TO SONIFY OR NOT TO SONIFY?: EDUCATOR PERCEPTIONS OF AUDITORY DISPLAY IN INTERACTIVE SIMULATIONS

Brett L. Fiedler[a], Bruce N. Walker[b], and Emily B. Moore[a]

[a]University of Colorado Boulder, Boulder, CO, USA
[b]Georgia Tech, Atlanta, GA, USA
brett.fiedler@colorado.edu, bruce.walker@psych.gatech.edu, emily.moore@colorado.edu

ABSTRACT

With the growing presence of auditory display in popular learning tools, it is beneficial to researchers to consider not only the perceptions of the students who use the tools, but the educators who include the tools in their curriculum. We surveyed over 4000 educators to investigate educator perceptions and preferences across four interactive physics simulations for the presence and qualities of non-speech auditory display, as well as surveying users' self-rated musical sophistication as potentially predictive of auditory display preference. We find that the majority of teachers preferred the simulations with auditory display and consistently rated aspects of the experience using simulations with sound positively over the without-sound variants. We also identify simulation design features that align with trends in educator ratings. We did not find the measured musical sophistication to be a predictor of auditory display preference.

1. INTRODUCTION

The use of auditory displays to support learning and accessibility within educational interactives, simulations, and games is growing. In this work, we investigate educator perceptions of the auditory display of widely-used physics simulations. This study provides insights into educator preferences when it comes to auditory displays within interactive learning resources, and to the potential benefits of auditory displays that are readily (and less readily) identified by educators.

As part of a large design and research project to implement auditory displays (including speech and non-speech sounds) within interactive science simulations to support learning and accessibility, there are currently ten physics simulations with sonifications and sound effects within the widely-used PhET Interactive Simulations collection [1]. The auditory displays were designed by an interdisciplinary team, with expertise in music and composition, physics, linguistics, education research, simulation and inclusive design, software development, and web accessibility. The iterative design process for the auditory display included feedback from physicists and teachers, and user interviews with youth, college students, and adults, including those with and without visual impairments [2, 3, 4, 5, 6]. Additionally, designs were also informed by authentic use of these simulations observed within formal and informal science classroom settings with middle school and high school youth, including students with learning disabilities [7, 8], those with visual impairments [9], and bilingual learners (primarily Spanish and English) [10].

Here, we investigate educators' perceptions of the resulting non-speech auditory display, which include sonifications and sound effects. Each simulation and its auditory display is unique, and sonifications emphasize the key mechanistic relationships represented within each simulation. From a survey of PhET educator users, we were interested in the following questions:

- Do educators prefer the simulations with or without auditory displays?
- Were some auditory displays preferred over others?
- For what teaching contexts were auditory displays considered feasible for use?
- Did educators' preferences for the auditory display correlate with musical sophistication?

2. METHODS

2.1. Participants

Participants were educators who use the PhET Interactive Simulation project website (`http://phet.colorado.edu`). Visitors to the PhET website can create a user account and opt-in to receiving email announcements. During account creation, they can provide information such as role (Teacher, Pre-service Teacher, Researcher, Student, etc.), STEM subject specialty, and grade level. We emailed an invitation to complete a research survey to the subset of users who selected one or more of the following options: Teacher, Pre-service Teacher, Teacher Educator, Other. In a second step to confirm participants were educators, an initial survey question asked participants to select their role, a selection of a non-educator role ended the survey allowing only those selecting educator roles to proceed. The survey was estimated to take about 15 minutes or less to complete. No compensation was provided. The total number of invited participants was 202,429. Of those invited, 4,658 responded to the survey beyond the role selection question, and 2,471 users completed the survey. We include partial responses in our statistical analysis.

This work is supported by NSF DRL-#1621363

2.2. Simulations and Auditory Displays

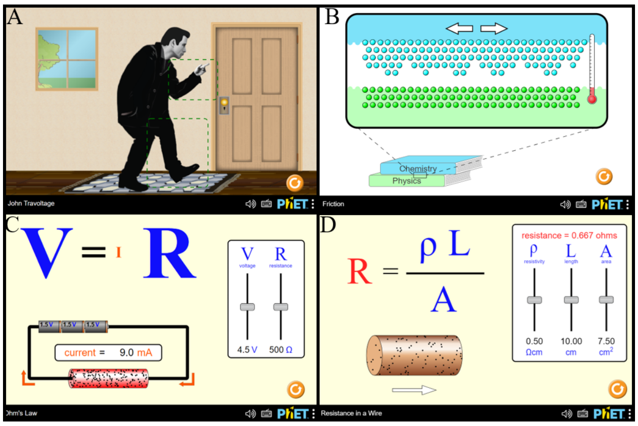

Figure 1: Four simulations used in the survey study: A) *John Travoltage*, B) *Friction*, C) *Ohm's Law*, D) *Resistance in a Wire*.

Simulations included in the survey (Figure 1) were selected from eight simulations published with sonifications at the time of survey creation (April 2020), representing the least complex simulations of the set.

John Travoltage. In John Travoltage [11, 12], Figure 1A, a character, John, stands on a rug by a door. Rubbing his foot results in negative charges transferring onto his body, and moving his arm towards the doorknob results in a shock. Learners can explore the relationship between the amount of charge on John's body and the distance between his hand and the doorknob that results in a shock. Highlighted in the auditory display is the sound of the foot rubbing on the rug, a pop sound as negative charges transfer onto John's body, a low continuous hum representing the charges on John's body, a ratchet-like sound when John's arm is moved, and an electrical zap sound as charges are discharged from John's body. Table 1A lists the auditory display features present for John Travoltage.

