Auditory Data Representation

Michelle R Carney, MIMS 18

Joyce S Lee, MIMS 19

With Advisor: Marti Hearst

UC Berkeley, School of Information

Masters of Information Management and Systems

Berkeley Center for New Media Certificate

May 2018

Acknowledgements

We would like to thank the Berkeley Center for New Media for helping fund this project, and our advisor, Professor Marti Hearst, for her guidance, feedback, patience, and support as we explored this new domain. Michelle would especially like to thank Joyce Lee for going above and beyond and being my collaborator on this project, and her unwavering enthusiasm for this project. We would also like to thank Peter Rowland for his help in creating the data sonification stimuli for the user testing, and Steve Fadden for his feedback and advice on our VUI user research. Thank you to everyone who have helped to make this project possible.

Overview

With the rise of both virtual assistants and software-embedded devices, audio-first interactions are becoming more prevalent in daily life. However, there is not yet an industry standard for designing sound experiences to communicate data generated by ubiquitous computing, particularly through emergent conversational interfaces. We imagine a future in which we will be able to explore data with a combination of the ear and a voice user interface: how might Alexa enable us to explore complex data? How might we hear the weather, stock market prices, or the status of our IoT-connected home – by sound alone?

We have conducted an in-depth literature review of prior work related to data sonification, shaping our analysis based on visual analogs of “auditory graphs” (i.e., histograms, scatterplots, pie charts, etc). Spanning across various disciplines – including literature from human-computer interaction (HCI), accessibility, music and art – these papers demonstrated a great range of approaches, including expressive, subjective design proposals as well as more functional, objective representations, vetted by experimental procedures.

A major difficulty encountered while reviewing the literature, however, is that much of it is dated and lacks access to the audio files associated with each of the papers. Therefore, we reproduced three different sonification methods for our own voice user interface (VUI), representing data from the 2010 Census and the 2015 American Community Survey. With this VUI, we conducted five in-person usability test sessions to evaluate the potential of auditory data exploration via a contemporary, conversational interface, as well as to present recommendations for future work.

This is an exciting and fun new project that combines learnings from tangible user interface design, natural language processing, and information visualization courses at UC Berkeley’s School of Information, with sponsorship from the Berkeley Center for New Media. While primarily aimed at smart devices, this project can have implications for developing accessible design patterns going forward to allow an alternative way to interact with data for everyone.

Table of Contents

Introduction

As of January 2018, one in six Americans own a smart speaker – and this population has more than doubled in the course of one year.[1] With this rise of virtual assistants and software-embedded devices, audio-first interactions are becoming more prevalent in daily life. However, there is not yet an industry standard on designing sound experiences that adequately communicate the datasets generated by ubiquitous computing, especially through conversational or voice user interfaces (VUIs). With the rise of these audio first, or even audio-only, VUIs – in tandem with the rise of data quantifying everything around us – there is a need to design best practices for the representation of data in auditory form.

Though it is increasingly timely, this is not a new problem. The accessibility community has long advocated to set standards for making websites accessible by screen reader: one such initiative, WebAIM,[2] was established in 1999. In the past few years, R[3] and Highcharts[4] have also developed projects to make their software and graphs more accessible. Furthermore, there have been a number of creative efforts to explore data sonification as art, including the work of Roger Molina[5] and Brian Foo.[6] This type of dedication to audio exploration remains relatively niche, even if sound is a valid – even rich – data source. Only certain professions require the training to extract and interpret meaning from audio: from piano tuners using the pitch to tune instruments, to physicians using a stethoscope to understand their patients' overall health. Hermann and Ritter noted early on that humans are "capable to detect very subtle patterns in acoustic sounds:"[7] so we

---

1 "The Smart Audio Report from NPR and Edison Research." National Public Media, www.nationalpublicmedia.com/smart-audio-report/
2 https://webaim.org/
3 https://zorkow.github.io/BrailleR/www/
4 https://www.highcharts.com/a11y.html
5 http://malina.diatrope.com/about/
6 http://brianfoo.com/
7 Hermann, Thomas, and Helge Ritter. "Listen to your data: Model-based sonification for data analysis." Advances in intelligent computing and multimedia systems (1999).

ask, how might we learn to hear structures or patterns in data, as we have learned to see them in visual representations?

Process

Our initial research question was fairly straightforward: *how has audio successfully been used to express data?* With prior exposure to work in the realms of accessibility and music, we began the project by focusing on understanding the space of auditory data representation in HCI literature. After searching for papers with keywords like “audio graphs,” “auditory data,” “musical graphs,” and “data sonification,” we found the works of prolific authors in this domain, including Bruce Walker[8] and John Flowers[9], as well as a conference expressly dedicated to this topic, the International Conference on Auditory Display (ICAD)[10]. After reading *The Sonification Handbook*[11] by various ICAD authors, and it was clear that much relevant research has been conducted, and that many experimental designs have been proposed on how to sonify data.

From our initial analysis, we noted that authors have tried to sonify different types of visual graphs. We used this as the basis of our subsequent research question during analysis, which was more granular and evaluative in nature: *what types of visual graphs have been sonified, and how well did they perform?* Excluding certain papers deemed adjacent to applying auditory data representation to a conversational interface, we selected the most relevant articles and mapped out the stimuli used, what experiments were attempted – if at all, and how well they worked.

A major difficulty encountered while reviewing the literature, however, is that much of it is decades old and lacks access to the audio files associated with each of the various papers. This is particularly problematic when audio stimuli were created with simulated data. As such, this

[8] http://sonify.psych.gatech.edu/~walkerb/index.html
[9] https://psychology.unl.edu/john-flowers
[10] http://icad.org/
[11] http://sonification.de/handbook/index.php/

motivated our last research question: *what is the possibility of auditory data exploration via a contemporary, conversational interface?* To investigate, we reproduced 3 different methods for data sonification into our own voice user interface, representing data from the 2010 Census and the 2015 American Community Survey. (See later section "Sonification of U.S. Census Data" for greater details.) After developing a VUI prototype, we conducted five in-person usability test sessions to evaluate this question and to present recommendations for future work.

