Understanding Pattern Recognition Through Sound with Considerations for Developing Accessible Technologies

Nicole Darmawaskita and Troy McDaniel(✉)

Arizona State University, Mesa, AZ 85212, USA
{ndarmawa,troy.mcdaniel}@asu.edu

Abstract. This work explores whether audio feedback style and user ability influences user techniques, performance, and preference in the interpretation of node graph data among sighted individuals and those who are blind or visually impaired. This study utilized a posttest-only basic randomized design comparing two treatments, in which participants listened to short audio clips describing a sequence of transitions occurring in a node graph. The results found that participants tend to use certain techniques and have corresponding preferences based on their ability. A correlation was also found between equivalently high feedback design performance and lack of overall feedback design preference. These results imply that universal technologies should consider avoiding utilizing design constraints that allow for only one optimal usage technique, especially if that technique is dependent on a user's ability.

Keywords: Assistive technology · Accessible education · Sonification

1 Introduction

In the U.S., educators are required to adapt their lessons to national accessible learning standards [1], but many students who are blind or visually impaired are unable to complete their education, with 22.3% not completing high school and only 15.7% earning a bachelor's degree in 2016 [2]. Accessible education, therefore, has much room for improvement to ensure those with visual impairments have equal opportunity. Node graphs are a commonly used educational aid that have yet to see a popular, modern, and accessible counterpart. In response, this research aims to develop effective accessible node graphs to improve the educational environment for these students.

Prior research found the use of nonspeech feedback, compared to the use of speech-only, in an accessible table, reduced a user's workload, but both feedback methods provided the same effectiveness in performance [3]. Additional research exploring the usage of multimodal feedback in the presentation of graphical information found most participants implemented the same generic technique in their interactions with accessible graphics [4]. A study comparing the usage of various feedback methods to present graph

C. Stephanidis et al. (Eds.): HCII 2020, LNCS 12426, pp. 208–219, 2020.
https://doi.org/10.1007/978-3-030-60149-2_17

concepts found no significant difference in preference nor effectiveness between feedback methods [5], demonstrating a correlation in which participants equally preferred all feedback methods that performed comparably and highly effectively. Additionally, all the highest performing participants used the same generic technique to interpret feedback, while the majority of lower performing participants used a different, albeit less performance-optimized, generic technique [5]. These observations suggest that certain design constraints encourage the usage of certain generic techniques in individuals based on their ability. When considering the wide variety of user ability, the correlation between performance and preference can be explained through individuals performing better and/or finding additional value—such as reduced workload—when interacting with certain design constraints, leading an overall userbase to equally prefer a variety of designs.

The current study explored whether audio feedback style and user ability influenced user techniques, performance, and preference in the interpretation of node graph data among sighted individuals and those who are blind or visually impaired. Specifically, this research seeks to find if relationships exist between feedback style, user techniques, performance, feedback preference, and user ability.

2 Method

2.1 Participants

47 participants were recruited for this IRB-approved study through email lists and referrals by colleagues and friends. Of the participants, 8 individuals were blind or visually impaired, and 39 were sighted. Most of the participants were local Arizona residents, with most of the sighted participants Arizona State University students. Sighted participants received $10 cash compensation, while those who were blind or visually impaired received $25.

2.2 Materials

Node Graph Structure. Participants listened to audio clips describing a sequence of transitions occurring in a unique node graph. For consistency, each node graph contained at most three nodes named 1, 2, and 3, respectively. Each sequence began at node 1 and contained four transitions. Each node in a node graph could transition by either taking a “Left” or a “Right.”

Patterns. Each sequence contained either a “Dead End” pattern, “Alternating” pattern, or “No Pattern.”

The “Dead End” pattern occurred in a sequence that contained a self-looping node on both “Left” and “Right.” Since the “Dead End” pattern could occur in any node, participants were provided with three sub-patterns to clarify which of the three nodes was the “Dead End”: “Dead End at 1”, “Dead End at 2”, and “Dead End at 3” (Fig. 1).

