SONIFICATION WITH MUSIC FOR CYBERSECURITY SITUATIONAL AWARENESS

Courtney Falk

Infinite Machines
12948 Cantigny Way
Carmel, IN, USA
courtney.falk@infinite-machines.com

Josiah Dykstra

U.S. Department of Defense
Cybersecurity Operations
Ft. George G. Meade, MD, USA
JDykstra@LTSnet.net

ABSTRACT

Cyber defenders work in stressful, information-rich, and high-stakes environments. While other researchers have considered sonification for security operations centers (SOCs), the mappings of network events to sound parameters have produced aesthetically unpleasing results. This paper proposes a novel sonification process for transforming data about computer network traffic into music. The musical cues relate to notable network events in such a way as to minimize the amount of training time a human listener would need in order to make sense of the cues. We demonstrate our technique on a dataset of 708 million authentication events over nine continuous months from an enterprise network. We illustrate a volume-centric approach in relation to the amplitude of the input data, and also a volumetric approach mapping the input data signal into the number of notes played. The resulting music prioritizes aesthetics over bandwidth to balance performance with adoption.

1. INTRODUCTION

The SOC is the heart of cyber defense for many industry and government organizations. The tactical cyber operators and analysts who work in these environments monitor and respond to threats against their organization's mission sometimes 24 hours a day. Their work is high-value and complex, and the workforce suffers from fatigue, frustration, and high cognitive workload [1]. The SOC aims to maximize the productivity of analysts detecting and mitigating cyber events, while accounting for the human limitations of such work. Cyber defenders also work outside of SOC environments where some or most of their time may even be spent on non-security related tasks.

The economic value of effective cybersecurity can be extraordinarily high. Organizations that quantify their expected losses in terms of data breaches, productivity, or intellectual property report that they routinely lose millions of dollars. Therefore, mitigations and controls are justified by the value they bring in lowering such risk. Data-driven fiscal decisions justify a robust and comprehensive approach to cybersecurity.

Security professionals have access to a plethora of software tools and seemingly endless volume of data that can provide insights about the health and status of computer networks. The data are commonly in text and binary formats, and visualization tools can help the human analysts more easily consume and analyze the data. Unfortunately, security analysts cannot afford the mental demand to visually monitor these displays continually.

Signals of information from cybersecurity data are both discrete and continuous. Depending on the situation, a security analyst may discern the status of security by considering one or more stream of real-time events including intrusion detection alarms, user login events, and changes in network traffic volume. The work is unpredictable and dynamic.

The focus of our research is to gently aid cybersecurity situation awareness. Situational awareness takes many forms, including detection of anomalies occurring in a variety of data sources such as user logins and remote exploitation attempts. In this paper, we consider the application of music with low information density as sonification of cyber activity. We envision our sonification as background music for cybersecurity professionals or groups, including the SOC. Ideally, the music would be a pleasant and subtle experience for those unaware of its information value.

Many sonification implementations, including those for network traffic and other cyber security data, prioritize information over aesthetics. That is, they seek to very clearly convey meaning of the input data, and sometimes to maximize the amount of information conveyed in sound. The result is an information-rich application, but one ill-suited for pleasant and continuous listening.

Music may offer the ability to aid SOC analysts in an appealing and complimentary manner to their existing environment. We support and extend Vickers and Hogg's intuition that aesthetics facilitates ease of listening [2], and propose that it is possible to use music to convey information that sacrifices some information bandwidth for improved aesthetics. Instead of mapping every sound and audio feature to the data directly, this approach relies on broad musical characteristics to subtly inform the user. In time through this line of research, we aim to show that the use of subtle musical cues improves the speed and accuracy of analytical tasks and achieves greater satisfaction from users compared with other alternatives [3].

This paper makes the following contributions:

- We propose a method for mapping discrete and continuous signals of events to music, with attention to application with cybersecurity data.
- We offer a research prototype implementation of this approach and demonstrate it by sonifying 708M authentication events from nine months on a live enterprise network.
- We evaluate the information bandwidth of this technique,

https://doi.org/10.21785/icad2019.014

and offer a research agenda for continued experimentation, development, and operations.

2. BACKGROUND AND RELATED WORK

In this section, we introduce the concepts and related work in music and cybersecurity necessary to understand the subsequent work.

