# CORRELATIONS AND SCATTERPLOTS : A COMPARISON OF AUDITORY AND VISUAL MODES OF LEARNING AND TESTING

*Michael A. Nees*

Lafayette College
Department of Psychology
Oechsle Hall, Easton, PA, 18042
**neesm@lafayette.edu**

## ABSTRACT

An experiment examined brief, computer-based modules for teaching and conducting achievement testing of introductory concepts related to correlations and scatterplots. Participants experienced either auditory or visual learning modules followed by either auditory or visual tests of the concepts presented in the modules. Visual modules and tests used on-screen text and visual scatterplots, whereas auditory learning modules and tests presented the same content with text-to-speech (TTS) presentations and auditory versions of scatterplots. Across learning and testing manipulations, no differences were found in the accuracy of responses on the tests, but both auditory learning and auditory testing resulted in longer response times. Results are discussed in the context of computer-based learning and auditory learning and testing as an accommodation.

## 1. INTRODUCTION

The auditory presentation of text via digital TTS has become more practical to implement in an array of devices, and interfaces that use TTS (e.g., the Siri digital assitant on Apple's iPhone) seem to be growing in popularity. The use of TTS for educational and learning activities also has become more feasible in computers and a multitude of digital devices. In a recent survey, over one third of e-reader users and over half of e-book users rated TTS functionality as "valuable" or "very valuable" [1], and at least some online TTS services (e.g., www.ispeech.org) explicitly tout the value of TTS for translating educational materials to the auditory modality.

TTS technologies can use sound to display text to visually-impaired learners and also to sighted learners with alternative learning preferences. Though computer-based and digital learning technologies have become ubiquitous, research to date largely has not addressed fundamental best-practice questions surrounding the implementation of these technologies for teaching and learning [2]. The redundant presentation of auditory and visual learning materials can be beneficial [3], but the equivalence of auditory-only versus visual-only presentations of materials for pedagogical purposes remains unclear. Similarly, very little research exists on the design of auditory-only tests.

The lack of research comparing the efficacy of visual and auditory presentations of learning and testing materials is problematic for several reasons. First, there seems to be a pervasive assumption that the auditory delivery of text—even very complex text associated with many learning activities—provides learning opportunities that are equivalent to the opportunities offered when the same text is presented visually. This assumption may be flawed when it is applied to the delivery of complex curricula in science and math education, as the demands placed on working memory by the transient nature of the auditory presentation may present memory difficulties (i.e., extraneous cognitive load, see [4]) not encountered by visual learners of the same text.

In addition to the delivery of curricula through sound, a common accommodation for test-takers with visual impairments (or other disabilities) has been the auditory presentation of test questions [5]. Despite the prevalence of this accommodation in both aptitude and achievement testing, researchers have yet to establish the validity of tests administered under oral accommodations. In current practice, often a human reader will administer oral examinations. This presents obvious complications for standardization of testing conditions, and researchers [6] have suggested that a better approach may be to develop "self-voicing" TTS systems to administer oral versions of tests [7]. The comparability of auditory and visual modes of information presentation should be established to ensure that equitable delivery of curricula and fairness in testing can be accomplished in both modalities.

Another known gap exists in the translation of graphical materials in visual texts into auditory representations [8] for both learning and testing. Geisinger [9] pointed out, "…the use of figures and graphs make tests more difficult and typically may alter the cognitive processes employed—because they must be described verbally to the test taker with visual impairment" (pp. 131). Auditory graphs offer a promising alternative to verbal descriptions for translating graphs into sound, as emergent percepts of data patterns may function similarly in auditory and visual graphs [10]. Research [11] has shown that auditory versions of scatterplots are as effective as visual representations for conveying correlations.