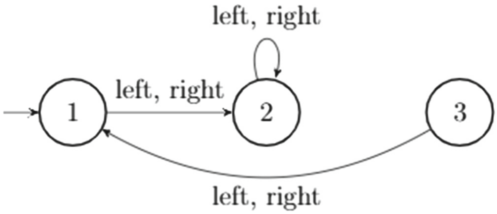

Fig. 1. A sample visualization of a node graph containing the “Dead End at 2” sub-pattern.

The “Alternating” pattern occurred in a sequence that contained the alternation between two nodes in succession through repetitions of either “Left” or “Right.” Since the “Alternating” pattern could occur with either transition direction, participants were provided with two sub-patterns to clarify which of the transition directions the sequence was alternating on: “Alternating Left” and “Alternating Right” (Fig. 2).

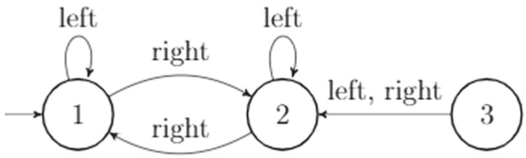

Fig. 2. A sample visualization of a node graph containing the “Alternating Right” sub-pattern.

“No Pattern” occurred in a sequence that contained neither the “Dead End” nor “Alternating” pattern (Fig. 3).

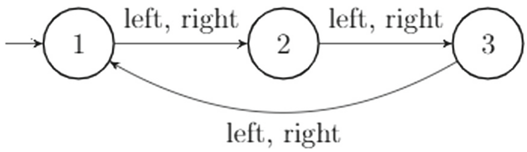

Fig. 3. A sample visualization of a node graph containing “No Pattern.”

Audio Feedback Styles. The sequences were presented in two audio styles: Speech-Only and Speech-And-Nonspeech-Sounds. The Speech-Only clips were presented in the form of: “Start at 1. <“Left” or “Right”> <“1”, “2”, or “3”>. <“Left” or “Right”> <“1”, “2”, or “3”>. <“Left” or “Right”> <“1”, “2”, or “3”>. <“Left” or “Right”> <“1”, “2”, or “3” >.” The phrase “Start at 1” indicated the sequence starts at node 1. The use of either a “Left” or “Right” indicated the direction of transition, and the number following indicated the destination node of the transition. The Speech-And-Nonspeech-Sounds clips were presented in a similar form, but “Left” was replaced with a piano tone of High C, and “Right” was replaced with a piano tone of High F#. The audio clips were generated through the combination of the Panopreter Basic text-to-speech software, recordings of piano tones, and the Audacity audio software editor.

For example, the Speech-Only phrase "Start at 1. Left 2. Right 2. Left 2. Right 2." would indicate the "Dead End at 2" pattern as the graph self-loops on node 2 on both transition directions. The Speech-And-Nonspeech-Sounds phrase "Start at 1. <High C> 2. <High C> 3. <High F#> 1. <High F#> 2." would indicate a "No Pattern" sequence as the graph contains neither a self-loop nor successive alternations between nodes.

Audio Feedback Style Training. Participants were introduced to each audio feedback style through respective training phases. Each training phase described its respective audio style structure, how the three patterns would be represented through the audio style and provided three sample audio clips showcasing each pattern for the participant to listen to.

Audio Feedback Style Testing. Participants tested each audio feedback style through respective testing phases. This phase contained nine multiple choice questions displayed in randomized order. Each question contained a unique audio clip of its condition's respective audio style and asked, "Which of the following patterns did you recognize from the audio clip?" The response options were "No pattern (or I don't know)," "Alternating Left," "Alternating Right," "Dead End at 1," "Dead End at 2," and "Dead End at 3."