Friction. In Friction [13], Figure 1B, two textbooks can be rubbed together. A thermometer indicates the books' temperature, which increases as the books are rubbed together. A zoomed-in view shows a pseudo-molecular-level view, with the "molecules" jiggling more or less depending on the (higher or lower) temperature. Highlighted in the auditory display is a rubbing sound when the books are rubbed together and a sound representing the "molecules" jiggling, which changes as temperature increases. Table 1B lists the auditory display features present for Friction.

Ohm's Law. Ohms Law [14] and Resistance in a Wire [15] (Figure 1C-D) are a pair of simulations with similar visual layouts but different auditory display approaches [2]. In the Ohm's Law simulation, there is an equation ($V = I \times R$), a circuit with batteries and a resistor, and two sliders to change voltage, V, and resistance, R. Moving the sliders (changing voltage or resistance) results in changes to the current, I, and this is indicated by changes to the equation (letter sizes change), and the circuit (battery and resistor representations change). When voltage or resistance is changed, a repeating, 2-second sound clip plays, with changes in pitch and tempo mapped to the changes in current. Table 1C lists the auditory display features present for Ohm's Law.

Resistance in a Wire. In Resistance in a Wire, there is an equation ($R = \rho \times L/A$), a piece of wire, and three sliders to change resistivity, rho, length, L, and area, A. Moving the sliders (changing resistivity, length, or area), results in changes to the resistance, R, and this is indicated by changes to the equation (letter sizes change), and the piece of wire (the length, area, and amount of black dots indicating impurities in the wire change). When resistivity, length, or area are changed, a short marimba tone is played, with changes in pitch mapped to the changes to the value of resistance. Table 1D lists the auditory display features present for Resistance in a Wire.

All simulations also include a short earcon that plays when the Reset All button is pressed to indicate the simulation has been reset.

2.3. Survey Design

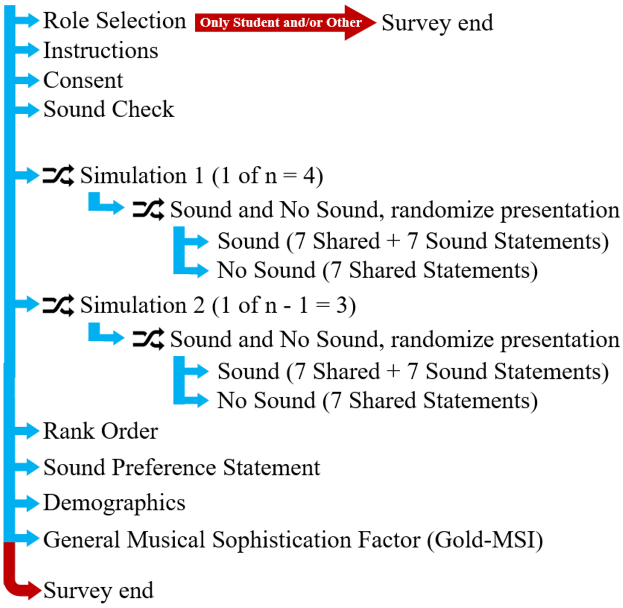

Figure 2: Survey flow chart. The selection of simulation, sound and no sound conditions, and order of statements was randomized.

A flow chart for the survey is presented in Figure 2. The survey is structured for participants to experience one simulation in sound (S) and no sound (NS) variants, and then repeat with a second simulation in S and NS variants, for a total of four simulation sections. The two simulations are randomly selected from a pool of four simulations, John Travoltage, Friction, Ohm's Law, and Resistance in a Wire, each described in Section 2.2. For example, one survey may include: 1. Resistance in a Wire (S), 2. Resistance in a Wire (NS), 3. Friction (NS), 4. Friction (S). To encourage simulation interaction, a minimum time of 30 seconds was required by the survey software once a new simulation variant appeared onscreen before educators were allowed to proceed with the survey. We further encouraged engagement with each simulation by following the appearance of each simulation variant with one simple multiple-choice conceptual question related to the simulation. For every simulation, educators were asked to rate seven Likert scale *shared statements* about their interaction with the simulation. If the simulation had sound, they were asked to

Feature	Sound Description	Type
A. ***John Travoltage***		
Leg swing	Carpet rubbing sound	Auditory Icon
Hand position	Ratchet, pitch increases as hand-doorknob distance decreases	Sonified Auditory Icon
Charge transfer	Pop sound, number, pitch increases/decreases as number increases/decreases	Sonified Earcon
Charges on body	Static-like, increasing number increases volume and playback rate	Sonification
Discharge	Electrical zap	Auditory Icon
Shock	“Ouch” and “Gazouch”	Speech
B. ***Friction***		
Grab/release micro-view book	Harp note, lower in pitch when released	Earcon
Grab/release macro view book	Instrument tone, lower in pitch when released	Earcon
Rubbing books together	Filtered noise generator; brush sound	Auditory Icon
“Molecules” Jiggling	Multiple versions of a marimba sound	Sonified Earcon
“Molecules” Evaporate	Synthesized tone spatialized; tinkling sound	Sonified Earcon
Books come in contact	Thud sound	Auditory Icon
“Molecules” cooling	White noise that slowly fades out simulating the sound of steam	Auditory Icon
C. ***Ohm’s Law***		
Current (I) value	Repeating 2-second sound clip, playback rate (pitch change) and tempo increases/decreases as value increases/decreases	Sonification
Voltage and Resistance sliders	Synthesized tick, neutral wood-block timbre	Earcon
D. ***Resistance in a Wire***		
Resistance (R) value	Discrete tone, marimba timbre, pitch increases/decreases as value increases/decreases	Sonification

Table 1: A. *John Travoltage*, B. *Friction*, C. *Ohm’s Law*, and D. *Resistance in a Wire* sound mapping.

rate an additional seven Likert scale *sound statements* about the sound in the simulation. Order of statements was randomized. After they had interacted with all four simulations, educators were asked to *rank* the four simulations (which would include two simulations with sound, two with no sound) and *rate* their agreement with a final preference statement regarding inclusion of sound features across all PhET simulations. Educators were then asked a set of 6 demographics questions. Lastly, they were presented with a set of 18 statements/questions from the *Goldsmith’s Music Sophistication Index*. The survey with all questions and survey logic is available [16].