Literature Review

We conducted an in-depth literature review of prior work on data sonification, shaping our analysis on based on visual analogs of the following “auditory graph” types: Histogram, Box Plot, Choropleth Map, Pie Chart, Scatter Plot, and Line Graph. For this review, we identified relevant papers in which the authors specified intention to sonify a type of graph, and to enable exploratory data analysis via audio. Not included among the selected publications in this analysis are those that include general auditory design principles,[12] [13] [14] examine audio signaling for alert purposes,[15] nor attempt to create tactile, accessible graphs that do not include an audio modality.[16]

The selected papers are presented below, grouped by the following representational characteristics: Distributions, Spatial Relationships, Clusters, and Trends. We summarize key findings from each representation here, and cover overarching themes in the “Analysis and Discussion” section that follows.

---

[12] Blattner, Meera M., Denise A. Sumikawa, and Robert M. Greenberg. "Earcons and icons: Their structure and common design principles." Human–Computer Interaction 4.1 (1989): 11-44.

[13] Kramer, Gregory; Walker, Bruce; Bonebright, Terri; Cook, Perry; Flowers, John H.; Miner, Nadine; and Neuhoff, John, "Sonification Report: Status of the Field and Research Agenda" (2010). Faculty Publications, Department of Psychology. 444.

[14] Hermann, Thomas, et al. The Sonification Handbook. Logos Verlag, 2011. http://sonification.de/handbook/index.php/

[15] Hearst, Marti A. "Dissonance on audio interfaces." IEEE Expert 12.5 (1997): 10-16.

[16] Wall, Steven, and Stephen Brewster. "Feeling what you hear: tactile feedback for navigation of audio graphs." Proceedings of the SIGCHI conference on Human Factors in computing systems. ACM, 2006.

Distributions: Histogram and Box Plot

Flowers and Hauer's early explorations evaluated the capability of auditory histograms and box plots to effectively communicate central tendency, variability, as well as shape (kurtosis and skewness).[17] As the authors explain, "These [audio histograms] consisted of strings of musical notes in which numeric values of observations (the midpoint of a class interval or bin) was represented by musical pitch, and the frequency of scores in each interval or bin was represented by the number of repetitions of a given note in the display." The box plots, by contrast, were simpler, ascending five-note sequences that mapped the extrema, quartiles, and median onto a musical scale. Participants were then asked to rate dissimilarity between pairs of 12 computer-generated data distributions, half of which were normal Gaussian distributions, while three were positively skewed and three were negatively skewed.

The authors found that despite conveying less information about distribution shape, auditory box plots were much "less effortful" for participants to perform the dissimilarity rating task. On the whole, however, the authors imply that both representations may be a useful alternative to traditional visual graphics, given that regression and multidimensional scaling analysis of the dissimilarity judgments suggest that auditory depiction provides "a highly effective means of conveying information about distributional characteristics." However, this experiment tested only among statistically knowledgeable subjects – advanced graduate students in experimental and social psychology – and with a relatively small sample (n=12). This representation may in fact be less effective for less numerically-skilled subjects, and is deserving of further study. Flowers and Hauer must have realized this as well, given the experimental design improvements in their

---

[17] Flowers, John H., and Terry A. Hauer. "The ear's versus the eye's potential to assess characteristics of numeric data: Are we too visuocentric?." Behavior Research Methods, Instruments, & Computers 24.2 (1992): 258-264.

subsequent paper,[18] which again contrasted the auditory histograms to auditory box plots, but evaluating participants that are both naive and informed, or familiar with statistics.

In this next paper, Flowers and Hauer focused on the auditory box plots, specifically explored (1) if these simplified depictions of distributions provide adequate information about distributional properties and (2) if combined auditory and visual presentation enhance sensitivity to differences in stimulus parameters. Over the course of three experiments, they found similar, if tempered, results as their prior study: the perceptual structure of the auditory box plots correspond well to the statistical properties it was intended to depict – even though it was a less familiar format to those without musical training, and perhaps more effortful to process than traditional visual graphics. Compared to the histogram, the box plot was again preferred for being succinct, but informative. The authors conclude that simpler is generally better when it comes to conveying basic distributional information during exploratory data analysis, but this is of particular importance to audio displays.

Spatial Relationships: Choropleth Map and Pie Chart

Our primary understanding of spatial relationships in auditory data representation came from Zhao et al's work to improve accessibility of geo-referenced statistical data to blind users.[19] In addition to an enhanced table, the authors created a spatial choropleth map (which we focus on in this report). Each state was assigned one of five categorical values, represented by a "value pitch," played for 200 milliseconds. For each state, this value pitch is followed by an "elevation pitch" lasting 100 milliseconds. (These piano pitches range from about one octave below middle C to about two octaves above middle C, and lower pitches indicate states further south.) The state

[18] Flowers, John H., and Terry A. Hauer. "“Sound” alternatives to visual graphics for exploratory data analysis." Behavior Research Methods, Instruments, & Computers 25.2 (1993): 242-249.
[19] Zhao, Haixia, et al. "Sonification of Geo-Referenced Data for Auditory Information Seeking: Design Principle and Pilot Study." ICAD. 2004.

pitches are played in order of 15 sweeping columns shown in Figure 1, from north to south then west to east. At the end of each column, a bell rings for 570 milliseconds, then is followed between a 100-millisecond pause. Three continuous bells indicating the end of the sequence, and the summary of all states, or “gist,” lasts about 25 seconds in total. Using the keyboard, participants could navigate via arrow keys and receive details – name and value of current state – on demand by pressing the spacebar key.

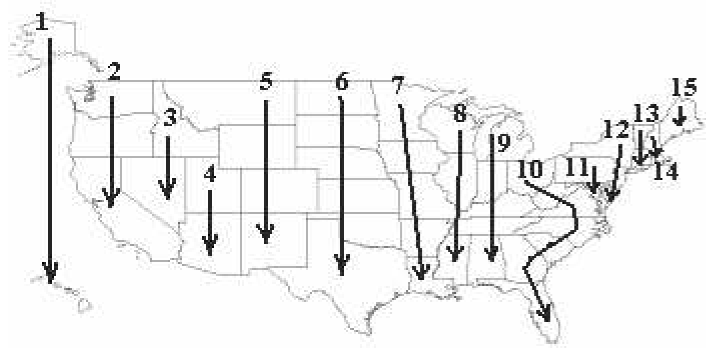

Figure 1. *Sweeping order in Zhao et al’s audio “gist,” with data played from north to south, columns to 15.*

While the authors concluded that the study participants successfully recognized geographic data distribution patterns in both the enhanced table and auditory map, the study could be improved with real – not simulated – data, as well as stronger sampling procedures: the pilot study featured a convenience sample of 9 sighted University of Maryland students and staff, of which 6 subjects had “excellent” knowledge of American geography, as demonstrated by pre-test. These findings may thus not be generalizable to non-sighted populations; furthermore, one must consider whether data distribution patterns were recognized due to participants’ existing familiarity with geography, rather than the auditory stimuli presented. The paper’s greatest strength is in the Auditory Information Seeking Principle (AISP) it offers for data sonification designs, parallel to

Scheiderman's visual information seeking mantra of "overview first, zoom and filter, then details-on-demand."[20] (See "Proposed Frameworks" for further discussion of AISP.)