There is a stark division in scholarly research between sonification and music. As a communications medium, music has been tied to emotion. Film scores, for example, provide useful insights because they use music as a passive medium to introduce or reinforce information. However, film scores are most often used to communicate emotion, not concrete data. Music in film is consumed passively, as a secondary stimulus to the visual display and spoken dialogue.

Research shows that music convey meaning in data. In a 2018 survey of sonified weather data, researchers revealed that musical characteristics contribute to meaningful data perception, analysis and interpretation [4]. They also found increased engagement levels with melodies that included pitch, timbre and rhythm. The same study participants reported pronounced differences between perceived usability and aesthetics.

Music offers a canvas for communicating a wide variety of diverse data. In theory, music can be generated from any source of input data. Research by Davis and Mohammad produced a system ("TransPose") that would generate music from text taken from literature [5]. Natural language text is a special case of data because it is by its very nature unstructured. Unstructured data presents its own unique challenges that continue to be addressed by natural language processing to this data. The research in this paper focuses on structured data, specifically logs of network login events.

The SOC typically serves to monitor trends and triage security events, and to determine whether they should be escalated for in-depth analysis. As a result, analysts only require enough data and granularity to make those decisions. Data feeds come from many internal and external sources. Some feeds, such as intrusion detection systems, can continually generate more than 100,000 events each day [6]. Data sources vary from one SOC to another, and are impacted by organization size, sector, and budget. Common data sources of security-relevant information include:

- Network intrusion detection systems
- Host intrusion detection systems
- Network traffic logs (raw and summarized)
- Operating system activity and audit logs
- Server and network device logs such as web server, proxy, and DNS logs
- Security device logs, such as firewalls and anti-virus
- Malware analysis and sandbox analysis
- User and entity behavior

Researchers have suggested that detecting anomalies in network traffic "has potential value as an anomaly-detection approach include long-term, continuous listening to the sonification for real-time detection of deviations" [7]. Current approaches to sonification, such as those pursued by Axon et al., emphasize encoding as much data in the music signal as possible [8]. The benefits from such an approach are two-fold. First, the human analyst who hears the sonification receives all the data as original received. Second, there is not filtering or processing involved, which greatly simplifies the process of transforming raw data into music. Examples of this work can be heard online (`https://soundcloud.com/user-71482294`).

Sonification is a viable technique for both real-time and historical analysis. In 2005, Childs explored auditory display for monitoring real-time data. Like cyber defenders, financial traders monitor and act on data streams from text and visual displays. Using simple and sparse musically-based sonification, they reported that a commercial prototype program was "effective" using a two-note scheme [9].

Music can be an intuitive medium requiring little or no training about how network events are mapped. Research shows that both trained and untrained humans can perceive the common elements of music. These elements are pitch (which governs melody and harmony), rhythm (and its associated concepts tempo, meter, and articulation), dynamics, and the sonic qualities of timbre and texture. Further, most listeners have learned to associate musical cues with emotions. In Vivaldi's *Four Seasons*, the composer uses music cues that attempt to auralize the sights, sounds, and events of the natural seasons. Film soundtracks also use music to prompt or bolster a desired feeling along with the video. Soundtracks are static and pre-selected, synchronized with the video. We are unaware of any attempt to use soundtrack music other than for art and emotion. That is, soundtrack music is not being used to communicate data in music.

In this paper, we focus on sonification through instrumental classical music. We limit ourselves to instrumental output without vocals to avoid extraneous variable, complexity, and mental demand on the listener. We describe our output as "classical" in the tradition of western music with established principles [10]. Classical music is found to be aesthetically pleasing by the general population [11].

An important variable in the ability to use music for sonification is the rate of information transfer. We define *auditory cognitive bandwidth* as the channel capacity of information that a human listener can perceive and process from an audio stream. We measure this bandwidth as bits per second. Intuitively, active listening maximizes the auditory cognitive bandwidth compared to passive, background music. In 2009, Ramakrishnan proposed an application of information theory to sonification design that allowed quantification of information communicated by a sonification [12]. This work focused on maximum rates and did not differentiate between active and passive listening. We hypothesize, but have not yet explored, that channel capacity is reduced with passive listening. However, in an 18-subject study of simple tasks, Hildebrandt et al. revealed that using sonification to monitor a process as a secondary task had no significant effect on performance in either task [13].