Shared Statements. Each simulation variant was followed by a randomized set of seven shared Likert scale statements, with response options ranging from strongly disagree (1) to strongly agree (5), regarding their performance beliefs (Statements 1-2), experience (Statements 3-4), and affect (Statements 5-7) during their simulation use. Six of the seven statements were positively worded; Statement 6, regarding “frustrating to use,” is negatively worded. Shared statements are listed in Section 3.1.

Sound-specific Statements. If the simulation variant included sound, seven additional statements were included specifically related to the auditory display, with response options ranging from strongly disagree (1) to strongly agree (5). These statements also related to their performance beliefs (Statements 1-2), experience (Statement 3) and affect (Statement 4), as well as statements regarding the feasibility of the sounds in three contexts: a “physical classroom” (Statement 5), “with my students” (Statement 6), and during “virtual learning” (Statement 7). The feasibility statements (Statements 5-7) also included a not applicable “N/A” option. Statement 4 regarding “unpleasant” quality of the sound is negatively worded. Sound-specific statements are listed in Section 3.2, feasibility statements in Section 3.3.

Rank Ordering and Sound Preference Statement. The simulation portion of the survey ended with educators ranking the simulations by first choice (1) to fourth choice (4) in preference for the four simulations in the survey. This was followed by an optional text field, where educators could provide an explanation for their ranking. Finally, there was a Likert scale (1-5) statement for the educator to rate their overall preference for sound features in PhET Interactive Simulations after interacting with the four simulations.

Demographics. Educators were asked to provide their age, gender identity, primary level of their students (Elementary to University), PhET simulation use prior to survey (Y/N), primary language, and country of residence.

Goldsmith’s Music Sophistication Index. Lastly, we asked educators to rate their agreement with statements from the General Sophistication sub-scale from the Goldsmith’s Music Sophistication Index (Gold-MSI) [17, 18]. The General Sophistication inventory includes statements from other sub-scales in the Gold-MSI that include the themes of Active Engagement, Perceptual Abilities, Musical Training, Singing Abilities, and Emotion. The General Sophistication inventory was presented in the form of 15 Likert-scale statements (1-7; Strongly Disagree - Strongly Agree) and three multiple choice responses (scored 1-7) with possible scores ranging from 18 to 126. This inventory was chosen specifically because it is validated for situations in which non-musicians are being scored for perceptions of their musical engagement, musical training, self-reported abilities, and emotional engagement with music [18]. Educators rated their agreement to the statements with scores summed to give each respondent a general sophistication score that correlates with their level of musical sophistication. Statements 7, 9, 11, 13, and 14 are negatively worded and were reverse scored before summing the total General Sophisication score.

2.4. Analysis

We generated descriptive statistics for each scale-style response and used Mann-Whitney Wilcoxon (MWW) statistical tests when comparing responses within the same simulation (i.e., same statement from Ohms Law (Sound) vs. Ohms Law (No Sound) or responses within Ohms Law (No Sound), etc.) to determine differences in rating distributions. While MWW test results are highlighted, all test results were similarly conclusive at the 95% confidence level for two-sample t-tests. ANOVA tests were conducted for differences in distributions between simulations and for the Gold-MSI scores. Statistics were generated in the Qualtrics platform and statistical tests were performed using R [19].

3. RESULTS

Survey results are presented in order of our research questions. When interpreting these results, we consider that educator responses reflect a combination of their own preferences and use of auditory displays, as well as their perceptions of their students' preferences and use of auditory displays.

3.1. Do educators prefer the simulations with or without auditory display?

To answer this question, we present results from three measures of educator preference between simulations with and without auditory display. First, we compare educator ratings of shared statements (Shared Statements 1-7) presented after interaction with the simulations with and without sound. Next, we compare the rankings educators assigned to the four total simulations they each interacted with (two different simulations, each with and without sound) over the course of the survey. Lastly, we present educators' ratings on the inclusion of sound features within interactive simulations in general.

Shared Statements. For two different simulations, educators encountered each simulation with sound and with no sound. After each, educators were asked to rate statements (Shared Statements 1-7) that were identically worded after both the sound and no sound conditions. Statements related to performance beliefs, experience, and affect during use of the simulation. Results from these seven statements for two of the simulations, John Travoltage and Ohm's Law, are shown in Figure 3.

Comparing all simulations with and without sound, including those shown in Figure 3, the sound condition has more favorable responses (statistically significant difference in the means - $p<0.01$) across all statements compared to the no sound condition, except for one. The single exception is in responses to Statement 6 "This simulation was frustrating to use" for Resistance in a Wire - comparison of the mean ratings for sound (1.65) and no sound (1.68) responses are not significantly different. See Table 2 for the mean and difference in mean for the sound and no sound conditions for all simulations.

Looking across responses to Shared Statements 1-7, for both sound and no sound variants, educators responded most favorably to Statements 1 and 3 with means exceeding 4 (Somewhat Agree), indicating that the simulations would help students learn (Statement 1) and were easy to use (Statement 3). The least favorable means were for Statement 4 "While interacting with this simulation, I felt enthusiastic". This statement also had the largest difference in means for sound (e.g., 4.07 for John Travoltage sim) and no sound (e.g., 3.50 for John Travoltage sim); educators on average indicated feeling more enthusiastic when using the simulation with sound compared to no sound. Statements most directly related to affect (Statements 5-7) show the largest difference in mean when comparing both conditions for all simulations; educators agreed more with positive affect statements when the simulation had sound than no sound. Full statistics for Shared Statements for both sound and no sound variants are available [16].