Roberts and Franklin also explored auditory representations of spatial relationships, but with five different designs of auditory pie charts.[21] Despite using sophisticated wraparound speaker setups, they experienced poor results when trying to use spatial qualities to communicate sound – and inversely, when trying to communicate spatial qualities with sound. Their experiment also required special equipment, making such technology less accessible to broader audiences. However, their most effective design depended least on spatial qualities, instead drawing from Morse code, with long dashes for 10% and short dots for 1% (see Figure 2 for an example). While this representation enables communication of precise values, its communicative ability is limited to representations with fewer categories. Furthermore, it requires counting and mental math of the end user, making it less flexible and not ideal for data exploration.

.

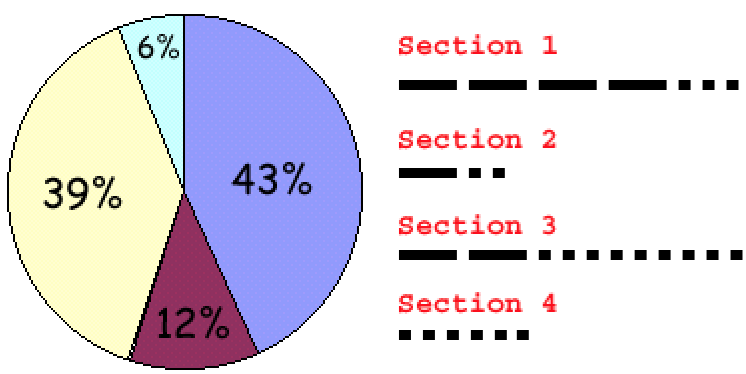

Figure 2. *Example of Franklin and Roberts's Morse code inspired pie chart sonification.*

[20] Ben Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations," Proceedings of the 1996 IEEE Symposium on Visual Languages.

[21] Franklin, Keith Michael, and Jonathan C. Roberts. "Pie chart sonification." Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on. IEEE, 2003.

Clusters: Scatter Plot

Scatter plots are a more popular form of data exploration that selected authors were interested in sonifying: Smith et al [22] point to early research by Bly, who initially tried to assess if participants could recognize translation, scaling, and correlation from just sound, sound and visual, or visual cue alone.[23] Among Bly's six-dimensional data samples, each of the six variables were mapped to properties of sound: pitch, duration, volume, waveshape, envelope, and overtone. By contrast, only two dimensions were mapped to the visual scatter plot display. After separating the data samples into two sets, test subjects were asked to discriminate which set the data was from when presented visually, aurally, and in a combination of both.

Trained subjects averaged 62% correct identification of samples with graphics alone, 64.5% correct with sound alone, and 69% correct with the combined graphics and sound presentation. Subjects who had participated in the sound-only trials were also given additional training months after the original experiment: in a subsequent experiment, these trained subjects went on to correctly identify 74% of the test samples." This is exciting because – as Bly outlines in her paper – shows the potential for sound to be used as a medium to help make sense of the data trends, both as an aid to vision and as its own medium.

[22] Smith, Stuart, R. Daniel Bergeron, and Georges G. Grinstein. "Stereophonic and surface sound generation for exploratory data analysis." Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 1990.

[23] Bly, Sara. "Presenting information in sound." Proceedings of the 1982 conference on Human factors in computing systems. ACM, 1982.

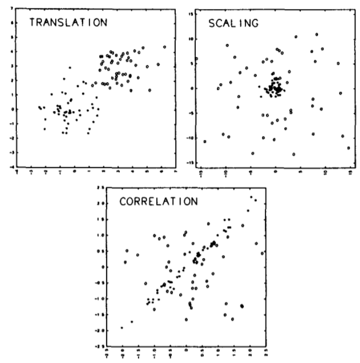

Figure 3: Phase 1 Data Sets

Figure 3. *Phase 1 data set from Bly (1982).*

Herrman and Ritter[24] described two design techniques for listening to data with the purpose of understanding patterns or trends: “(1) particle trajectories in a ‘data potential’ is a sonification model to reveal information about the clustering of vectorial data and (2) ‘data sonograms’ is a sonification for data from a classification problem to reveal information about the mixing of distinct class.” While applicable to Bly’s work, the “data sonograms” concept draws from Madhyastha and Reed’s “sound scatter plots,” [25] where they proposed designs based on the relative location of each of the data points.

Unlike the experimental approaches taken by most of the literature reviewed, Tomlinson et al.[26] conducted a unique, two-year longitudinal study: focusing on the education of math and graphing among non-sighted students, it employed auditory graphing software developed by the Georgia Institute of Technology Sonophication Lab, and it incorporated the perspective of both learners and educators. With a mixed methods approach, they used interviews to understand challenges faced by a middle school math teacher at a school for the blind with traditional auditory

---

[24] Hermann, Thomas, and Helge Ritter. "Listen to your data: Model-based sonification for data analysis." Advances in intelligent computing and multimedia systems (1999).

[25] Madhyastha, T.m., and D.a. Reed. “Data Sonification: Do You See What I Hear?” IEEE Software, vol. 12, no. 2, 1995, pp. 45–56.

[26] Tomlinson, Brianna J., et al. “Exploring Auditory Graphing Software in the Classroom.” ACM Transactions on Accessible Computing, vol. 9, no. 1, 2016, pp. 1–27

graphing software: this included pain points around lesson planning, execution, and reviewing student work. The researchers also evaluated individual feedback from students with a 27-question survey, including free-response and ranked-response questions. The authors found that students were able to learn the trends and tasks asked by the learning software, in addition to orienting themselves with the data via sonified scatter plots.