3. METHOD

In this section, we introduce methods for encoding discrete and continuous signal from cybersecurity events into musical forms. Multiple different data signals, both continuous and discrete, can be encoded in such a way that they all fit in a coherent musical framework. Parseihian and Katz strive for a similar, generative and modular approach in their research to produce sonification cues that represent a physical environment [14]. A hypothetical example scenario is data center monitoring. The processing loads

of machine could provide a continuous input data signal. Events of when machines crash could provide a discrete input data signal. A system administrator could listen to the output music and quickly gain a general sense about the health of equipment in the server room.

3.1. Discrete Signals

Some cybersecurity data occurs as a discrete signal. Discrete data signals occur infrequently relative to other inputs. For example, network login events in a sufficiently large network occur frequently and regularly enough that they appear as a continuous signal. Login failures, however, are significantly less frequent, and the gaps between their signals make each event appear separate and discrete. These discrete signals should stand out from the other, competing data signals in the music and demand attention. Other examples of cybersecurity data that imitate a discrete data signal include anti-virus alerts and abrupt machine shutdowns, which may signal faulty hardware or malicious software.

We offer two approaches to encoding discrete signals into music: the intrusive noise, and harmonic tension. These two approaches are not mutually exclusive and could be combined or used for different signal simultaneously.

3.1.1. Intrusive noise

One approach to encoding a discrete event is an intrusive sound to represent the event. The sound could be either a musical or non-musical element. This noise may noticeably stand out from the ongoing music, making itself apparent. Take for instance the "BRAAAM" effect (aka the "BWONG") popularized in the movie, *Inception* [15]. The more distinctive and invasive the sound, the more likely that a listener will notice. However, overuse of the noise will cause fatigue and annoyance.

3.1.2. Harmonic tension

A second approach would be to take the music already being played at a given timestep and introduce harmonic tension. The idea is to create something that is still aesthetically pleasing while being noticeable. For instance, adding a minor seventh to a major chord produces a dominant seventh chord. In Western music, dominant sevenths create a sense of musical tension from the dominant key. The upside to this approach is that it is aesthetically pleasing and less fatigueinducing than a "BRAAAM." However, the downside is that it would not be as noticeable, and may require some amount of training on the part of the analyst in order to properly recognize its signal.

3.2. Continuous Signals

Some cybersecurity data occurs as a continuous signal. A continuous data signal is one that is present for most or all timesteps. Examples of cybersecurity data in this form include network traffic and web server logs. One way of encoding these signals is to create a continuous music stream and alter one parameter in relation to the data signal. Arpeggiating over chords is one technique to create such a musical stream.

We offer two approaches to encoding continuous signals into music: increasing and decreasing volume, and adding and removing voices. These two approaches could also be combined or used for different signal simultaneously.

3.2.1. Increasing and decreasing volume

As the data signal increases, increase the volume of the musical stream. Start at a pianissimo for the lowest level, going up to fortissimo for the highest level. This could be fatiguing for a signal that stays at high levels for long periods of time, causing the music to be loud over long periods of time.

3.2.2. Adding and removing voices

A second approach to creating a musical stream is to select multiple independent musical lines, or voices. Then quantize the data signal into equal levels. When the data signal is within the lowest level, only play that corresponding voice. As the data signal increases and crosses boundaries, add the other voices. Vice versa, the data signal decreasing removes voices. Adding and removing voices could be done in conjunction with the volume approach where the volume for each voice is set to correspond to where the data signal is in that particular quantized layer.

4. CASE STUDY

To illustrate the application of our approach, we consider a SOC task of monitoring authentication events for anomalies. We use the public network authentication dataset from the Los Alamos National Laboratory (LANL) [16]. This anonymized data set encompasses nine continuous months and represents 708,304,516 successful authentication events from real users to computers collected from the LANL enterprise network. The data are representative of other similar networks, and offer a non-critical indicator of activity on the network. The stream of login events produces a continuous data signal that varies according to the day of week and time of day, as seen clearly in Figure 1 where Monday is the first day of the week.

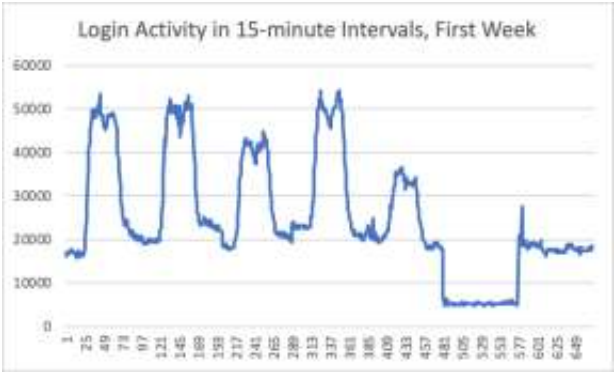

Figure 1: Graph of LANL login events ordered chronologically for the first week of time.