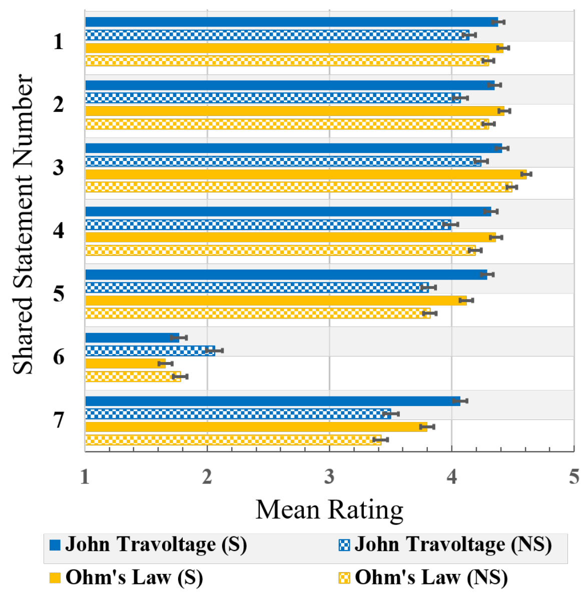

Figure 3: Mean ratings for Ohm's Law and John Travoltage simulations, with sound (S) and with no sound (NS). Shared Statement 6 is negatively worded. See [16] for complete tables for all simulations and conditions.

John Travoltage and Friction have the lowest mean ratings overall for Shared Statements 1-7 for simulations without sound (though still favorable) and they also show the largest difference (comparing with sound mean rating to no sound mean rating) in comparison to Ohm's Law and Resistance in A Wire. In comparison, Ohm's Law and Resistance in a Wire have higher mean ratings without sound when compared to Friction or John Travoltage without sound. There are also smaller difference in mean change in responses between Ohm's Law and Resistance in a Wire.

Rank Ordering. As a direct indicator of preference between simulations with and without sound, we asked educators to rank the four simulations they had experienced from their First Choice (1) to Fourth Choice (4). For example, if the survey had included John Travoltage and Ohm's Law, each educator participant receiving this survey variant would have experienced both of these simulations with and without sound, for a total of four simulations. Average rankings of the simulations are displayed in Table 3. Mean values closer to 1 indicate the simulation was more frequently ranked higher (more preferred) than other simulations in a participant's subset. MWW nonparametric test p-values are <0.01 for all simulation pairs, indicating a statistically significant increase in mean ranking of all simulations with sound. Each mean represents the aggregate of all participants who experienced a particular simulation (e.g., Ohm's Law (No Sound)). Note, participants were presented with randomized sets of simulations, all experiencing the simulation with and without sound for two out of the pool of four simulations, but not necessarily experiencing the same two simulations for comparison.

Simulation	Mean (No Sound)	Mean (Sound)	Difference (Col. 2 - Col. 3)
1. This [sim] would help students learn this concept.			
John Travoltage	4.14	4.38	0.24
Friction	4.01	4.3	0.29
Ohm's Law	4.30	4.42	0.12
Resistance in a Wire	4.38	4.46	0.08
2. It was easy to understand the concepts in this [sim].			
John Travoltage	4.07	4.35	0.28
Friction	3.92	4.21	0.29
Ohm's Law	4.3	4.43	0.13
Resistance in a Wire	4.39	4.47	0.08
3. It was easy to interact with and use this [sim].			
John Travoltage	4.24	4.41	0.17
Friction	4.07	4.26	0.19
Ohm's Law	4.49	4.61	0.12
Resistance in a Wire	4.54	4.64	0.10
4. I would recommend this [sim] to students and teachers.			
John Travoltage	3.99	4.32	0.33
Friction	3.86	4.2	0.34
Ohm's Law	4.19	4.36	0.17
Resistance in a Wire	4.29	4.36	0.07
5. This [sim] was enjoyable to use.			
John Travoltage	3.81	4.29	0.48
Friction	3.66	4.13	0.47
Ohm's Law	3.82	4.12	0.30
Resistance in a Wire	3.92	4.16	0.24
6. This [sim] was frustrating to use. (negative)			
John Travoltage	2.17	1.77	-0.29
Friction	2.17	1.95	-0.22
Ohm's Law	1.78	1.66	-0.12
Resistance in a Wire	1.68	1.65	-0.03*
7. While interacting with this [sim], I felt enthusiastic.			
John Travoltage	3.50	4.07	0.57
Friction	3.35	3.89	0.54
Ohm's Law	3.42	3.8	0.38
Resistance in a Wire	3.58	3.85	0.27

Table 2: Summary of ratings statistics for Shared Statements 1-7 that appeared on both no sound (NS) and sound (S) for each displayed simulation. *MWW test for S6 of Resistance in a Wire was statistically non-significant. See [16] for tables for all statements.

Simulation (Condition)	Ranking Mean / Median
John Travoltage (Sound)	1.65 0.05 / 1
Friction (Sound)	1.87 0.05 / 2
Ohm's Law (Sound)	2.12 0.06 / 2
Resistance in a Wire (Sound)	2.23 0.06 / 2
Ohm's Law (No Sound)	2.83 0.05 / 3
Resistance in a Wire (No Sound)	2.83 0.05 / 3
John Travoltage (No Sound)	3.23 0.05 / 3
Friction (No Sound)	3.23 0.05 / 3

Table 3: Mean and median ranking of simulations from First Choice (1) to Fourth Choice (4). Total number of participant rankings for each simulation were n=1,234 for John Travoltage, n=1,244 for Friction, n= 1,266 for Ohm's Law and n=1,278 for Resistance in a Wire.