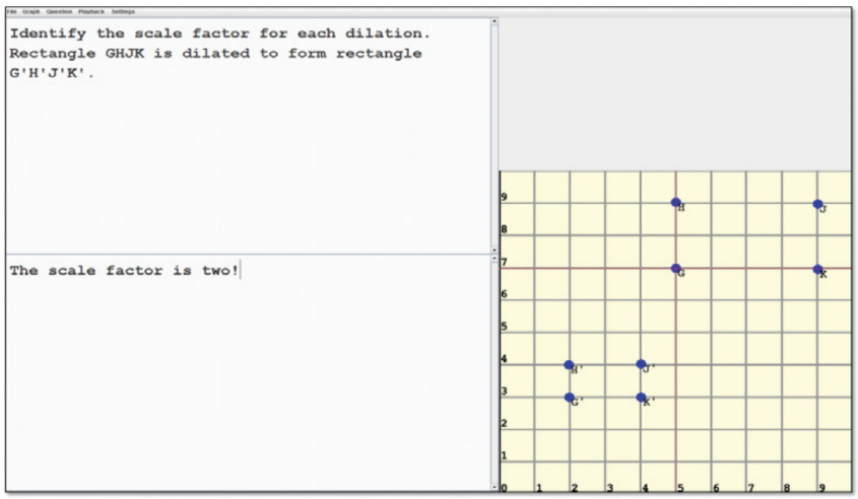

Figure 4. *A screenshot of the software used by Tomlinson et al. and the types of tasks asked.*

Trends: Line Graph

Among the auditory data representations considered, line graphs may be the most logical for some: each stream of data can be represented as a continuous note, like the line itself, with changes in pitch, volume, or any other mapping to mirror changes along the Y axes. This is especially common for time-series representations, such as weather temperature and stock market prices. For instance, Flowers et al (2002) sonified daily weather patterns from past records of temperature and precipitation in Lincoln, Nebraska based on “basic principles of auditory perception, attention and cognition."[27] While they did not test the design on participants, they did note several design choices including: (1) balancing the MIDI to center around the temperature range for the area of interest for artistic reasons, and (2) choosing a grand piano MIDI #1 as the timbre evokes the feeling of rainfall, versus the octave arpeggio on the dulcimer MIDI #15 to indicate "snow piling up."

Taking a more functional approach, Mansur et al.[28] explored using pitch to map two-dimensional line graphs with sighted and non-sighted subjects; participants were evaluated on their ability to understand concepts like the slope of a line, symmetry, and identification of straight lines versus exponentials, which was then compared to results from completing similar problems using tactile graph. The authors argue that these "sound graphs" were more flexible than the tactile graphs as they allowed "the ability to free the user's hands for other tasks...more importantly, they can be used effectively regardless of the individual's orientation or distance from the source of sound." Furthermore, they concluded that participants were able to easily interpret new sounds in a meaningful way quicker than a tactile representation.

Also working with both sighted and non-sighted participants, Walker and Lane at the Georgia Institute of Technology's Sonification Lab did some work focusing on the directionality of slopes of the line through audio representation, or "polarity."[29] Using semantically meaningful data – like dollars, size, and temperature mapped to frequency, tempo, and brightness, they measured if both pools of participants perceived data trends in the same way. While there were no differences between the two populations for most mappings, like size-brightness or temperature-tempo, the frequency-dollars mapping had different polarities for the two populations. As the authors explained: "Sighted participants might be considering that more expensive items within a class,

[27] Flowers, John H., and Douglas C. Grafel. "Perception of Sonified Daily Weather Records." Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 46, no. 17, 2002, pp. 1579–1583.
[28] Mansur, Douglass L., et al. "Sound Graphs: A Numerical Data Analysis Method for the Blind." Journal of Medical Systems, vol. 9, no. 3, 1985, pp. 163–174.
[29] Walker, Bruce N., and David M. Lane. "Sonification mappings database on the web." Georgia Institute of Technology, 2001.

such as automobiles or airplanes, tend to be faster and therefore have higher pitched sounds associated, hence the positive polarity for frequency-dollars, while visually impaired listeners may be more in tune with the everyday sounds of the money itself, noting that a dropped coin makes a high-pitched clink, while a roll of quarters or a bag of notes makes a lower-pitched thud."

With Nees, Walker later explored stock prices on a time-series "tone graph," where the pitch of the stock changed with the price.[30] Their experiment compared two groups: one received feedback on their estimations, while the other had an alternative task to perform. The authors found that the group with feedback performed better, and that it made heuristic sense to map the tone to a single stock price fluctuations over time of day.

In one of the most unique experimental approaches encountered, Brewster and Brown tested participants' understanding of data trends with different instrument timbre and pitch – but by having them draw the graph they just heard.[31] While they only had six participants, they still found that having separate timbre for two different data series helped participants keep the lines differentiated when they redrew them. They also concluded that pitch was a good indication of the trend of the line, e.g. a higher pitch would translate to a higher value on the Y axis.

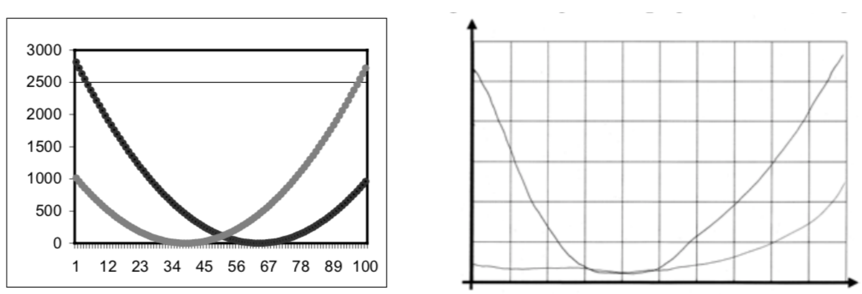

Figure 5. *An example from Brewster and Brown (2003), comparing of the stimuli (left) and the participant's drawing of what they heard (right).*

[30] Walker, Bruce N., and Michael A. Nees. "Conceptual versus perceptual training for auditory graphs." Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Vol. 49. No. 17. Sage CA: Los Angeles, CA: SAGE Publications, 2005.

[31] Brown, Lorna M., and Stephen A. Brewster. "Drawing by ear: Interpreting sonified line graphs." Georgia Institute of Technology, 2003.