One drawback to the dataset is that it contains only successful login events. An analyst monitoring a computer network may want to monitor login traffic and investigate anything unsuccessful. To produce a more real-world scenario, we utilize a Poisson distribution to artificially insert unsuccessful login events.

The LANL data was quantized into 15-minute increments. This aggregation compresses the time required to analyze the data, and smooths the signal. Since each 15-minute increment is then encoded into a quarter note, the music is in common (4/4) time,

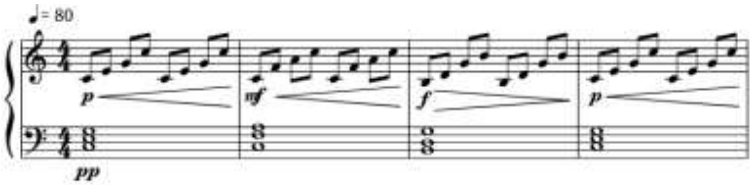

Figure 2: Sonification of a continuous data signal where differences in the input signal are encoded into the volume of the notes.

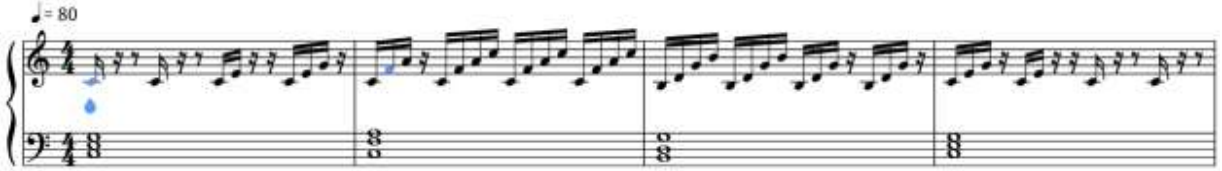

Figure 3: Sonification of a continuous data signal where the differences in the input signal are encoded into the number of sixteenth notes present in an arpeggio starting from the bottom.

each measure equates to one hour of time in the input data. Data about the user-computer pairing for each login event were abandoned, keeping only the information about the time when the login occurred. The goal was to sonify the number of logins in each 15-minute block into a musical structure that describes its state. Two techniques were applied: volume modulation, and changing the number of notes in a corresponding beat.

One of the driving goals of this research is to provide concrete artifacts that demonstrate the described techniques in action.[1] All source code is freely available under the GPLv3 license. The source code is written in Python 3 with all module dependencies being readily available via the Python Package Index (PyPI). MIDI is the generated as the output, which is playable via a wide variety of software audio players and synthesizers [17].

5. RESULTS AND DISCUSSION

This project experimented with two different novel sonification approaches. Both approaches utilized a common framework. An aesthetically-pleasing musical sequence was chosen to constrain the structure of the notes. There are numerous common chord progressions in Western music. This first system used the I-IV-V-I chord progression of a major key, which is extraordinarily common. Both approaches play the chord sequence as whole notes two octaves lower than the music from the data signal. This produces the carrier signal that tells a listener that the system is indeed working as expected even if there is a lack of data signal. Additionally, both approaches equate each 15-minute block of input data as one quarter note beat in the output music. The system outlined here has a few basic hyperparameters that could be changed to produce different results:

- Harmonic components:
  - Chords in the progression.
  - Musical key.
- Tempo components:
  - Length of each input time block.
  - Beats per minute.
  - Time signature.
- Timbre components:
  - Number of musical voices
  - Instruments/samples associated with particular voices.
- Genre/style [18].
- Length of quantization for the input data.

[1] https://github.com/CalmLogarithm/sonification

5.1. Loudness-centric Approach

The first sonification approach is to vary the volume of the output musical notes in direct relation to the amplitude of the input data. As the values of the input data increase, so too do the volume of the music it produces. A first pass through the data set provides the maximum signal level. Each step of the input data is then converted to a value between 20 and 127 that corresponds to its value relative to the maximum value. The values 20 and 127 were chosen because 127 is the loudest volume level for a MIDI signal and 20 is still barely audible. Figure 2 shows what the musical notation for such a volume-based encoding would look like.