The mean ranking for the simulations with sound is ranked higher (closer to 1) when compared to the same simulation without sound for all simulations with statistically significant differences in the mean and median rankings. The difference in ranking for each variant was greatest for John Travoltage and Friction. Despite the randomized presentation of simulations, it is notable the larger overall preference for John Travoltage and Friction with sound, in comparison to a more modest preference for Ohm's Law and Resistance in a Wire with sound.

Sound Preference Statement. After experiencing the simulations with and without sound, educators were asked to rate their agreement with the statement: I believe as many PhET sims as possible should have sound features. 77.5% of educators rated the statement favorably (selecting 4 or 5 from the five-point scale).

Overall, all four simulations were on average rated favorably for all Shared Statements, but the presence of an auditory display did increase ratings in almost every case. With the additional evidence provided by the higher rank order of simulations with auditory display and the highly favorable rating of the Sound Preference Statement, we find that educators do prefer the presence of auditory display in simulations. However, we also find disparities in the ratings among simulations and trends were found that grouped the simulations, often pairing John Travoltage/Friction and Ohm's Law/Resistance in a Wire. We will consider this for discussion further in Section 4.

3.2. Were some auditory displays preferred over others?

To answer this question, we investigate educator ratings of a four-question subset of the seven sound-specific statements shown immediately after simulations with sound. This subset of Sound Statements consisted of: 1) The sounds were helpful, 2) It was easy to match the sounds to their meanings, 3) The sounds were interesting, 4) The sounds were unpleasant. Three additional sound-specific statements related to teaching with each simulation were included in the survey; results from these teaching-focused statements are provided in Section 3.3.

Mean ratings for the four-question subset of sound-specific statements are shown in Table 4. Full statistics for sound-speicifc statements are available [16]. Most notably, the mean ratings were favorable for all simulations with sound. The mean ratings for the simulation John Travoltage were most favorable across all statements, followed closely by ratings for Friction. Ohm's Law and Resistance in a Wire were consistently rated similarly, and always less favorably than John Travoltage and Friction. The frequency distribution of ratings, shown for ratings in response to the statement: The sounds were helpful, is shown in Figure 4, highlights the overall favorable ratings for all four simulations. Figure 4 is also presented to highlight that the greater width of the distributions for Ohm's Law and Resistance and a Wire comes from an increase in unfavorable ratings, but notably no presence of a bimodal distribution.

The disparities in auditory-specific ratings that again appear to group John Travoltage/Friction and Ohm's Law/Resistance in a Wire are interpreted in the Discussion (Section 4).

3.3. For what learning contexts did educators consider the use of simulations with auditory displays to be feasible?

To answer this question, we look to educator ratings of the three learning context-specific statements shown immediately after simulations with sound. These statements all began with the phrase

Simulation	Mean Rating / SD
1. The sounds were helpful.	
John Travoltage	4.23 0.05 / 0.95
Friction	4.02 0.06 / 1.07
Ohm's Law	3.74 0.06 / 1.18
Resistance in a Wire	3.59 0.07 / 1.23
2. It was easy to match the sounds to their meanings.	
John Travoltage	4.37 0.05 / 0.93
Friction	4.03 0.06 / 1.10
Ohm's Law	3.85 0.06 / 1.16
Resistance in a Wire	3.65 0.07 / 1.27
3. The sounds were interesting.	
John Travoltage	4.22 0.05 / 0.93
Friction	4.01 0.05 / 0.97
Ohm's Law	3.70 0.06 / 1.13
Resistance in a Wire	3.68 0.06 / 1.12
4. The sounds were unpleasant. (negative)	
John Travoltage	1.99 0.06 / 1.14
Friction	1.99 0.06 / 1.15
Ohm's Law	2.40 0.07 / 1.27
Resistance in a Wire	2.34 0.07 / 1.25

Table 4: Mean ratings and standard deviations for subset of auditory-specific Sound Statements 1-4 provided after simulation experiences with sound.

"Listening to these sounds would be feasible for use..." and continue with "in my physical classroom", "with my students", and "in virtual learning". To maintain a moderate length survey, we opted for a short list of statements that included two basic teaching contexts, in the classroom and virtual, along with a third less-specific statement "with my students" that was intended to be inclusive of all contexts educator's support student learning. Mean ratings in response to these statements are shown in Table 5.

Simulation	Mean Rating / Std. Dev.
Listening to these sounds would be feasible...	
...in my physical classroom.	
John Travoltage	4.29 0.07 / 1.24
Friction	4.13 0.07 / 1.29
Ohm's Law	3.90 0.07 / 1.37
Resistance in a Wire	3.93 0.08 / 1.41
...with my students.	
John Travoltage	4.47 0.06 / 1.03
Friction	4.35 0.06 / 1.10
Ohm's Law	4.16 0.06 / 1.16
Resistance in a Wire	4.16 0.07 / 1.24
...for virtual learning.	
John Travoltage	4.62 0.05 / 0.98
Friction	4.49 0.06/ 1.02
Ohm's Law	4.36 0.06 / 1.04
Resistance in a Wire	4.32 0.06/ 1.13

Table 5: Mean ratings and standard deviations for feasibility subset of sound-specific statements provided after simulation experiences with sound.

Educators' responses indicated that use of simulations with sound was feasible across multiple learning contexts, with virtual learning rated most favorable for feasibility. We share some further interpretations of this finding in the Discussion (Section 4)

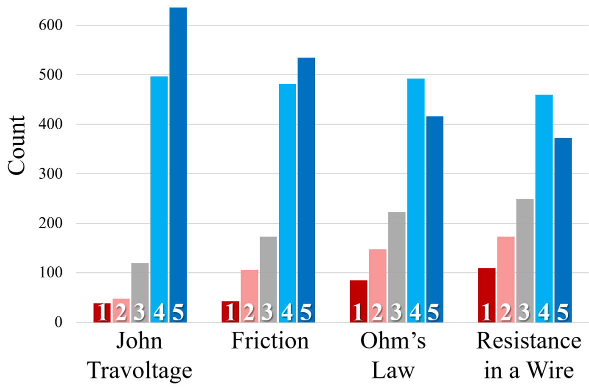

Figure 4: Histogram of frequencies of ratings (1-5) for the sound-specific statement "The sounds were helpful".