Post-experiment Survey. A seven-question post-experiment survey was given to participants. Participants were asked to rate on a 5-point scale, from 1 very unintuitive to 5 very intuitive, how intuitive they found each audio feedback style to be in enabling them to identify patterns. Participants were also asked to rate on a 5-point scale, from 1 very difficult to 5 very easy, how easily they were able to identify the nonspeech sounds in the Speech-And-Nonspeech-Sounds audio style. Participants were asked to elaborate on any techniques they used in detecting patterns, which audio style they preferred, and to elaborate on their preference. These responses were used to analyze the participants' spatial hearing abilities, the intuitiveness of the audio styles, common techniques used, design preferences, and the relationship between these components.

2.3 Procedure

Design. This study utilized a posttest-only basic randomized design comparing two treatments. The treatments were the audio feedback styles of Speech-Only and Speech-And-Nonspeech-Sounds. The dependent variables were audio feedback preference and intuitiveness, techniques used, relative pitch ability, and accuracy in the identification of node graph data. Accuracy was used to determine participant and audio feedback performance level. Since the participant pool of individuals who are blind or visually impaired tends to be small, the study utilized a repeated measures design to maximize the sample size for each treatment level. Participants were randomly placed into groups through a trickle process to determine the completion order of the two audio feedback conditions, compensating for any order effects.

Procedure. At the beginning of the study, the participant digitally signed the consent form after reading through it with the study facilitator. The participant then received their cash compensation for the study, and the study began with a short demographic survey.

The participant then listened to short audio clips describing a sequence of transitions occurring in a node graph. Each of these audio clips presented a certain pattern. These transition sequences were presented in two conditions, each with a unique audio feedback style as mentioned above.

The start of each condition included a training phase providing an overview of the condition's audio feedback style, patterns, and sample audio clips of patterns in the condition's audio style. Once the participant understood the information conveyed through the audio style and was able to identify the patterns, they continued onto the testing phase of the study, in which they listened to a total of nine audio clips. Upon listening to each clip, participants were asked to identify, to the best of their ability, the pattern they heard through the audio. Within each condition, the order of the audio clips was also presented in a randomized order to counterbalance any order effects within the conditions. The participant was able to ask for the clips to be repeated as many times as they requested. The participant was scored on their accuracy in identifying patterns but remained unaware of their scores.

Once the two conditions were completed, the participant was asked to complete the post-experiment survey. The participant was then informed that the study had been completed, had the opportunity to ask any further questions, and was thanked for their participation.

3 Results

In analyzing these results, all instances of F-tests were F-tests for equality of variance, all instances of t-tests were 2-tailed paired t-tests, and the significance level of all tests was 0.01.

To verify the order of the study conditions were randomized, a Sign Test (2-tailed 1-Sample Binomial Test with a probability of 0.5) was performed (Table 1). Out of the 47 total participants, 29 completed the Speech-Only condition first, while 18 completed the Speech-And-Nonspeech-Sounds condition first. Out of the 39 sighted participants, 24 completed the Speech-Only condition first, while 15 completed the Speech-And-Nonspeech-Sounds condition first. Out of the 8 participants who are blind or visually impaired, 5 completed the Speech-Only condition first, while 3 completed the Speech-And-Nonspeech-Sounds condition first. Performing a Sign Test determined the order of the study conditions were randomized with a p-value of 0.14, 0.72, and 0.19, respectively.

Table 1. Sign test to determine if condition order was randomized.

Sign test for order ($\alpha = 0.01$)	
Sighted	$p = 0.1996$
Blind or visually impaired	$p = 0.7266$
Total	$p = 0.1439$

3.1 Preference

Of the 47 total participants, 24 preferred the Speech-Only audio style, 20 preferred the Speech-And-Nonspeech-Sounds audio style, and 3 had no preference. Of the 39 sighted participants, 20 preferred the Speech-Only audio style, 16 preferred the Speech-And-Nonspeech-Sounds audio style, and 3 had no preference. Of the 8 participants who are blind or visually impaired, 4 preferred the Speech-Only audio style, 4 preferred the Speech-And-Nonspeech-Sounds audio style, and 0 had no preference (Table 2).