The benefits of the volume-centric approach is that it guarantees a continuous, coherent musical stream. One drawback is that it may prove difficult for the human ear to distinguish subtle difference in the volume. So for a continuous data signal where minor changes are significant, this may not be the best approach to use.

5.2. Volumetric Approach

The second sonification approach encodes the input data signal into the number of sixteenth notes played in a given quarter node interval. Recall from earlier that each quarter node interval corresponds to a fiften minute block of time in the LANL login event data. Each sixteenth note is a step in an arpeggio much like the volume-based approach. But for the number-based approach, each fifteen minute interval of input data is quantized by four. The lowest 25% of the input signal would only cause the lowest note of the

arpeggio to be rendered while the highest 25% of the input signal would cause all four notes of the arpeggio to be rendered. So as the input data signal increases, the number of notes played increases, and due to the structure of the arpeggio, also goes higher (Figure 3).

There are two drawbacks to using the number-based approach. First, a signal that is mostly in the bottom quarter of the input range will create music that is mostly a sequence of sixteenth notes that occur on the beat. This creates a somewhat percussive effect. The second drawback is the coarse granularity because there are only four notes used in the arpeggio then the data can only be divided into four parts. There are workarounds that could improve this. One workaround is to stack multiple different voices on top of one another so that once the lowest voice has rendered all of its arpeggio then the next voice up begins rendering its arpeggio. A second workaround is to analyze the distribution of the input data signal and divide up the input value range in unequal blocks to make it more likely that more than one of each note in the arpeggio is playing, creating more of a sense of diversity in the output music.

5.3. Evaluation

We sought to maintain a low-bandwidth information channel. This implementation focused on a single input variable: network logins. At the peak, there were 125,008 login events encoded in a single 15-minute increment. That value requires a 17-bit value to encode. The music is 80 beats per minute, or 1.33 beats per second. As a result, this music communicates 22.66 bits per second. We have not yet conducted user testing for these prototypes. However, our initial assessment from listening to the sonifications is that they are both aesthetic and effective in communicating the authentication events. We invite the reader to listen to audio clips of this dataset.[2]

6. CONCLUSIONS AND FUTURE WORK

In this work, we proposed a novel sonification process for mapping data related to cyber defense into music. The resulting music prioritizes aesthetics to balance performance with adoption. The musical cues relate to notable events in such a way as to minimize the amount of training time a human listener would need in order to make sense of the cues. We demonstrated our technique on a dataset of 708M authentication events over nine continuous months on an enterprise network.

One limitation is that the software implementations linked to this paper only operates on a static data set. A fully-featured tool that is ready for enterprise use must also operate over a continuous data stream. Changes are required to the source code for it to function correctly with data of a dynamic and previously unknown range. However, we believe that the approach is applicable to any input data that can be quantized.

Future work should consider music generation via neutral nets [19]. Neural networks can be trained on a variety of musical styles and genres, creating dynamic yet coherent compositions to accompany the sonification techniques described in this paper. Due to how network data is of variable length, specific neural network architectures such as long short-term memory (LSTM) are the most applicable to this problem [20].

We intend to experiment with our approach in real SOC settings and to validate the utility and satisfaction of the music on their work. Axon et. Al have both developed a sonification approach and tested its usefulness in a realistic scenario [21]. Realistic tests should utilize network analysts as the pool of test subjects. If possible, network analysts both with and without backgrounds in music education should be used in order to test how easily the sonification techniques described in this paper are interpreted by people of varying degrees of musical skill.

We look forward to continued research, development, and operational use of this technique for improved security.

[2] https://soundcloud.com/user-679831789/sets/sonification

7. REFERENCES

[1] C. L. Paul and J. Dykstra, "Understanding Operator Fatigue, Frustration, and Cognitive Workload in Tactical Cybersecurity Operations," *J. Info. Warfare*, vol. 16, no. 2, pp. 1–11, 2017.

[2] P. Vickers and B. Hogg, "Sonification abstraite/sonification concrete: An 'aesthetic perspective space' for classifying auditory displays in the ars musica domain," in *Proc. of the 12th Int. Conf. on Auditory Display*, London, UK, 2006, pp. 210–216.

