ARE USERS BETTER ABLE TO CORRECTLY INTERPRET SINGLE OR CONCATENATED AUDITORY ICONS THAT CONVEY A COMPLEX MESSAGE?

Jinwoo Choi[1], Natalie Lodinger[1], Keith S. Jones[1],
Akbar Siami Namin[2], Miriam Armstrong[1], and David Sears[3]
[1]Department of Psychological Sciences, [2]Department of Computer Science, [3]School of Music
Texas Tech University, Lubbock, TX

Auditory icons are naturally occurring sounds that systems play to convey information. Systems must convey complex messages. To do so, systems can play: 1) a single sound that represents the entire message, or 2) a single sound that represents the first part of the message, followed by another sound that represents the next part of that message, etc. The latter are known as concatenated auditory icons. To evaluate those approaches, participants interpreted single and concatenated auditory icons designed to convey their message well and poorly. Single auditory icons designed to convey their message well were correctly interpreted more often than those designed to convey their message poorly; that was not true for concatenated auditory icons. Concatenated auditory icons should not be comprised of a series of sounds that each represents its piece of a message well. The whole of a concatenated auditory icon is not the sum of its parts.

Sonifications are non-speech sounds that convey information to users (Walker & Nees, 2011; Kramer et al., 2010). One type of sonification is the *auditory icon*, which consists of a naturally occurring sound that is used to convey information about computer interface events (Gaver, 1989).

Auditory Icons

Gaver published a series of seminal papers concerning auditory icons and their potential role in computerized systems (1986; 1989; 1993; 1995). In those papers, Gaver describes various examples of auditory icons. For example, Gaver (1989) details the SonicFinder, an auditory interface that complemented the Macintosh's graphical interface. In the SonicFinder, when an object was dropped into the trash, the sound of shattering dishes was played; when the trash was emptied, a crunching sound was played.

Gaver argued that auditory icons can effectively convey information to users because users have formed associations between certain sounds and certain events based on their day-to-day lives. For example, users have broken dishes and then placed them in the trash. Accordingly, the sound of dishes breaking is associated with putting things in the trash.

Gaver's research inspired a considerable amount of research on auditory icons. Cabral and Remijn (2019) provide a relatively recent overview of that literature with an emphasis on understanding the use and design of auditory icons. Brazil and Fernström (2011) discuss how auditory icons convey information, various applications of auditory icons, and how to design them.

Using Auditory Icons to Convey Complex Messages

Systems may sometimes need to convey complex messages. An example of a complex message is "the attacker disrupted your computer." One way to convey this message is by using a single auditory icon that represents the entire message, such as the sound of a computer malfunctioning. A second way to convey that message is to play one auditory icon, e.g., the sound of an automobile running poorly, followed by another auditory icon, e.g., the sound of a computer. The combination of two different sonifications is sometimes referred to as concatenation, which has been implemented with auditory icons (Kostiainen, 2011) as well as other sonifications (Blattner et al., 1989).

To date, research has not yet investigated whether one of these approaches is more effective than the other. It is possible that single and concatenated auditory icons may differ in their interpretability. For example, it may be more difficult to identify single sounds that can effectively convey complex messages than to identify single sounds that effectively convey the constituents of a complex message that can be concatenated to form the entire message. If so, then single auditory icons that convey complex messages may be less interpretable than concatenated auditory icons that convey those same messages.

Present Study

The goal of this study was to determine whether participants were better able to correctly interpret single or concatenated auditory icons, both of which conveyed the same complex message. Participants listened to single and concatenated auditory icons that were designed to convey the message "the attacker disrupted your computer." After listening to each auditory icon, participants selected the message they thought the auditory icon was meant to convey from a list of possibilities.

Due to the exploratory nature of this study, no formal hypothesis is offered regarding the relative interpretability of single and concatenated auditory icons. The present study is important because it is the first systematic comparison of these types of auditory icons.

METHOD

Participants

Six hundred and nine participants completed the experiment and received partial course credit. Twenty-four participants were removed from the sample due to missing data. One hundred and fifteen participants were removed from the sample because they self-reported that they did not devote their full attention to the study. Forty-two participants were removed from the sample because they completed the experiment in an extraordinarily long (Cut-off value = 663 s) amount of time. Cut-off values were determined via the "elbow method", i.e., inspecting a plot of study completion times to identify points at which completion times changed markedly (DeSimone et al., 2014; Meade & Craig, 2012; Curran, 2016).