Performing a Sign Test determined a significant difference for having no preference, but no significant difference in preference for Speech-Only or Speech-And-Nonspeech-Sounds audio (Table 3). Since the p-value for no preference was extremely low (<0.0000001) for the total and sighted participants, the choice of No Preference was considered to be an outlier, making the appropriate probability of the Binomial Test to be 0.5 rather than 0.33. Although the p-value for no preference was significant but not extremely low for the participants who were blind or visually impaired, the choice of No Preference may still be considered an outlier due to the small sample size and the tendency of this study's results for this group of participants to reflect the same as those for sighted participants.

Table 2. Participant preference for audio feedback style.

Audio feedback preference			
Participants	Speech-Only	Speech-And-Nonspeech-Sounds	No preference
Sighted	20	16	3
Blind or visually impaired	4	4	0
Total	24	20	3

Table 3. Sign test for audio feedback style preference.

Sign test for audio feedback style preference ($\alpha = 0.01$)			
Participants	Speech-Only	Speech-And-Nonspeech-Sounds	No preference
Sighted	$p \approx 1$	$p = 0.3368$	$p < 0.0001$
Blind or visually impaired	$p = 1$	$p = 1$	$p = 0.0078$
Total	$p \approx 1$	$p = 0.3817$	$p < 0.0001$

3.2 Intuitiveness

The intuitiveness of each audio feedback style was ranked on a scale of 1–5, with 1 indicating very unintuitive and 5 indicating very intuitive. Of the total participants, the average rank for the Speech-Only audio style was 3.06 with a median of 3 and a variance of 1.23. The average rank for the Speech-And-Nonspeech-Sounds audio style was 2.87 with a median of 3 and a variance of 1.72 (Table 4). An F-test determined no significant difference between the variances with a p-value of 0.26. A t-test determined no significant difference between the two rankings with a p-value of 0.31 (Table 5).

Of the sighted participants, the average rank for the Speech-Only audio style was 3 with a median of 3 and a variance of 1.42. The average rank for the Speech-And-Nonspeech-Sounds audio style was 2.84 with a median of 3 and a variance of 1.65 (Table 4). An F-test determined no significant difference between the variances with a p-value of 0.63. A t-test determined no significant difference between the two rankings with a p-value of 0.42 (Table 5).

Of the participants who are blind or visually impaired, the average rank for the Speech-Only audio style was 3.37 with a median of 3 and a variance of 0.26. The average rank for the Speech-And-Nonspeech-Sounds audio style was 3 with a median of 3.5 and a variance of 2.28 (Table 4). An F-test determined a potentially marginal significance between the variances with a p-value of 0.011. However, this marginal significance may be due to the small sample size of 8 and can be considered as not significant when considering the tendency of similar results between both groups of participants. A t-test determined no significant difference between the two rankings with a p-value of 0.58 (Table 5).

Table 4. Average, median, and variance of audio feedback style intuitiveness rankings.

Intuitiveness ranking of audio feedback styles						
	Speech-Only			Speech-And-Nonspeech-Sounds		
Participants	Average	Median	Variance	Average	Median	Variance
Sighted	3	3	1.4210	2.8461	3	1.6599
Blind or visually impaired	3.375	3	0.2678	3	3.5	2.2857
Total	3.0638	3	1.2349	2.8723	3	1.7224

Table 5. F-test and t-test results comparing audio feedback style intuitiveness rankings.

Comparison of intuitiveness rankings ($\alpha = 0.01$)		
Participants	F-test	t-test
Sighted	$p = 0.6344$	$p = 0.4213$
Blind or visually impaired	$p = 0.0112$	$p = 0.5837$
Total	$p = 0.2628$	$p = 0.3164$

3.3 Accuracy

The total average accuracy of the Speech-Only audio style condition was 78.01% with a standard deviation of 25.64, and that of the Speech-And-Nonspeech-Sounds audio style condition was 74.23% with a standard deviation of 26.12. Of the sighted participants, the average accuracy of the Speech-Only condition was 77.20% with a standard deviation of 27.44, and that of the Speech-And-Nonspeech-Sounds condition was 73.50% with a standard deviation of 28.22. Of the participants who are blind or visually impaired, the average accuracy of the Speech-Only condition was 81.94% with a standard deviation of 14.47, and that of the Speech-And-Nonspeech-Sounds condition was 77.77% with a standard deviation of 11.87 (Table 6).