[3] S. Barrass and P. Vickers, "Sonification Design and Aesthetics," in T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*, pp. 145-171, Berlin: Logos Publishing House, 2011. Retrieved from https://sonification.de/handbook/download/TheSonificationHandbook-chapter7.pdf

[4] J. Middleton, J. Hakulinen, K. Tiitinen, J. Hella, T. Keskinen, P. Huuskonen, J. Linna, M. Turunen, M. Ziat, and R. Raisamo, "Sonification with Musical Characteristics: A Path Guided by User Engagement," in *24th Int. Conf. on Auditory Displays*, Houghton, MI, 2018.

[5] H. Davis and S. M. Mohammad, "Generating Music from Literature," 2014. Retrieved from https://arxiv.org/pdf/1403.2124.pdf.

[6] C. Zimmerman, "Ten Strategies of a World-Class Cybersecurity Operations Center," 2014. Retrieved from https://www.mitre.org/sites/default/files/publications/pr-13-1028-mitre-10-strategies-cyber-ops-center.pdf.

[7] B. N. Walker and M. A. Nees, "Theory of Sonification," in T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*, pp. 9–39, Berlin: Logos Publishing House, 2011. Retrieved from https://sonification.de/handbook/download/TheSonificationHandbook-chapter2.pdf

[8] L. Axon, J. R. Nurse, M. Goldsmith, and S. Creese, "A formalised approach to designing sonification systems for network-security monitoring," in *Int. J. on Advances in Security*, 2017, 10(1–2).

[9] E. Childs, "Auditory Graphs of Real-Time Data," in *11th Int. Conf. on Auditory Displays (ICAD)*, Limerick, Ireland. July 2005.

[10] "classical, n.2." OED Online, Oxford University Press, March 2019. Retrieved from www.oed.com/viewdictionaryentry/Entry/33881. Accessed 9 March 2019.

[11] D. K. Simonton, "Aesthetic success in classical music: A computer analysis of 1935 compositions," in *Empirical Studies of the Arts*, vol. 4, pp. 1–17, 1986.

[12] C. Ramakrishnan, "Sonification and information theory," in *Proc. of the 6th Int. Conf. on Auditory Display (ICAD)*, pp. 121–142, 2009.

[13] T. Hildebrandt, T. Hermann, and S. Rinderle-Ma, "Continuous sonification enhances adequacy of interactions in peripheral process monitoring," *Int. J. of Human-Computer Studies*, vol. 95, pp. 54–65, 2016.

[14] G. Parseihian and B. F. G. Katz, "Morphocons: A new sonification concept based on morphological earcons," *J. of the Audio Engineering Society*, vol 60, no. 6, pp. 409– 18, July 2012.

[15] K. Jagernauth, "Who Really Created The 'Inception' BRAAAM? Composer Mike Zarin Sets The Record Straight," 13 November 2013. Retrieved from IndieWire: https://www.indiewire.com/2013/11/whoreally-created-the-inception-braaam-composer-mikezarin-sets-the-record-straight-91690/

[16] A. D. Kent, *User-Computer Authentication Associations in Time*, Los Alamos Laboratory, 2014. doi:10.11578/1160076

[17] The MIDI Association, "The Complete MIDI 1.0 Detailed Specification," 1996. Retrieved from https://www.midi.org/specifications/item/the-midi-1-0-specification

[18] C. J. Carr and Z. Zukowski, "Generating Albums with SampleRNN to Imitate Metal, Rock, and Punk Bands," 2018. Retrieved from https://arxiv.org/abs/1811.06633

[19] N. Kalchbrenner, E. Elsen, K. Simonyan, S. Noury, N. Casagrande, E. Lockhart, F. Stimberg, A. van den Oord, S. Dieleman, and K. Kavukcuoglu, "Efficient Neural Audio Synthesis," 2018. Retrieved from https://arxiv.org/abs/1802.08435

[20] M. Kaliakatsos-Papakostas and A. Gkiokas, "Interactive Control of Explicit Musical Features in Generative LSTM-based Systems," in *Proc. of the Audio Mostly 2018 on Sound in Immersion and Emotion*, ACM, September 2018.

[21] L. Axon, B. Alahmadi, J. Nurse, M. Goldsmith, and S. Creese, "Sonification in Security Operations Centres: What do Security Practitioners Think?" in *Proc. of the Workshop on Usable Security (USEC)*, February 2018.