Analysis and Discussion

The interdisciplinary selection of papers demonstrated great variety: some demonstrated more creative intent to evoke with more memorable, musical designs, while many focused on accurate, scientific representations. (Perhaps a timeless dichotomy between art and science, this contrast is not unlike the divergence in approaches to visual representation of information, between Nigel Holmes and Edward Tufte.) Observed differences in the data sonification literature manifested in two primary areas:

- *The absence – or rigor – of the experimental procedure.* Some papers proposed frameworks for evaluating or designing an auditory data representation, while others also tested designs with empirical results. Among those papers that included experimental data, however, there was a range of variation in sample sizes and procedures, thus impacting generalizability of the findings.
- *The quality of stimuli used.* Some researchers used abstract, MIDI sounds without any semantic representation, while others mapped audio to semantic meaning some form of representative meaning (like weather patterns). Papers also demonstrated differences in whether or not the researchers mapped sounds to real or simulated data.

For our analysis of existing auditory data representations in graphical form, we thus plotted the 12 reviewed papers along two dimensions: abstract vs. functional data on the X-axis and objective vs. subjective approach on the Y-axis. (See Figure 6 for a summary of paper placements within each quadrant.)

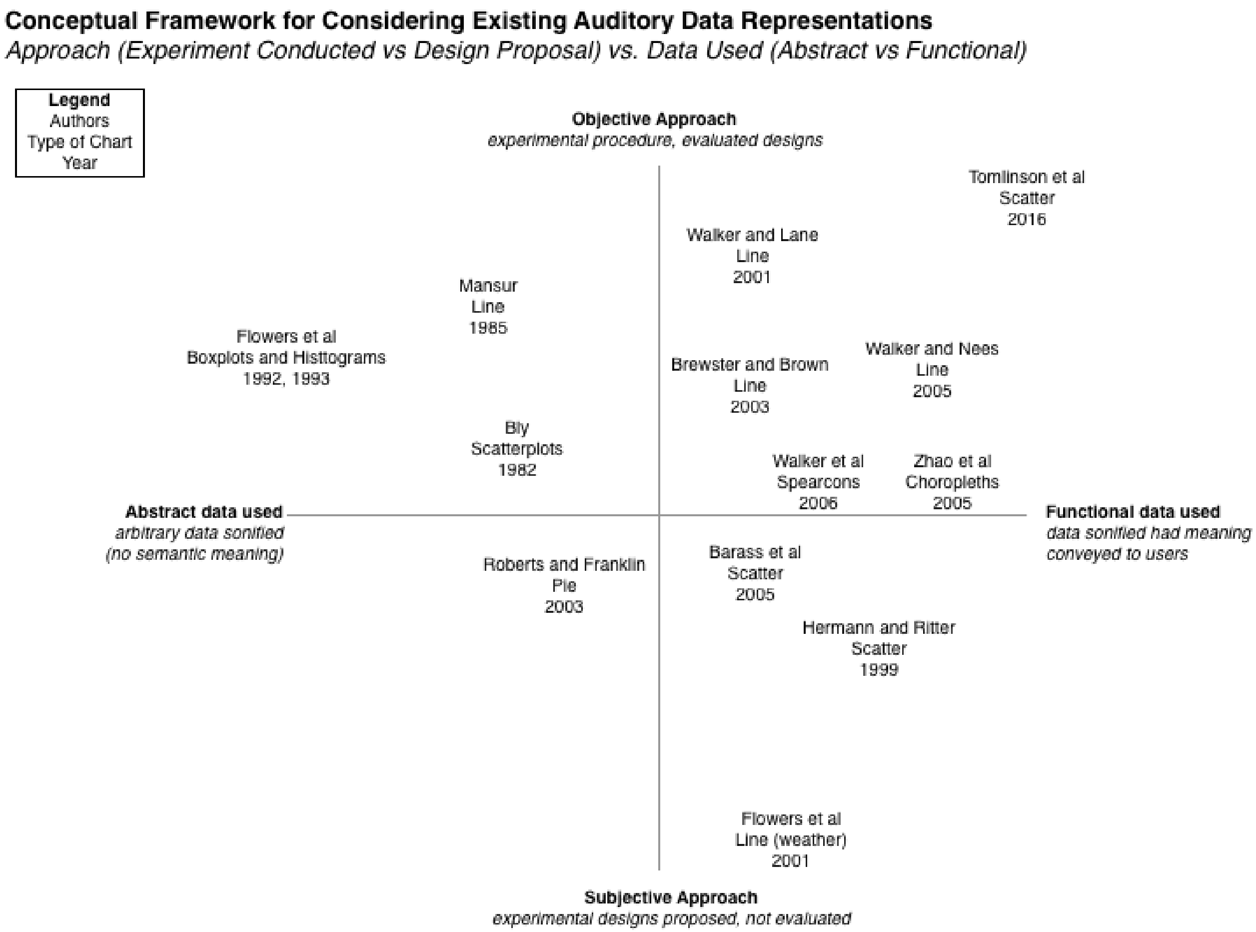

Figure 6. *Framework for considering existing auditory data representations in graphical form. X-axis is the type of data used (Abstract vs. Functional) and the Y-axis is the approach of the paper (Objective Experiment Performed vs Subjective Design Proposed and not evaluated).*

Objective vs. Subjective Approaches

One striking feature of these papers were some were explicitly proposing a new design, while others tested various designs to quantify their results. Depending on their intent and level of rigor, some researchers also ran experiments with sighted and low- to no-vision participants. A few extremes worth pointing out are Flowers et al, 2002,[32] which only proposed possible design patterns, without conducting any experimental evaluation – versus Tomlinson et al, 2016,[33] who conducted a longitudinal study to thoroughly understand the lasting implications of their work.

Abstract vs. Functional Data

Depending on the type of data representation being explored, there were different sources of data used. For instance, Flowers (2002) used weather pattern data to explore if the design made semantic sense, while Flowers (1992, 1993)[34] [35] used simulated data for histogram and box plot skew to test if participants could hear the difference in trends between the plots, although the data did not have semantic meaning. We hypothesize that the use of functionally meaningful data might help ground users and understand the trends better, and we explored this later in our own prototype (see the “Sonification of U.S. Census Data” section below).

Sonification Stimuli Used

There are multitudes of different ways to create auditory stimuli for sonifying data, including:

---

[32] Flowers, John H., and Douglas C. Grafel. “Perception of Sonified Daily Weather Records.” Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 46, no. 17, 2002, pp. 1579–1583.