3.4. Did educators' preferences for the auditory displays correlate with musical sophistication?

With this survey we investigated characteristics potentially predictive of whether or not a user will prefer auditory displays. The last set of questions presented in the survey was the General Sophistication sub-scale of the Goldsmith's Musical Sophistication Index (Gold-MSI) [17]. A histogram of individual total scores for the General Sophistication sub-scale is shown in Figure 5A. The histogram shows a roughly normal distribution skewed slightly towards lower scores.

First, we were interested in possible trends in general musical sophistication score and ratings for the sound-specific statements. For example, do educators who Strongly Agree (5) that the sounds in Resistance in a Wire are "unpleasant", tend to have a higher or lower musical sophistication score? For each sound-specific statement, we looked for differences in statement rating as a factor of an individual's total General Sophistication sub-scale score. Significance was determined by analysis of variance (ANOVA) to reject the hypothesis that individuals' statement ratings were random with respect to general sophistication score. As an example, Figure 5B shows a box plot with a density overlay (width $\sim$ density) for ratings of "The sounds in this simulation were unpleasant" for Resistance in a Wire against individuals' Gold-MSI scores. With this analysis, we did not find any trend for significant differences across ratings based on scoring for all responses.

Secondly, we were interested if the educators who scored at the extremes of the distribution of the General Sophistication sub-scale might have more varied or extreme survey responses that would otherwise be masked when considering the full data set in aggregate. We separated out the respondents based on high and low General Sophistication sub-scale scores, taking those one standard deviation above (>87) and below (<45) the mean and running the analysis again, looking for trends in ratings to the sound-specific statements. We found a few statements that showed a significance among one rating group, but no trends were observed across statements that would indicate high/low scorers consistently showed some variance in opinions from the rest of the population.

Lastly, we hypothesized that sub-themes within the statements provided in the General Sophistication sub-scale might be better predictors of statement rating. Prior research has suggested that the specific statements associated with the Active Engagement sub-

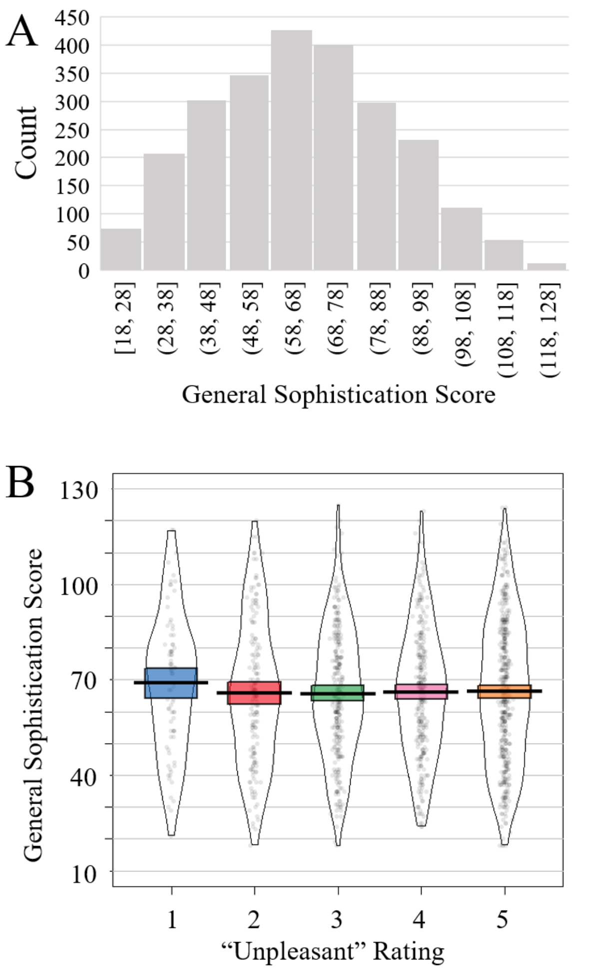

Figure 5: A) Histogram of participants' MSI scores from the General Sophistication sub-scale of the Gold-MSI. Mean = 66, St. Dev. = 21, Range 18 - 125. B) Example box plot of general sophistication MSI score vs. rating of sound as "unpleasant" for *Resistance in a Wire*.

scale could be used as a significant predictor of listener performance scores in a sonification mapping task [18]. The General Sophistication sub-scale consists of statements from other subscales (e.g., Active Engagement, Perceptual Abilities, or Emotions from Section 2.3), therefore we were able to use individual statements (rated 1-7) or groups of statements to see if another theme presented significant differences among ratings [17]. Again, significance was found sporadically for individual statements. For example, responding favorably on "I can compare and discuss differences between two performances or versions of the same piece of music" correlated with responding "Strongly Disagree" to the "It was easy to match the sounds to their meanings" statement for John Travoltage. However, there are no trends in significance across groupings of simulations or groups of statements, so we do not find any subset of statements to be adequate predictors of shared or sound statement rating.

Overall, we were unable to find meaningful correlations between responses related to the simulations auditory display and the Gold-MSI General Sophistication Sub-scale. We did not find that educator preferences correlated with musical sophistication.

4. DISCUSSION

We found that all four simulations were on average rated favorably with regard to respondent's perceptions of performance, experience, and affect regardless of the presence of auditory display. The presence of an auditory display resulted in increased favorable ratings on average. The perceptions of the auditory display itself was also rated favorably for every simulation. However, simulations were not rated identically with auditory ratings, rankings, and responses typically averaging more favorably for the John Travoltage and Friction pair in comparison to the Ohm's Law and Resistance in a Wire pair.