Performing F-tests and t-tests determined no significant difference between the variance and accuracy of each condition in total and based on participant group (Table 7).

Table 6. Average participant accuracy and standard deviation by audio feedback style.

Average accuracy statistics by audio feedback style				
Participants	Speech-Only		Speech-And-Nonspeech-Sounds	
	Accuracy	Stdev.	Accuracy	Stdev.
Sighted	77.20%	27.44	73.50%	28.22
Blind or visually impaired	81.94%	14.47	77.77%	11.87
Total	78.01%	25.64	74.23%	26.12

Table 7. F-test and t-test results comparing accuracy between audio feedback style conditions.

Comparison of condition accuracy ($\alpha = 0.01$)		
Participants	F-test	T-test
Sighted	$p = 0.8640$	$p = 0.2382$
Blind or visually impaired	$p = 0.6813$	$p = 0.1970$
Total	$p = 0.9006$	$p = 0.1528$

3.4 Relative Pitch vs. Nonspeech Sound Interpretation Technique vs. Feedback Preference

Participant relative pitch ability was determined by their responses in the post-experiment survey. Participants considered to have weaker relative pitch provided responses indicating they had difficulty distinguishing the nonspeech sounds, needed to concentrate more on the nonspeech sounds, and/or found the nonspeech sounds confusing or distracting. Participants considered to have stronger relative pitch provided responses indicating they found the nonspeech sounds instinctively "clicked" in their minds, acted as an easily identifiable label, and/or required less concentration.

30 of the 39 sighted participants provided sufficient responses to determine their relative pitch ability. Of the total participants who provided sufficient responses, 18 had weaker relative pitch, and 20 had stronger relative pitch. Of the sighted participants, 14 had weaker relative pitch, and 16 had stronger relative pitch. Of the participants who are blind or visually impaired, 4 had weaker relative pitch, and 4 had stronger relative pitch (Table 8).

A Sign Test was performed to determine if participant relative pitch ability was disproportionate. The Sign Test found the number of participants with weaker and stronger relative pitch were equivalent in total and for both participant groups (Table 9).

Table 8. Participant relative pitch ability.

Participant relative pitch ability		
Participants	Weaker relative pitch	Stronger relative pitch
Sighted	14	16
Blind or visually impaired	4	4
Total	18	20

Table 9. Sign test to determine if participant relative pitch ability was disproportionate.

Sign test for relative pitch ability proportions ($\alpha = 0.01$)	
Sighted	$p = 0.8555$
Blind or visually impaired	$p = 1$
Total	$p = 0.8714$

Participants uniquely utilized a combination of techniques to interpret each audio feedback style. Some commonly used techniques included visualization, memorization, repeated listening of audio clips to focus on different information, and associating aspects of the audio with a body part and/or movement. In particular, the "translation" technique was unique only to the interpretation of nonspeech sounds. The "translation" technique was used when a user translated nonspeech sounds into language, while its counterpart, the "no translation" technique, was used when a user directly comprehended the nonspeech sounds. Participant usage of the "translation" or "no translation" technique was determined by their responses in the post-experiment survey. All the participants utilizing the "translation" technique had weaker relative pitch, while all the participants utilizing the "no translation" technique had stronger relative pitch (Table 10).

Similarly, all the participants utilizing the "translation" technique also preferred the Speech-Only audio style. Most of the participants utilizing the "no translation" technique preferred the Speech-And-Nonspeech-Sounds audio style, with only 2 having no preference (Table 11). Since having no preference was considered an outlier, the results could be interpreted as essentially all participants utilizing the "no translation" technique preferred the Speech-And-Nonspeech-Sounds audio style.