[33] Tomlinson, Brianna J., et al. “Exploring Auditory Graphing Software in the Classroom.” ACM Transactions on Accessible Computing, vol. 9, no. 1, 2016, pp. 1–27

[34] Flowers, John H., and Terry A. Hauer. “‘Sound’ Alternatives to Visual Graphics for Exploratory Data Analysis.” Behavior Research Methods, Instruments, & Computers, vol. 25, no. 2, 1993, pp. 242–249

[35] Flowers, John H., and Terry A. Hauer. “The Ear’s versus the Eye’s Potential to Assess Characteristics of Numeric Data: Are We Too Visuocentric?” Behavior Research Methods, Instruments, & Computers, vol. 24, no. 2, 1992, pp. 258–264

- the pitch, or frequency, can be shifted to indicate higher or lower amounts,
- rhythmic pattern can be changed to indicate quantity or different categories,
- timbre (i.e., the different “characteristic” or instrument of a sound – like clarinet or trumpet) can signify different categories, or
- tones can be selected for semantic meaning – like footsteps to indicate a more populous area, and crickets to indicate a less populous area.

While it is still unclear from the papers we explored which types of sonifications work the best, there are metaphors that persist in music and map well with mental models. These include higher pitch meaning more, and lower pitch meaning less (when communicating trends or relative distances); alternately, “Morse code” like rhythmic patterns are effective in signifying quantity (see Franklin and Roberts, 2003)[36].

Proposed Frameworks

Apart from our own analytical framework, two auditory representation concepts were proposed in the aforementioned papers explored. Zhao et al[37] articulated the “Auditory Information Seeking Principle” (AISP) as a set of interactions needed to easily explore data through sonification. This consists of four types of data interactions:

- *Gist:* quick grasp of the overall data trends and patterns from a short auditory message.
- *Navigate:* fly through the data collection and closely examine portions of interest.
- *Filter:* seek data items satisfying certain criteria.
- *Details-on-demand:* obtain details of groups or an individual item for comparison.

[36] Franklin, Keith Michael, and Jonathan C. Roberts. "Pie chart sonification." Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on. IEEE, 2003.
[37] Zhao, Haixia, et al. "Sonification of Geo-Referenced Data for Auditory Information Seeking: Design Principle and Pilot Study." ICAD. 2004.

This framework offers an auditory parallel to Scheiderman's visual information seeking mantra of "overview first, zoom and filter, then details-on-demand."

Similarly, Barrass[38] also introduced a new way to consider sonification of data, drawing inspiration from prior descriptions of "everyday sounds" by "material type, size, and hardness,"[39] and "speech sounds" by "sex, maturity, and expression."[40] Though not precisely an "auditory graph," Barrass expands upon the link between pitch circularity and the color wheel – or hue circle – recognizing, that "both timbre and hue perception are qualitative and categorical." He thus proposed the Timbre-Brightness-Pitch framework (CTBP) from psychoacoustics to a perceptual sound space (PSS), arguing that this technique could be used to "analyze the display, compare displays, and optimize data mapping sequences for the display."

Multimodal Approaches

It is important to note that all of these papers we explored either were passive listening experiences, or used a mixed-modality approach, so it is difficult to measure the efficacy of the designs comparatively. Many designs used keyboard, mouse, or joystick interactions in addition to sound, and many of the papers did not use speech feedback to help guide participants. However, the work at the Georgia Tech Sonification Lab by Bruce Walker used some speech cues that we were inspired by for creating our VUI prototype. Particularly interesting are the spearcon and spindex explorations, which use different volumes of speech as one navigates through an address book of names.[41]

---

[38] Stephen Barass, "A Perceptual Framework for the Auditory Display of Scientific Data," 2005

[39] Gaver, W. 1986. Auditory icons. Human-Computer Interaction, 2(2) Lawrence Erlbaum Assoc., Mahwah, NJ.

[40] Gibson, J. J. 1966. The Senses Considered as Perceptual Systems. Houghton Mifflin, Boston, MA.

[41] Jeon, M. and Walker, B. N. 2011. Spindex (speech index) improves auditory menu acceptance and navigation performance. ACM Trans. Access. Comput. 3, 3, Article 10 (April 2011).

Sonification of U.S. Census Data

Context

We had difficulty understanding and assessing the efficacy of different auditory data representations because many of the stimuli were unavailable to be heard, and quite a few papers' results made it unclear how effective these designs truly were. We decided to create and test a prototype of our own to a curated selection of government data.[42] We first performed usability testing with remote users to assess what types of interactions people would want to have with governmental data, and we built our prototype around existing HCI research literature, as well as insights from expert interviews with Lucy Greco, Accessibility Evangelist at the University of California. Our prototype reproduced three different methods for data sonification with 2010 Census data for population by state[43] and data from the 2015 American Community Survey for employment and unemployment – by age[44] as well as education.[45]

We sought and recruited a five-person sample that would embody our articulated "citizen analyst" persona (one male, four female), demonstrating a range in terms of familiarity with: smart speakers, musical knowledge, U.S. geography, and census data. During each moderated session, the participant was presented each of the three different audio representations to explore, then asked follow-up questions about initial impressions, perceived difficulty, and user expectations. We also randomized the presentation of the designs to each participant: three participants received the employment data sonification first, while two participants received the population data sonification

[42] This was for our Needs and Usability INFO 214 Final Project, final paper can be found here: https://docs.google.com/document/d/1DCne77MQdPmzygdHyEnj0awnP5rMcLxbDewCktepLkI/

[43] https://www.census.gov/2010census/data/apportionment-dens-text.php

[44] https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=DEC_10_DP_DPDP1&src=pt

[45] https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_16_5YR_S2301&prodType=table

first. At the end of the session, participants were asked if they had any favorite auditory representations, overall impressions, and suggested improvements. (See Appendix 3 for full usability testing script.)

Stimuli

Population - Choropleth Map (Zhao et al., 2004)[46]

For this design, we replicated a technique used by Zhao et al. to sonify a choropleth map. We grouped the states populations into the five major regions in the United States (excluding Hawaii, Alaska, Puerto Rico, and Washington DC) and mapped each population to a pitch – the higher the pitch, the larger the population; the lower the pitch, the lower the population. For this design, we first play the states by region from north to south, then west to east where only the region is read aloud. Participants are then able to get "details on demand" and zoom in, hearing the region read followed by the individual state label and population sound.