We can interpret the pairings first by the similarities in the visual design and interaction design of the pairs. John Travoltage and Friction share design attributes that are, in turn, distinct from those shared by Ohm's Law and Resistance in a Wire. John Travoltage and Friction each visually display a more real-world scenario for exploration than Ohm's Law and Resistance in a Wire (the interactive objects are legs, arms, and books compared to sliders and buttons), John Travoltage and Friction each can be used with younger learners (middle school to college, compared to high school to college learners for Ohm's Law and Resistance in a Wire) and the conceptual goals of John Travoltage and Friction are less visually explicit than those of Ohm's Law and Resistance in a Wire (though all address introductory level physics content). It is possible that, for this population of educators, the visually explicit nature of Ohm's Law and Resistance in a Wire may explain the smaller difference between sound and no sound, in the Shared Statements, when compared to John Travoltage/Friction.

Aside from content and visual design, all four simulations have unique auditory display designs, but multiple similarities and differences in the auditory displays can be identified. For example, Friction and Ohm's Law both have more musical qualities in their auditory displays (Friction with the background marimba tones of the "molecules" jiggling, Ohm's Law with the repeating 2-second sound clip associated with changes in Current value), while Friction and Resistance in a Wire both share significant use of marimba tones. John Travoltage and Friction both use auditory icons which are absent from the other two simulations. The groupings that emerged and the consistency of these groupings align with the more real-life scenarios and associated auditory icons found in John Travoltage and Friction (e.g., brushing sounds for physical contact between foot/carpet or book/book), that are not found in Ohm's Law and Resistance in a Wire. Preliminary analysis of short answer responses from educators show an emerging theme consistent with this interpretation: auditory icons and sounds associated with objects and relationships that have sound in everyday life seem to be considered more favorably and more beneficial for student learning in comparison to more abstract or less "real life" sounds and sonifications.

In responses to the question, "Please explain why you ranked the simulations in the order above", following the simulation ranking task, many educators offered explanations for their rankings that related to how helpful, easy to understand, and pleasant the sounds were or were not. Qualitative analysis of the 2187 responses to the optional text prompt is in progress and will be reported elsewhere. Nevertheless, a theme that is emerging from our qualitative analysis is that educators' perceptions of sounds and sonifications in the simulations are more positive (and rated as more appropriate for student learning) when those sounds more easily map onto their lived experience. In contrast, sounds are less preferred in the simulations that are more "abstract", without sounds associated with them in real life. Notably, there are few educator responses indicating the auditory display being useful in different ways for individual students' needs or preferences.

Here, we share two quotes from educator responses, highlighting this theme. The first quote is from an educator whose survey included the simulations John Travoltage and Resistance in a Wire. This educator rated all sound-specific statements for John Travoltage favorably (Sound Statements 1-4: 5, 5, 5, 5) and most sound statements for Resistance in a Wire unfavorably (2,1,2,4), and wrote:

> ...(2) The sounds in John Travoltage relate to concrete physical actions and sollicite [sic] experience, helping studente [sic] connect the new concept to old knowledge. (3) The Resistance app is a pure visual representation of an abstract relationship, and the sounds are meaningless. Worse, they add a layer of extra information that needs to be deciphered and that distracted me from focusing on what was happening elsewhere, like the visual changes in the wire which I only really looked at when there was no sound.

This educator explicitly calls out the auditory display associated with "concrete physical actions" found in John Travoltage as a positive feature. In contrast, this educator indicates that the Resistance in a Wire simulation is about an "abstract relationships" and the the associated auditory display is "meaningless". Further, this educator found the addition of the auditory display in this case to be distracting.

In a second quote from a different educator who rated all sound-specific statements for Ohm's Law unfavorably (Sound Statements 1-4: 1,1,2,2), writing about the simulation Ohm's Law:

> The sounds don't indicate any meaning to me. A lower pitch goes with decreased quantity, but that seems arbitrary. The sounds are one more thing to distract students from the CONCEPT to be learned.

In this quote, the educator indicates that they correctly identified the sonification mapping in Ohm's Law (as the value of current decreases, the pitch of the repeating 2-second sound clip decreases). Similar to the first quote, this educator perceives the mapping unfavorably, writing that it is "arbitrary" and may not contribute to conceptual understanding for learners, specifically the sounds being "one more thing to distract" from conceptual learning.

Feasibility for use in Teaching Contexts. Educators' responses to the open-ended question after the simulation ranking task may provide further insight into the findings regarding feasibility in different learning contexts. From our qualitative analyses so far, we have found that some educators had concerns about managing the auditory display within classroom settings with many students, with some educators writing about potential issues related to too many sounds in a classroom, or a lack of headphones for all of their students. These concerns likely contributed to the comparatively lower ratings for the statement referring to feasibility "in my physical classroom". Given the high ratings for feasibility of using simulations with sound in virtual learning contexts, educators may perceive that challenges related to managing sound may be less in virtual contexts, or potentially that there are greater benefits to using simulations with sound in virtual learning contexts. Given the feasibility of using simulations with sound "with my students", the most general learning context statement, consistently was rated with a mean between "in my physical classroom" and "in virtual learning", educators may have been consistently imagining more in-person contexts such as in one-on-one discussions, or as supporting resources, etc. – contexts that may be perceived as intermediate to whole-class teaching and fully virtual.

Notably, the survey was distributed in April 2020, in the midst of the initial expansion of virtual learning in response to the COVID-19 pandemic. Some educators taking the survey may have been experiencing teaching in virtual learning contexts for the first time, or had not yet experienced it but would go on to engage in teaching virtual classes later in 2020/21. It would be interesting to distribute a variation of the original survey in a post-COVID-19 school year, and investigate potential differences in responses regarding feasibility as presumably more educators would have had recent personal experiences with virtual learning.