The resultant sample included 428 participants (153 males, 275 females). Their average age was 20.09 years (*SD* = 2.55), ages ranged 18-37 years.

Creating Auditory Icons

Auditory icons for the present study were created based on ratings collected during a large-scale pilot study. Pilot participants listened to sounds and rated their perceived relatedness to several cybersecurity-related terms, "email", "computer", "Internet", "disrupt", "gain", "alter", "access", "function", and "content", on a 5-point scale. Those terms were chosen because the pilot study and the present study are part of a line of research in which we are exploring the use of sonification to inform users about cybersecurity threats.

A mean and 95% confidence interval was computed for the ratings related to each term. Confidence interval widths varied widely, which raised concerns about selecting sounds based on means. Instead, sounds were selected based on confidence interval lower limits, i.e., the low-end of the estimated range into which the rating should fall. That way, we could be confident that participants thought the sound represented a given term at least that well.

Twenty auditory icons from the pilot study were selected for use in the current experiment. Ten of the selected icons were target icons designed to convey the message "disrupt computer." The remaining icons were filler icons designed to convey cybersecurity related messages such as "alter email."

The target set included two versions of each of five auditory icon types: 1) Single-Best, 2) Concatenated-Best, 3) Concatenated-Match Single-Best, 4) Single-Weak, and 5) Concatenated-Weak. Single-Best auditory icons were single sounds whose confidence interval lower limits for "disrupt" and "computer" were as high as possible (3.17 – 3.26 for "disrupt"; 3.04 – 3.11 for "computer"). Concatenated-Best auditory icons were comprised of a sound whose confidence interval lower limit for "disrupt" was as high as possible (3.52 – 3.60), 1 second of silence, and then a sound whose confidence interval lower limit for "computer" was as high as possible (4.04 – 4.38). Please note that Concatenated-Best auditory icons were rated as representing the message better than Single-Best auditory icons. Concatenated-Match Single-Best auditory icons were comprised of a sound whose confidence interval lower limit for "disrupt" was as similar as possible to that for one of the Single-Best auditory icons (3.06 – 3.27), 1 second of silence, and then a sound whose confidence interval lower limit for "computer" was as similar as possible to that for one of the Single-Best auditory icons (3.11 – 3.12). Single-Weak auditory icons were single sounds whose confidence interval lower limits for "disrupt" and "computer" were as low as possible (1.23 – 1.38 for "disrupt"; 1.25 – 1.27 for "computer"). Concatenated-Weak auditory icons were comprised of a sound whose confidence interval lower limit for "disrupt" was as low as possible (1.50 – 1.63), 1 second of silence, and then a sound whose confidence interval lower limit for "computer" was as low as possible (1.35 – 1.42).

The filler set included four Single and six Concatenated auditory icons. Ratings for terms other than "disrupt" and "computer" were quite low and variable, so individual filler items were not matched to counterparts in the target set.

Procedure

The participant completed the experiment on Qualtrics. They first read and agreed to the informed consent statement and received instructions. Participants then listened to the target and filler auditory icons, for a total of 20 sounds. For each, they selected the message that they thought the sound was meant to convey from a list of seven messages: the attacker a) disrupted your computer, b) accessed your computer, c) altered your email, d) affected the function of your Internet, e) disrupted your Internet, and f) accessed your email. The order of the auditory icons was randomized for each participant. Participants then answered questions about their age and sex, as well as about whether they devoted their full attention to the experiment. Finally, participants read a debriefing statement.

RESULTS

Our goal was to determine whether participants were better able to correctly interpret single or concatenated auditory icons. To answer that question, we performed four sets of comparisons: 1) Single-Best vs. Single-Weak auditory icons, 2) Concatenated-Best vs. Concatenated-Weak auditory icons, 3) Single-Best vs. Concatenated-Match Single-Best auditory icons, and 4) Single-Best vs. Concatenated-Best auditory icons.

The first two sets of comparisons evaluated whether the ratings on which the auditory icons were selected reflected participants' abilities to correctly interpret the auditory icons. If so, then participants should correctly interpret Best auditory icons more often than Weak auditory icons. The third set of comparisons evaluated whether participants were better able to correctly interpret single or concatenated auditory icons whose ratings were matched. As noted earlier, the highest ratings for single auditory icons were lower than the highest ratings for concatenated auditory icons. Thus, it was necessary to compare single and concatenated auditory icons whose ratings were as similar as possible so that the difference in ratings could not confound the comparison. The fourth set of comparisons evaluated whether participants were better able to correctly interpret single or concatenated auditory icons whose ratings were as high as possible.