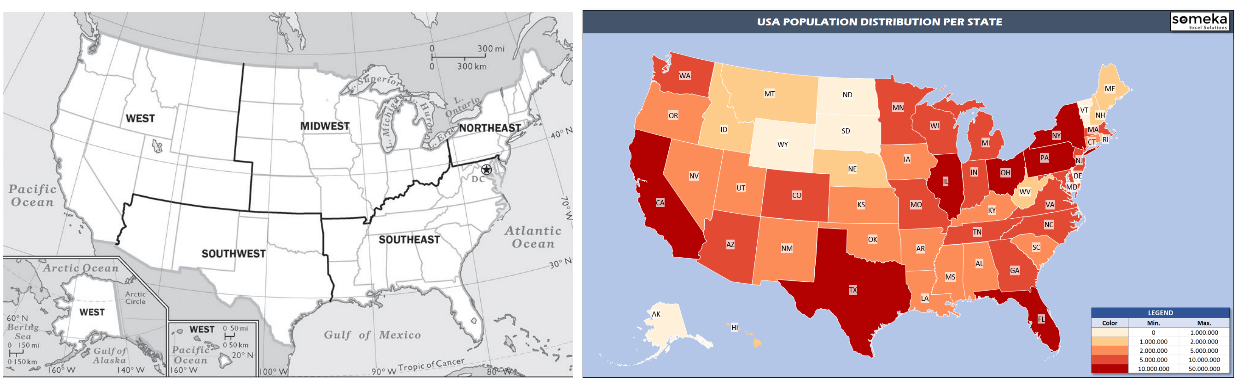

Figure 7. *Visual representation of population data. (left) Image of U.S. map split into different geographical regions that we used for our prototype.*[47] *(right) Choropleth (heat map) of the U.S. Population.*[48]

[46] Zhao, Haixia, et al. "Sonification of Geo-Referenced Data for Auditory Information Seeking: Design Principle and Pilot Study." ICAD. 2004.

[47] © 2009 National Geographic Society, Washington, D.C.

[48] Generated with https://www.someka.net/excel-template/usa-heat-map-generator/.

Employment and Unemployment by Age - Line Graph (Brewster and Brown, 2003)[49]

This design used timbre to illustrate two different data categories: employment (piano, positive) and unemployment (trumpet, negative). Similar to the previous design, we used pitch to indicate the percentage of employment or unemployment for each age group. For this design, Alexa first plays the employment dataset, then then unemployment dataset. After, the participant is able to play the individual employment or unemployment subsets again, and Alexa reads out the order of the categories by age (“Age 16-24…” etc.) and then plays the entire trend of the data.

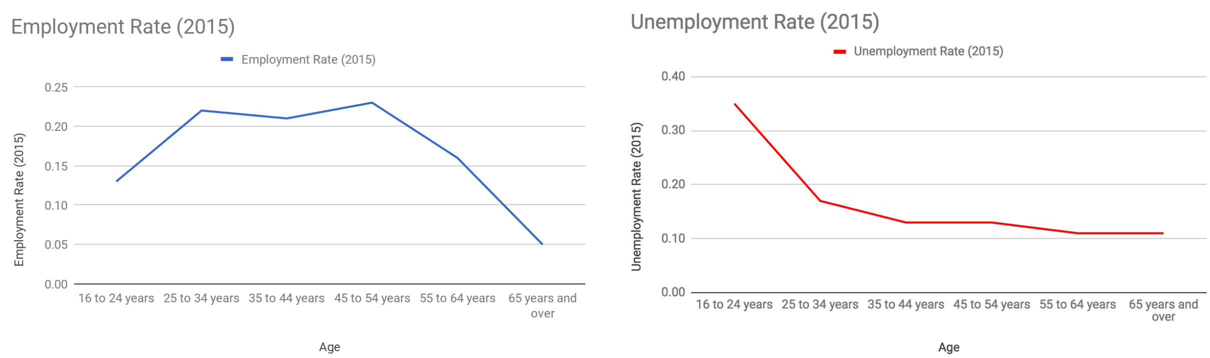

Figure 8. *Visual representation of employment by age data. (left) Employment rate by age groups (right) Unemployment rate by age groups.*

[49] Brown, Lorna M., and Stephen A. Brewster. "Drawing by ear: Interpreting sonified line graphs." Georgia Institute of Technology, 2003.

Employment and Unemployment by Education - Pie Chart (Franklin & Roberts, 1992)[50]

This design also employs the use of timbre to illustrate two different data categories: employment (piano) and unemployment (trumpet). Unlike the previous designs, we did not use pitch to indicate the percentage value, instead we used a design from the author's paper for sonifying pie charts where the "...inspiration from Morse code, where a combination of long and short beeps, known as dashes and dots respectively, construct individual letters. A value of 10% is indicated by a dash, whilst a dot symbolizes a value of 1%."

For this design, Alexa first plays the employment dataset, then then unemployment dataset. After, the participant is able to play the individual employment or unemployment subsets again, and Alexa reads out each individual category label by education level (e.g. "Less than high school…"), plays that specific stimuli in the data, then reads out the next label and plays that stimuli, etc.

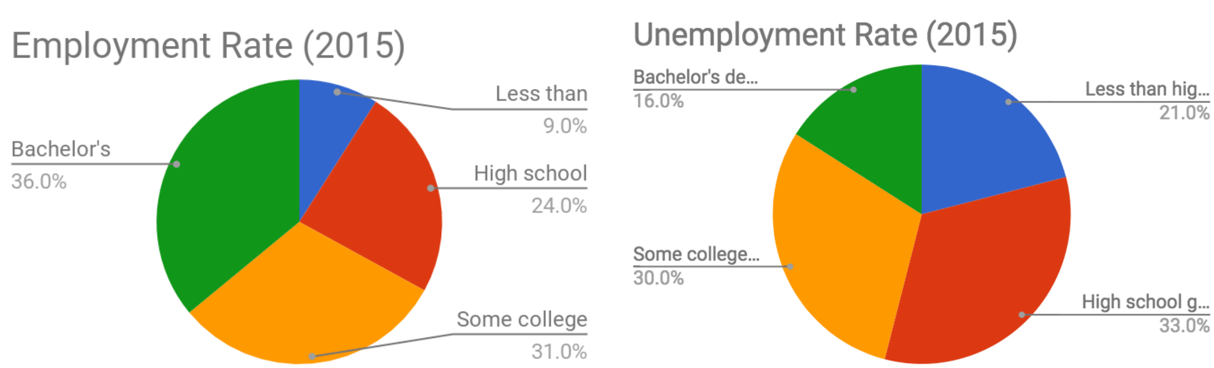

Figure 9. *Visual representation of Employment by education level data, used as sources for our stimuli. (left) Employment rate by education level. (right) Unemployment rate by education level.*

[50] Franklin, Keith Michael, and Jonathan C. Roberts. "Pie chart sonification." Information Visualization, 2003. IV 2003. Proceedings. Seventh International Conference on. IEEE, 2003.

