härdle, and Antony Unwin (Eds.). Springer, Berlin, Heidelberg, 15–56.

[14] Florian Grond and Thomas Hermann. 2012. Aesthetic strategies in sonification. *AI and Society* 27, 2 (2012), 213–222. https://doi.org/10.1007/s00146-011-0341-7

[15] Iwan Gulenko. 2014. Improving passwords: influence of emotions on security behaviour. *Information Management & Computer Security* 22, 2 (2014), 167–178. https://doi.org/10.1108/imcs-09-2013-0068

[16] Marc Hassenzahl and Andreas Beu. 2001. Engineering Joy. *IEEE Software* 18, 1 (2001), 1–8.

[17] Thomas Hermann. 2008. Taxonomy and Definitions for Sonification and Auditory Display. In *Proc. International Conference on Auditory Display*. 1–8.

[18] Thomas Hermann, Tobias Hildebrandt, Patrick Langeslag, and Stefanie Rinderle-Ma. 2015. Optimizing aesthetics and precision in sonification for peripheral process-monitoring. In *Proc. 21st International Conference on Auditory Display*. 317–318.

[19] Thomas Hermann, Jiajun Yang, and Yukie Nagai. 2016. EmoSonics - Interactive sound interfaces for the externalization of emotions. In *Proc. AudioMostly*. 61–68. https://doi.org/10.1145/2986416.2986437

[20] Neil Hillman and Sandra Pauletto. 2014. The Craftsman: The use of sound design to elicit emotions. *The Soundtrack* 7, 1 (2014), 5–23. https://doi.org/10.1386/st.7.1.5_1

[21] Charles Hutchins, Holger Ballweg, Shelly Knotts, Jonas Hummel, and Antonio Roberts. 2014. Soundbeam: A Platform for Sonyfing Web Tracking. *Proceedings of the International Conference on New Interfaces for Musical Expression* (2014), 497–498. http://www.nime.org/proceedings/2014/nime2014{_}447.pdf

[22] David Jaeger, Chris Pelchen, Hendrik Graupner, Feng Cheng, and Christoph Meinel. 2016. Analysis of Publicly Leaked Credentials and the Long Story of Password (Re-)use. *Proceedings of the 11th International Conference on Passwords (PASSWORDS2016)* (2016), 1–19. http://mykayem.org/pdfs/Jaeger2016.pdf

[23] Patrik N Juslin. 2019. *Musical Emotions Explained: Unlocking the Secrets of Musical Affect.* Oxford University Press.

[24] Mariam M. Kassim and A. Sujitha. 2013. ProcurePass: A user authentication protocol to resist password stealing and password reuse attack. *Proceedings - 2013 International Symposium on Computational and Business Intelligence, ISCBI 2013* (2013), 31–34. https://doi.org/10.1109/ISCBI.2013.14

[25] Otto Hans-Martin Lutz, Jacob Leon Kröger, Manuel Schneiderbauer, and Manfred Hauswirth. 2019. Surfing in Sound: Sonification of hidden web tracking. In *Proc. International Conference on Auditory Display*. 306–309. https://doi.org/10.21785/icad2019.071

[26] Otto Hans-Martin Lutz and Manuel Schneiderbauer. 2020. Password Strength Sonification Video. https://www.youtube.com/watch?v=f7WitRVSElM

[27] Yotam Mann. 2015. Interactive Music with Tone.js. *Proceedings of the International Web Audio Conference* (2015).

[28] Mattt. 2020. Sonic Hash. https://github.com/mattt/Sonic-Hash

[29] Daniel Miessner. 2020. Top 100.000 commonly used password list. https://github.com/danielmiessler/SecLists/tree/master/Passwords/Common-Credentials

[30] NASA. 2003. Space Shuttle Final Approach Radio Transmission 2003/02/01. https://commons.wikimedia.org/wiki/File:Space{_}shuttle{_}columbia{_}final{_}approach{_}20030201.ogg

[31] John G. Neuhoff. 2019. Is Sonification Doomed to Fail?. In *Proc. 25th International Conference on Auditory Display*. 327–330. https://doi.org/10.21785/icad2019.069

[32] Lawrence O'Gorman. 2003. Comparing Passwords, Tokens, and Biometrics for User Authentication. *Proc. IEEE* 91, 12 (2003), 2019–2020. https://doi.org/10.1109/JPROC.2003.819605

[33] Sarah Pearman, Shikun Aerin Zhang, Lujo Bauer, Nicolas Christin, and Lorrie Faith Cranor. 2019. Why people (don't) use password managers effectively. *Proceedings of the 15th Symposium on Usable Privacy and Security, SOUPS 2019* (2019), 319–338.

[34] Rosalind W Picard. 1999. Affective Computing for HCI. In *Proceedings of the 8th HCI International Conference on Human-Computer Interaction: Ergonomics and User Interfaces*. 829–833. http://dl.acm.org/citation.cfm?id=647943.742338

[35] James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39 (1980), 345–356.

[36] Tobias Seitz, Manuel Hartmann, Jakob Pfab, and Samuel Souque. 2017. Do differences in password policies prevent password reuse? *Conference on Human Factors in Computing Systems - Proceedings* (2017), 2056–2063. https://doi.org/10.1145/3027063.3053100

[37] Akbar Siami Namin, Rattikorn Hewett, Keith S. Jones, and Rona Pogrund. 2016. Sonifying Internet Security Threats. *Proc. 2016 Conference on Human Factors in Computing Systems Extended Abstracts* (2016), 2306–2313. https://doi.org/10.1145/2851581.2892363

[38] Blase Ur, Felicia Alfieri, Maung Aung, Lujo Bauer, Nicolas Christin, Jessica Colnago, Lorrie Faith Cranor, Henry Dixon, Pardis Emami Naeini, Hana Habib, Noah Johnson, and William Melicher. 2017. Design and evaluation of a data-driven password meter. *Conference on Human Factors in Computing Systems - Proceedings* 2017-May (2017), 3775–3786. https://doi.org/10.1145/3025453.3026050

[39] Paul Vickers, Christopher Laing, Mohamed Debashi, and Tom Fairfax. 2014. Sonification Aesthetics and Listening for Network Situational Awareness. (2014). https://doi.org/10.13140/2.1.4225.6648 arXiv:1409.5282

[40] Bruce Walker, Gregory Kramer, and David M. Lane. 2000. Psychophysical scaling of sonification mappings. *Proceedings of ICAD 2000* (2000), 99–104. http://dev.icad.org/Proceedings/2000/WalkerKramer2000.pdf

[41] Bruce N. Walker and Michael A. Nees. 2011. Theory of Soification. In *The Sonification Handbook*, Thomas Hermann, Andy Hunt, and John Neuhoff (Eds.). 9–39.

[42] Daniel Lowe Wheeler. 2016. ZxCVBN: Low-budget password strength estimation. *Proceedings of the 25th USENIX Security Symposium* (2016), 157–173.