Each set of comparisons was comprised of four tests, one for each possible combination of the relevant auditory icons (e.g., Single-Best_1 vs. Single-Weak_1, Single-Best_1 vs. Single-Weak_2, etc.). Each set of comparisons was designated a "family," and family-wise error was maintained at .05 via a Bonferroni correction (alpha per test = .05/4 tests = .01). Each test was a McNemar's Test, which can be thought of as a repeated-measures Chi-Squared Test (Adedokun & Burgess, 2012). Specifically, a McNemar's Test evaluates whether the number of participants who correctly rated Stimulus A and incorrectly rated Stimulus B is equal to the number of participants who incorrectly rated Stimulus A and correctly rated Stimulus B (Adedokun & Burgess, 2012). If so, participants rated Stimulus A and B equally well. If not, a significant test result would indicate that participants rated one stimulus more accurately than the other. The tests were performed on IBM SPSS Statistics software.

Did ratings reflect participants' abilities to correctly interpret Single auditory icons? Figure 1 presents the four 2x2 contingency tables for our first set of comparisons, Single-Best vs. Single-Weak auditory icons. Inspection of Figure 1 suggests Single-Best auditory icons were interpreted correctly more often than Single-Weak auditory icons. For example, the red table indicates that 158 participants correctly identified Single-Best_1 and incorrectly identified Single-Weak_1. However, only 37 participants incorrectly identified Single-Best_1 and correctly identified Single-Weak_1.

Figure 1
Contingency Tables for Single-Best (SingleBest) and Single-Weak (SingleWeak) Auditory Icons

		SingleWeak_1		SingleWeak_2	
		Correct	Incorrect	Correct	Incorrect
SingleBest_1	Correct	23	158	15	166
	Incorrect	37	210	19	228
SingleBest_2	Correct	17	162	15	164
	Incorrect	43	206	19	230

Formal analyses corroborated that impression. Specifically, each of the four McNemar's Tests in this family was statistically significant. See Table 1 for details.

Table 1
McNemar Chi-Square Statistics for Single-Best and Single-Weak Auditory Icons

		SingleWeak_1	SingleWeak_2
SingleBest_1	Chi-Square	75.08	116.81
	P-Value	<.001	<.001
SingleBest_2	Chi-Square	69.08	114.89
	P-Value	<.001	<.001

These outcomes suggest Single-Best auditory icons were interpreted correctly more often than Single-Weak auditory icons. Accordingly, it appears that the ratings on which these auditory icons were selected reflected participants' abilities to correctly interpret single auditory icons.

Did ratings reflect participants' abilities to correctly interpret Concatenated auditory icons? Figure 2 presents the four 2x2 contingency tables for our second set of comparisons, Concatenated-Best vs. Concatenated-Weak auditory icons. Inspection of Figure 2 suggests Concatenated-Best auditory icons were not interpreted correctly more often than Concatenated-Weak auditory icons. For example, the red table indicates that 75 participants correctly identified Con-Best_1 and incorrectly identified Con-Weak_1. Similarly, 79 participants incorrectly identified Con-Best_1 and correctly identified Con-Weak_1.

Figure 2
Contingency Tables for Concatenated-Best (ConBest) and Concatenated-Weak (ConWeak) Auditory Icons

		ConWeak_1		ConWeak_2	
		Correct	Incorrect	Correct	Incorrect
ConBest_1	Correct	47	75	29	93
	Incorrect	79	227	72	234
ConBest_2	Correct	24	56	21	59
	Incorrect	102	246	80	268

Formal analyses corroborated that impression. Specifically, three of the four McNemar's Tests in this family were not statistically significant. The fourth test was statistically significant, but revealed that Concatenated-Best auditory icons were interpreted correctly *less* often than Concatenated-Weak auditory icons. See Table 2 for details.

Table 2
McNemar Chi-Square Statistics for Concatenated-Best and Concatenated-Weak Auditory Icons

		ConWeak_1	ConWeak_2
ConBest_1	Chi-Square	0.10	2.67
	P-Value	.747	.102
ConBest_2	Chi-Square	13.39	3.17
	P-Value	<.001	.075

These outcomes suggest Concatenated-Best auditory icons were not interpreted correctly more often than Concatenated-Weak auditory icons. Accordingly, it appears that the ratings on which each of the individual auditory icons was selected do not reflect participants' abilities to correctly interpret the message the concatenated auditory icons convey.