Auditory Display to Support Inclusive Learning Environments. An additional related finding emerging from our analysis of educators' open-ended responses is a lack of indications that educators were considering potential benefits of auditory display for learners with different needs (e.g., sensory disabilities, print-related disabilities, or cognitive impairments), or regarding the presence of auditory display as supporting inclusive learning environments. In particular, with respect to educators' comparatively lower ratings for Ohm's Law and Resistance in a Wire, which also have the most abstract sonifications of the group of four simulations, we wonder if educators were not considering the possible ways that auditory displays could prove to be beneficial (or in some cases, such as with learners with significant visual impairments, a necessary) complement to the visual display. We are investigating this further in our analysis of the open-ended responses to this survey.

5. CONCLUSIONS

This study found that the majority of teachers preferred the simulations with auditory display compared to the same simulation without auditory display and teachers consistently rated the sound variants of the simulations slightly more helpful, easy to understand, and enjoyable than the without-sound versions of the simulations. Educators also found the simulations with sound to be slightly more feasible in virtual, rather than physical contexts. Lastly, musical sophistication did not appear to be a significant predictor of auditory-specific ratings. Building upon the differences in ratings between simulations revealed in this work, future work includes the continued analysis of the open-text responses and further investigating the difference in user preference of abstract and real-life sound design. Finally, we are continuing to investigate user characteristics potentially predictive of auditory preference in interactive simulations, such as users' familiarity with the simulation's disciplinary content.

6. ACKNOWLEDGEMENT

We are grateful to the survey participants, Brianna Tomlinson for her early contributions to the study and survey design, and Tiara Sawyer for her efforts in the preliminary qualitative analysis mentioned.

7. REFERENCES

[1] PhET Interactive Simulations (2021). John Travoltage. https://phet.colorado.edu/en/accessibility

[2] Tomlinson, B. J., Walker, B. N. & Moore, E. B. Auditory Display in Interactive Science Simulations: Description and Sonification Support Interaction and Enhance Opportunities for Learning. Conf. Hum. Factors Comput. Syst. - Proc. 112 (2020). doi:10.1145/3313831.3376886

[3] Smith, T. L., Greenberg, J., Reid, S., & Moore, E. B. (2018). Parallel DOM architecture for accessible interactive simulations. In Proceedings of the Internet of Accessible Things (pp. 1-8).

[4] Winters, R. M., Tomlinson, B. J., Walker, B. N., & Moore, E. B. (2019). Sonic Interaction Design for Science Education. Ergonomics in Design, 27(1), 5-10.

[5] May, K. R., Tomlinson, B. J., Ma, X., Roberts, P., & Walker, B. N. (2020). Spotlights and Soundscapes: On the Design of Mixed Reality Auditory Environments for Persons with Visual Impairment. ACM Transactions on Accessible Computing (TACCESS), 13(2), 1-47.

[6] Tomlinson, B. J., Kaini, P., Harden, E. L., Walker, B. N., & Moore, E. B. (2019). A Multimodal Physics Simulation: Design and Evaluation with Diverse Learners. The Journal on Technology and Persons with Disabilities, 88.

[7] Harden, E. L., & Moore, E. (2019, June). Co-adapting a Design Thinking Activity to Engage Students with Learning Disabilities: Insights and Lessons Learned. In Proceedings of the 18th ACM International Conference on Interaction Design and Children (pp. 464-469).

[8] Tomlinson, B. J., Kaini, P., Walker, B. N., Batterman, J. M., & Moore, E. B. (2018). Supporting Simulation Use for Students with Intellectual and Developmental Disabilities. The Journal on Technology and Persons with Disabilities, 202.

[9] Winters, R. M., Harden, E. L., & Moore, E. B. (2020, October). Co-Designing Accessible Science Education Simulations with Blind and Visually-Impaired Teens. In The 22nd International ACM SIGACCESS Conference on Computers and Accessibility (pp. 1-4).

[10] Fiedler, B. L., Bennett, M. B., Johnson, N. E., & Moore, E. B. Coordinating epistemic frames in informal physics: Agency, support, and technology.

[11] PhET Interactive Simulations. (2020a). John Travoltage. https://phet.colorado.edu/sims/html/john-travoltage/latest/john-travoltage_en.html

[12] Tomlinson, B., Kaini, P., Harden, E. L., Walker, B., & Moore, Emily B. (2019) Design and evaluation of a multimodal physics simulation. Journal on Technology People with Disabilities. Journal on Technology and Persons with Disabilities, Vol 7 (pp 88-102).

[13] PhET Interactive Simulations. (2020b). Friction. https://phet.colorado.edu/sims/html/friction/latest/friction_en.html

[14] PhET Interactive Simulations. (2020c). Ohm's Law. https://phet.colorado.edu/sims/html/ohms-law/latest/ohms-law_en.html

[15] PhET Interactive Simulations. (2020d). Resistance in a Wire. https://phet.colorado.edu/sims/html/resistance-in-a-wire/latest/resistance-in-a-wire_en.html

[16] Fiedler, B. L., & Moore, E. B. (2021). Educator Perceptions of Auditory Display in Interactive Simulations [materials and data]. Open Science Framework. osf.io/h84vz

[17] Mllensiefen, D., Gingras, B., Stewart, L. & Musil, J. (2014). The Goldsmiths Musical Sophistication Index (Gold-MSI): Technical Report and Documentation v1.0. London: Goldsmiths, University of London.

[18] Schuett, J. (2019). Measuring the effects of display design and individual differences on the utilization of multi-stream sonifications. PhD dissertation. School of Psychology, Georgia Institute of Technology. Atlanta, GA. Georgia Tech SMARTech Repository. URL: https://smartech.gatech.edu/handle/1853/61808

[19] R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/