The current study examined auditory and visual learning and testing of introductory statistical concepts about correlation and scatterplots in a sample of university students with no prior formal education in statistics. The use of TTS and auditory versions of scatterplots was compared to visual presentation of text and graphical scatterplots with a 2 (learning module: visual or auditory) X 2 (test format: visual or auditory) X 2 (question type: scatterplot or no scatterplot) mixed design. The study was designed to examine: 1) the efficacy of both auditory and visual learning; 2) the comparability of auditory and visual testing; 3) the possibility of interactions between modes of learning and testing; and 4) the possibility of differential effects for test questions that required or did not require judgments about (auditory or visual) scatterplots.

## 2. METHOD

### 2.1. Participants

Participants ($N$ = 41; 20 females; $M$ age = 20.0, $SD$ = 1.9 years) were recruited from undergraduate psychology courses at the Georgia Institute of Technology. Participants were excluded if they had taken a statistics course at the high school or college level, and all participants reported normal or corrected-to-normal hearing and vision.

### 2.2. Stimuli

A brief (approximately 3000 word) script of a lesson on correlations and scatterplots was prepared. The lesson covered basic concepts such as the direction and strength of correlations, interpreting $r$ values, and reading bivariate scatterplots of data. This lesson was the basis of the respective auditory and visual learning modules, both of which were approximately 20 min in duration.

For the visual module, the text was presented on the computer screen in complete sentences (from one up to several sentences at once) at a pace that was controlled by the computer. The duration of the text presentations was yoked to the duration of the corresponding TTS audio file from the auditory module (described below), thus the learning modules were exactly matched in duration. Visual examples of scatterplots used in the module were made using Microsoft Excel and were displayed alone on the screen for 5 s. To ensure that the information contained in the visual scatterplots was commensurate with that of the auditory scatterplots, all visual scatterplots were stripped of axis labels; only data points showing the relationship between the two variables depicted, which were described in the text, were displayed.

For the auditory learning module, TTS conversions of the text of the visual module were made with the demo function of the TTS engine at http://www.ispeech .org in early 2011[1]. TTS was created using the "English male" voice (now "US English male") at the "normal" (i.e., default) speed setting. The text of the visual module was converted to mp3 files from the website. Exact text from the visual modules was entered into the TTS engine with two exceptions: 1) where appropriate, the text was modified to reflect the auditory nature of the module (e.g., "scatterplot" was changed to "auditory scatterplot"); and 2) numbers and symbols were entered into the TTS engine as words to ensure that the auditory speech was intelligible for all text elements from the visual module (e.g., "$r$ =" was voiced as "*are equals*"). The data from each scatterplot in the visual module were sonified into auditory graphs using the Sonification Sandbox [12] software. All scatterplots were sonified to be 5 s in duration in the range of notes C4 (MIDI note 60, 262.6 Hz) to C8 (MIDI note 108, 4186.0 Hz) using a positive polarity mapping and the MIDI piano timbre.

The visual test consisted of 20 multiple-choice questions. Test questions were comparable to the types of practice and test questions on correlations found in introductory statistics texts. Each question was displayed on the screen in its entirety with each of the four possible answers visible. The auditory test presented the exact same questions and answers with TTS. Each question was read in its entirety, followed by each of the four possible answers in succession. The test was designed such that half of the questions were conceptual in nature and did not display a scatterplot representation of the data, while half of the questions displayed one or more scatterplots as part of the question or answers. At the beginning of each test, participants were given a brief (one paragraph) overview of either auditory or visual scatterplots (depending on the test format condition). The overview was necessary to explain the respective representations to participants who had experienced the learning module in a different modality from the test format (e.g., participants who experienced the visual learning model needed a brief description of how the auditory scatterplots represented data).

### 2.3. Procedure

Following informed consent, participants were randomly assigned to one of the four factorial combinations of the 2 (learning: auditory versus visual) x 2 (test: auditory versus visual) between-subjects independent variables. Participants were seated at a computer in front of a 17 in (43.2 cm) Dell LCD computer monitor. A computer program made with Adobe Director presented stimuli and collected data. Auditory stimuli were presented with Sennheiser HD 202 headphones. All participants wore headphones during the study, though no sounds were presented to participants assigned to visual learning conditions and visual testing conditions. Similarly, the computer screen was blank during auditory conditions of the study. Participants experienced either the auditory or visual learning module, followed by either the auditory or visual test. The 20 test questions were presented in a random order for each participant, and both responses and response times were recorded. The response time for a trial was operationally defined as the duration between the onset of the question (i.e., the appearance of the question on the screen for the visual test or the beginning of the TTS audio reading of the question for the auditory test) and the logging of a response to the test question. Participants in either condition could log a response at any time; participants in the auditory test condition were not obligated to listen to the entire question and set of answers. Following the test, participants completed the NASA-TLX [13] measure of subjective workload.