Did participants interpret Single or Concatenated auditory icons better when their ratings matched? Figure 3 presents the four 2x2 contingency tables for our third set of comparisons, Single-Best vs. Concatenated-Match Single-Best auditory icons. Inspection of Figure 3 suggests Single-Best auditory icons were interpreted correctly more often than Concatenated-Match Single-Best auditory icons. For example, the red table indicates that 132 participants correctly identified Single-Best_1 and incorrectly identified Con-Match_1. However, only 55 participants incorrectly identified Single-Best_1 and correctly identified Con-Match_1.

Figure 3
Contingency Tables for Single-Best (SingleBest) and Concatenated-Match Single-Best (ConMatch) Auditory Icons

		ConMatch_1		ConMatch_2	
		Correct	Incorrect	Correct	Incorrect
SingleBest_1	Correct	49	132	54	127
	Incorrect	55	192	64	183
SingleBest_2	Correct	41	138	49	130
	Incorrect	63	186	69	180

Formal analyses corroborated that impression. Specifically, each of the four McNemar's Tests in this family was statistically significant. See Table 3 for details.

Table 3
McNemar Chi-Square Statistics for Single-Best and Concatenated-Match Single-Best Auditory Icons

		ConMatch_1	ConMatch_2
SingleBest_1	Chi-square	31.71	20.78
	P-Value	<.001	<.001
SingleBest_2	Chi-square	27.99	18.70
	P-Value	<.001	<.001

These outcomes suggest Single-Best auditory icons were interpreted correctly more often than Concatenated auditory icons when the latter were constructed from individual auditory icons whose ratings were as similar as possible to those for the single auditory icons. Accordingly, it appears that Single auditory icons convey their message more clearly than Concatenated auditory icons.

Did participants interpret Single or Concatenated auditory icons better when their ratings were as high as possible? Figure 4 presents the four 2x2 contingency tables for our fourth set of comparisons, Single-Best vs. Concatenated-Best auditory icons. Inspection of Figure 4 suggests Single-Best auditory icons were interpreted correctly more often than Concatenated-Best auditory icons. For example, the red table indicates that 129 participants correctly identified Single-Best_1 and incorrectly identified Con-Best_1. However, only 70 participants incorrectly identified Single-Best_1 and correctly identified Con-Best_1.

Figure 4
Contingency Tables for Single-Best (SingleBest) and Concatenated-Best (ConBest) Auditory Icons

		ConBest_1		ConBest_2	
		Correct	Incorrect	Correct	Incorrect
SingleBest_1	Correct	52	129	35	146
	Incorrect	70	177	45	202
SingleBest_2	Correct	50	129	34	145
	Incorrect	72	177	46	203

Formal analyses corroborated that impression. Specifically, each of the four McNemar's Tests in this family was statistically significant. See Table 4 for details.

Table 4
McNemar Chi-Square Statistics for Single-Best and Concatenated-Best Auditory Icons

		ConBest_1	ConBest_2
SingleBest_1	Chi-square	17.49	53.41
	P-Value	<.001	<.001
SingleBest_2	Chi-square	16.16	51.31
	P-Value	<.001	<.001

These outcomes suggest Single-Best auditory icons were interpreted correctly more often than Concatenated-Best auditory icons. Recall that ratings for each of the individual auditory icons that were concatenated were higher than the ratings for each of the single auditory icons. Specifically, Con-Best_1 had a confidence interval lower limit of 3.60 for its rating of the term "Disrupt", and 4.04 for that of the term "Computer". Con-Best_2 had a confidence interval lower limit of 3.52 for its rating of the term "Disrupt", and 4.38 for that of the term "Computer". On the other hand, Single-Best_1 had a confidence interval lower limit of 3.17 for its rating of the term "Disrupt", and 3.11 for that of the term "Computer". Single-Best_2 had a confidence interval lower limit of 3.26 for its rating of the term "Disrupt", and 3.04 for that of the term "Computer". Accordingly, it appears that Single auditory icons convey their message more clearly than Concatenated auditory icons, despite that ratings for the individual auditory icons that were concatenated were higher than the ratings for each of the single auditory icons.