Table 10. Comparison of participant relative pitch ability to nonspeech sound interpretation technique.

Relative pitch vs. Nonspeech sound interpretation technique			
Participants	Relative pitch	Translation	No translation
Sighted	Stronger	0	16
	Weaker	14	0
Blind or visually impaired	Stronger	0	4
	Weaker	4	0
Total	Stronger	0	20
	Weaker	18	0

Table 11. Comparison of participant nonspeech sound interpretation technique to audio feedback style preference.

Nonspeech interpretation technique vs. Audio feedback style preference				
Participants	Technique	Speech-Only	Speech-And-Nonspeech-Sounds	No preference
Sighted	Translation	14	0	0
	No translation	0	14	2
Blind or visually impaired	Translation	4	0	0
	No translation	0	4	0
Total	Translation	18	0	0
	No translation	0	18	2

4 Conclusion

The study found no significant difference between overall preference for, intuitiveness when using, or accuracy in identifying patterns through either audio feedback method. Both audio feedback methods allowed users to perform equally well with high accuracy, suggesting that both feedback methods are viable avenues for information presentation. The study found “translation” and “no translation” to be generic techniques participants often used when listening to nonspeech audio feedback. The following relationships were found between performance, feedback style, user techniques, participant preference of feedback style, and strength of participant relative pitch:

1. When interacting with the Speech-And-Nonspeech-Sounds audio style, individuals with strong relative pitch utilized the “no translation” technique and found this style more intuitive than the other due to reduced cognitive workload, preferring this style to the other.
2. When interacting with the Speech-And-Nonspeech-Sounds audio style, individuals with weak relative pitch utilized a “translation” technique and found this style less intuitive than the other due to increased cognitive workload, preferring the alternative style.
3. Participants grouped by relative pitch strength utilized the same generic technique, but uniquely employed specific techniques.
4. A correlation was found between equivalently high design performance and lack of overall design preference.

These relationships support the results of prior research [3–5] and imply that:

1. Certain design constraints encourage the usage of certain generic techniques in individuals based on their ability.
2. Users will uniquely adapt to the constraints of their context to find their own ideal strategy.
3. Technology should not be developed with design constraints that allow for only one optimal generic technique. Varying user ability can explain the correlation between equivalently high design performance and lack of overall design preference. Users will perform better when interacting with a design constraint that is conducive to an optimal generic technique that they are able to harness. As such, users will prefer certain design constraints based on their ability. Significant proportions of users have varying ability, which will counterbalance any differences in design performance and preference.

Acknowledgments. The authors would like to thank the National Science Foundation and Arizona State University for their funding support. This material is partially based upon work supported by the NSF under Grant No. 1828010.

References

1. U.S. Government Printing Office: Individuals with disabilities education improvement act of 2004: conference report (to accompany H.R. 1350). U.S. Government Printing Office, Washington, D.C. (2004)

2. Statistics January 2019. https://www.nfb.org/resources/blindness-statistics. Accessed 6 February 2020
3. Ramloll, R., Brewster, S., Yu, W., Riedel, B.: Using non-speech sounds to improve access to 2D tabular numerical information for visually impaired users. In: Blandford, A., Vanderdonckt, J., Gray, P. (eds.) People and Computers XV—Interaction Without Frontiers, pp. 515–529. Springer, London (2001). https://doi.org/10.1007/978-1-4471-0353-0_32
4. Goncu, C., Marriott, K.: GraVVITAS: generic multi-touch presentation of accessible graphics. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011. LNCS, vol. 6946, pp. 30–48. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23774-4_5
5. Toennies, J.L., Burgner, J., Withrow, T.J., Webster, R.J.: Toward haptic/aural touchscreen display of graphical mathematics for the education of blind students. In: 2011 IEEE World Haptics Conference, pp. 373–378. IEEE (2011). https://doi.org/10.1109/whc.2011.5945515