## 3. RESULTS

Analyses were conducted using mixed 2 (learning module: visual or auditory) X 2 (test format: visual or auditory) X 2 (question type: scatterplot or no scatterplot) ANOVAs on both the number of correct answers on the test and the response times to test questions. For the number of correct answers, the main effects of learning module, $F(1,37) = 0.37$, $p = .55$, and test format, $F(1,37) = 3.26$, $p = .08$, were not statistically

[1] At the time of this submission, the website appeared to have made minor modifications to the interface and TTS algorithms since the stimuli were created.

significant. The effect of question type was significant, $F(1,37) = 13.24$, $p = .001$, $\eta^2_P = .26$; participants correctly answered more questions without scatterplots ($M = 6.37$, $SE = .32$) than with scatterplots ($M = 5.28$, $SE = .24$). There were no statistically significant interactions ($p$ values ranged from .14 to 89). Of note, the test exhibited neither ceiling nor floor effects; chance performance would have resulted in $M = 2.5$ correct answers in each condition. Results are shown in Figure 1.

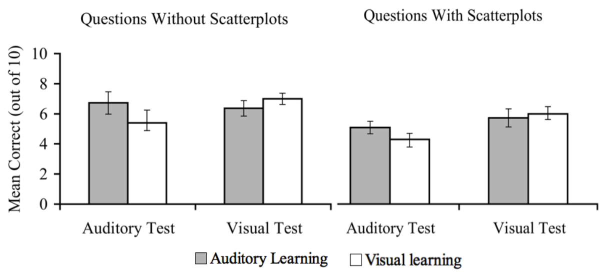


Figure 1: Mean numbers of correct answers across conditions. Error bars represent standard error.

For the response times (reported in s) to test questions, all main effects were significant. For the learning module independent variable, $F(1,37) = 5.23$, $p = .03$, $\eta^2_P = .12$, participants' mean response times were significantly faster if they had learned the material from the visual module ($M = 30.3$, $SE = 1.4$) as compared to the auditory module ($M = 25.8$, $SE = 1.5$). For the test format independent variable, $F(1,37) = 54.37$, $p < .001$, $\eta^2_P = .60$, participants' mean response times were significantly faster if they took the visual version of the test ($M = 20.7$, $SE = 1.4$) as compared to the auditory version ($M = 35.3$, $SE = 1.4$). For the question type independent variable, $F(1,37) = 13.00$, $p < .001$, $\eta^2_P = .26$, participants' mean response times were significantly faster for questions without scatterplots ($M = 26.5$, $SE = 1.1$) as compared to questions with scatterplots ($M = 29.6$, $SE = 1.1$). The interaction of test format with question type was also significant, $F(1,37) = 30.14$, $p < .001$, $\eta^2_P = .45$. The interaction was reflected in the fact that participants taking the auditory version of the test were slower to provide a response to questions with scatterplots ($M = 31.4$, $SE = 1.5$) as compared to questions without scatterplots ($M = 39.3$, $SE = 1.5$), but participants taking the visual version of the test did not show a difference for questions without scatterplots ($M = 21.6$, $SE = 1.5$) as compared to questions with scatterplots ($M = 19.9$, $SE = 1.6$). Results are shown in Figure 2.