DISCUSSION

In the present study, Concatenated Best auditory icons were comprised of a sound that represented half of the to-be-conveyed message well, 1 second of silence, and another sound that represented the other half of the to-be-conveyed message well. The present results suggest Concatenated-Best auditory icons did not effectively convey their intended message. Specifically, Concatenated-Best auditory icons were not interpreted correctly more often than Concatenated-Weak auditory icons; the latter were designed to convey their intended message poorly. Further, Concatenated-Best auditory icons were not interpreted correctly more often than Single-Best auditory icons, despite ratings for each of the individual auditory icons that were concatenated being higher than the ratings for each of the single auditory icons.

Why Didn't Our Concatenated Auditory Icons Convey Their Intended Message Well?

The present results make clear that our concatenated auditory icons did not convey their message well. However, they unfortunately do not speak to why they did not do so.

There are likely a number of possibilities. For example, it is possible that hearing the second auditory icon in the sequence disrupted participants' efforts to interpret the first auditory icon in the sequence. If so, then the concatenated auditory icon would convey only part of its message. Future research should evaluate such possibilities and whether concatenation can be accomplished in such a way so as to avoid such disruptions.

Implications for Concatenated Auditory Icon Design

The present results suggest one cannot construct effective concatenated auditory icons by presenting a series of individual sounds that each represents its piece of the to-be-conveyed message well. Rather, when designing concatenated auditory icons, it appears the whole is not the sum of its parts.

Consequently, when designing concatenated auditory icons, we recommend that one should rely on ratings of how well the collection of sounds represents the to-be-conveyed message, rather than on ratings of how well each individual sound represents its piece of that message. Future research should evaluate how well concatenated auditory icons designed in that way convey their intended messages.

ACKNOWLEDGEMENT

This research was supported by the U.S. National Science Foundation (Award # 1564293). Opinions, findings, and conclusions are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

Adedokun, O. A., & Burgess, W. D. (2012). Analysis of Paired Dichotomous Data: A Gentle Introduction to the McNemar Test in SPSS. *Journal of MultiDisciplinary Evaluation, 8*(17), 125-131.

Blattner, M., Sumikawa, D., & Greenberg, R. (1989). Earcons and Icons: Their Structure and Common Design Principles. *Human-Computer Interaction*, *4*(1), 11–44. https://doi.org/10.1207/s15327051hci0401_1

Cabral, J. P., & Remijn, G. B. (2019). Auditory Icons: Design and Physical Characteristics. *Applied Ergonomics, 78*, 224-239. doi:10.1016/j.apergo.2019.02.008

Curran, P. G. (2016). Methods for the Detection of Carelessly Invalid Responses in Survey Data. *Journal of Experimental Social Psychology, 66*, 4-19. doi:10.1016/j.jesp.2015.07.006

DeSimone, J. A., Harms, P. D., & DeSimone, A. J. (2014). Best Practice Recommendations for Data Screening. *Journal of Organizational Behavior, 36*(2), 171-181. doi:10.1002/job.1962

Brazil, E., & Fernström, M. (2011). Auditory Icons. In *The Sonification Handbook* (pp. 325-338). Logos.

Gaver, W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human-Computer Interaction*, *2*(2), 167–177. https://doi.org/10.1207/s15327051hci0202_3

Gaver, W. (1989). The SonicFinder: An Interface that Uses Auditory Icons. *Human-Computer Interaction, 4*(1), 67-94. doi:10.1207/s15327051hci0401_3

Gaver, W. W. (1993). Synthesizing Auditory Icons. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '93*. doi:10.1145/169059.169184

Gaver, W. W., & Smith, R. B. (1995). Auditory Icons in Large-scale Collaborative Environments. *Readings in Human–Computer Interaction,* 564-569. doi:10.1016/b978-0-08-051574-8.50059-5

Kostiainen, J. (2011). *Mobile Auditory Guidance for Public Transportation.* [Master's thesis, Aalto University].

Kramer, G., Walker, B., Bonewright, T., Cook, P., & Flowers, J. H. (2010). Sonification Report: Status of the Field and Research Agenda. *Faculty Publications, Department of Psychology*.

Meade, A. W., & Craig, S. B. (2012). Identifying Careless Responses in Survey Data. *Psychological Methods, 17*(3), 437-455. doi:10.1037/a0028085

Walker, B. N., & Nees, M. A. (2011). Theory of Sonification. In *The Sonification Handbook* (pp. 9–39). Logos.