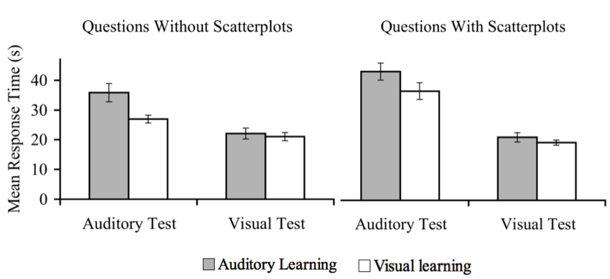


Figure 2: Mean response times across conditions. Error bars represent standard error.

One obvious interpretation of the disparities in response times is that participants in the visual condition simply read the questions and answers faster than the TTS presented the questions and answers in the auditory version of the test. This interpretation would be supported if the difference between mean auditory and visual test response times increased as the duration of the auditory questions and answers increased. To examine this possibility, an exploratory correlation showed that, across the 20 different questions, the duration of the audio version of the question and answers (i.e., the time required for participants to hear the question and all four answers in the auditory test condition) and the difference in mean response times for the visual versus auditory test conditions were not related, $r(19) = .38$, $p = .10$. Though this relationship (see Figure 3) might have reached statistical significance with a larger sample of questions, the pattern of results showed that the duration of the auditory test questions alone did not account for tendency of auditory test-takers to require a longer response time.

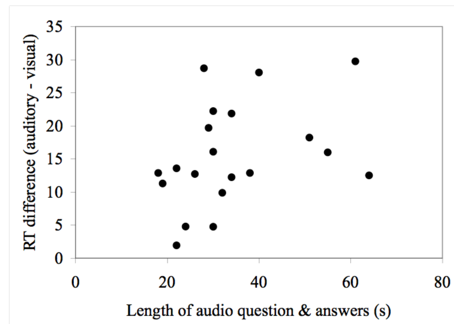


Figure 3: The length of audio questions and answers as a function of the mean response time difference (visual format subtracted from auditory format) for each of the test questions.

Finally, a 2 (learning module: visual or auditory) X 2 (test format: visual or auditory) ANOVA was performed on the NASA-TLX composite scores. The main effects of learning module, $F(1,37) = 0.12$, $p = .73$, and the interaction of learning module with test format, $F(1,37) = 0.89$, $p = .35$, were not statistically significant. The main effect of test format was significant, $F(1,37) = 8.04$, $p = .007$, $\eta^2_P = .18$. Participants in the auditory test condition experience greater perceived workload ($M = 10.73$, $SE = 0.64$) than participants in the visual test condition ($M = 8.14$, $SE = 0.65$). Results are shown in Figure 4.

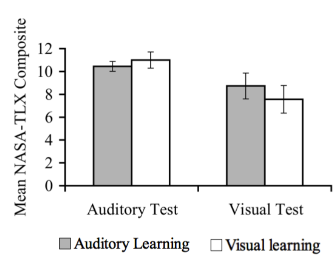


Figure 4: Mean NASA-TLX composite scores. Error bars represent standard error.

## 4. DISCUSSION

Participants with no prior educational background in statistics learned simple concepts about bivariate correlations equally well—as measured by correctness of responses to test questions—from visual modules (that used text with traditional visual scatterplots) and auditory modules (that used TTS with auditory scatterplots). Auditory learning modules and auditory versions of achievement tests may represent a viable alternative to visual presentation of materials during learning and testing. Accommodated versions of standardized tests and achievement tests that use oral presentation of questions to date have not found an adequate solution to translating test questions involving graphs and diagrams to the auditory modality. Verbal descriptions of visual figures may be insufficient to offer comparable information, but auditory representations of graphical data may help to fill this considerable gap.

Participants who learned the material with the auditory module took about 5 s longer on average per question on the test than participants who had learned the material with the visual modules, and this effect was present regardless of (i.e., collapsed across) the format of the test. Participants taking auditory versions of the test took about 15 s longer per question on average to register a response to the test question, and responses were even slower with auditory testing for questions that featured an auditory scatterplot as part of the question (as opposed to conceptual questions that featured only spoken words with no scatterplot). The data suggested that longer response times for the auditory version of the test were not simply attributable to the durations of the auditory test questions.

Learners who are assessed with auditory tests may need to be given longer to complete the test. Extended time during testing is another common accommodation that is often implemented in conjunction with auditory presentation of questions. Often, the amount of extra time given seems to be arbitrarily chosen as “time and a half” or “double time.” Studies like this one may be able to offer empirical guidelines for the amount of extra time needed to achieve comparable mean performance across testing formats. The significant difference in perceived workload did not correspond to an objective decrease in test performance as measured by the number of correct answers, but the perceived workload could potentially have detrimental effects on test performance in assessment scenarios that run longer than the brief test here.

## 5. CONCLUSIONS

TTS auditory versions of learning materials with auditory graphs may offer a comparable alternative to traditional visual learning materials for teaching basic statistics concepts. Perhaps even more importantly, TTS versions of tests with auditory graphs may offer a standardized means of assessing achievement of basic statistics (and perhaps other math) concepts in the auditory modality that is comparable to the visual tests currently used in learning assessment, though the current study’s results suggest that TTS test-takers may require more time to complete assessments. The finding that auditory test-takers perceived higher subjective workload warrants further investigation.

## 6. ACKNOWLEDGMENT

This research was conducted at the Georgia Tech Sonification Lab with support from Bruce Walker. Zia Drakhshandeh created some stimuli for this experiment, and Michelle Han collected data. Their contributions are gratefully acknowledged.

## 7. REFERENCES

[1] N. Foasberg, "Adoption of e-book readers among college students: A survey," *Information Technology and Libraries,* vol. September, pp. 108-128, 2011.

[2] D. A. Cook, "The research we still are not doing: An agenda for the study of computer-based learning," *Academic Medicine,* vol. 80, pp. 541-548, 2005.

[3] R. Moreno and R. E. Mayer, "Verbal redundancy in multimedia learning: When reading helps listening," *Journal of Educational Psychology,* vol. 94, pp. 156-163, 2002.

[4] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," *Learning and Instruction,* vol. 4, pp. 295-312, 1994.

[5] S. G. Sireci, S. E. Scarpati, and S. Li, "Test accommodations for students with disabilities: An analysis of the interaction hypothesis.," *Review of Educational Research,* pp. 457-490, 2005.

[6] E. G. Hansen, M. J. Lee, and D. C. Forer, "A "self-voicing" test for individuals with visual impairment," *Journal of Visual Impairment & Blindness,* pp. 273-275, 2002.

[7] R. P. Dolan, T. E. Hall, M. Banerjee, E. Chun, and N. Strangman, "Applying principles of universal design to test delivery: The effect of computer-based read-aloud on test performance of high school students with learning disabilities.," *Journal of Technology, Learning, and Assessment,* vol. 3, 2005.

[8] M. A. Nees and B. N. Walker, "Data density and trend reversals in auditory graphs: Effects on point estimation and trend identification tasks," *ACM Transactions on Applied Perception,* vol. 5, Article 13, 2008.

[9] K. F. Geisinger, "Psychometric issues in testing students with disabilities," *Applied Measurement in Education,* vol. 7, pp. 121-140, 1994.

[10] M. A. Nees and B. N. Walker, "Listener, task, and auditory graph: Toward a conceptual model of auditory graph comprehension," in *International Conference on Auditory Display (ICAD2007)*, Montreal, Canada, 2007, pp. 266-273.

[11] J. H. Flowers, D. C. Buhman, and K. D. Turnage, "Cross-modal equivalence of visual and auditory scatterplots for exploring bivariate data samples.," *Human Factors,* vol. 39, pp. 341-351, 1997.

[12] B. K. Davison and B. N. Walker, "Sonification Sandbox reconstruction: Software standard for auditory graphs," in *ICAD 07 - Thirteenth Annual Conference on Auditory Display*, Montreal, Canada (26-29 June), 2007, p. TBD.

[13] S. G. Hart and L. E. Staveland, "Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research," in *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds. Amsterdam: North Holland Press, 1988, pp. 239-250.