We sonified the data using Ableton Live and Audacity, and we developed the skill in Storyline and launched it on the Alexa Skills store (pending approval). We did all of our usability testing using the Echo Tap, and our skill, “Tally Ho.” You can view the script for our skill in the appendix, and our actual skill mock up here:

https://getstoryline.com/shared/projects/d6d7badf1378536c3a6b34cd964d4aab4cbf9a66

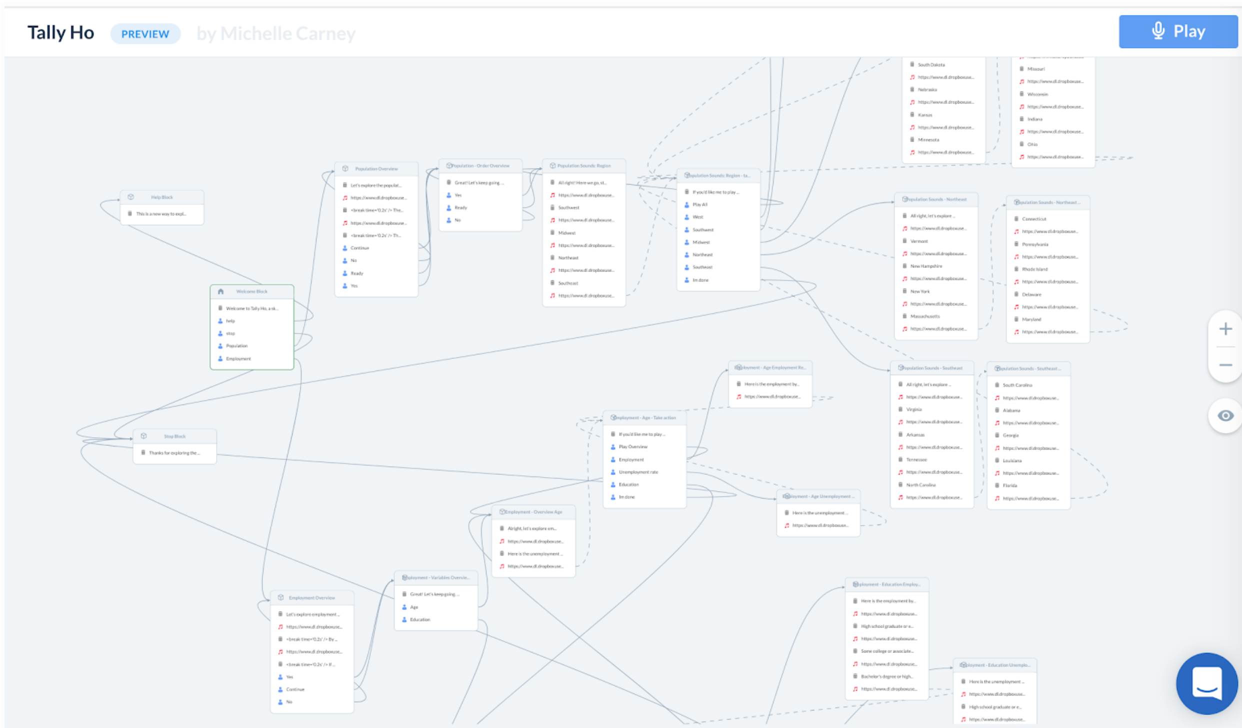

Figure 10. *Overview of Storyline skill created from script (Appendix 1) and mp3s.*[51]

[51] https://www.dropbox.com/home/USpop/audio_v2

Results

From our usability testing, we found a few key important findings. We can note the findings into two groups: reactions to the designs, and what participants would change.

Reactions

One of the immediate findings was understanding people's initial reactions: we did not preface the stimuli with any explanation as to what the participants would hear, and it was interesting what they found important. Participants noted that the labels and instruments were the most salient, rather than specific tones – they could tell the facilitator exactly what the instrument meant, and possibly the trend, but they could not remember which pitches were higher. They also appreciated the instructions in the skill being repeated multiple times, as this was a new exercise in listening for them, and the repetition helped prepare them for listening (the facilitators did not repeat any prompts). Even with the repetition, 4 out of the 5 participants noted that it was "confusing" to remember what they had just heard, and they had no precise recall. However, even with that, all 5 participants noted they thought it was "interesting," "cool," or "powerful" and that there was something novel and exciting about the designs, as they had never "listened to data."

What participants would change

For the population map, while 3 participants stated they enjoyed listening to the data ambiently, there was a request from 2 participants to have the labels of state names first to make it more functional, rather than only upon “zooming in” to each region. It was clear that there was a mental tradeoff between communicating the discrete data points individually or as an overall trend.

For the employment designs, we drew from Brewster and Brown’s 2003 paper and had consciously chosen to map piano to employment and trumpet to unemployment. It appears that the sound metaphor was consistent with our participants. However, one person specifically called it out saying that they really enjoyed it, while another said it might be perceived as insensitive to map the trumpet to unemployment – since people who are unemployed may not be so by choice. It was clear that the “meaningful” sounds were more salient and cognitively easier to remember, but were also open to subjective interpretation.

Lastly, the employment data sonified for education – the Morse code inspired design – was one of the most difficult designs for participants. All participants (5 out of 5) said it was too challenging and they got lost; participants wanted labels for every example since the pause between each was too ambiguous, and they were not sure when to stop counting. Participant backgrounds and usability test results per participant are summarized in Tables 1 and 2 below.

Table 1. *Screener results of participant backgrounds*

Participant Number	Own Smart Speaker?	Used Smart Speaker?	Musical Training	U.S. Census Knowledge
1	No	Yes	Moderate	High
2	No	No	Moderate	High
3	Yes	Yes (Daily)	Moderate	Low
4	Yes	Yes (Daily)	Moderate	Moderate
5	No	No	High	High

Table 2. *Usability test results per participant*

Participant Number	Favorite Design	How many explanations of sound experience correct	How many times explored “details on demand”
1	Population	3/3	2/3
2	Employment - Age (Morse Code)	3/3	1/3
3	Employment - Education (pitches)	3/3	2/3
4	Population	3/3	1/3
5	Population	2/3	0/3

Conclusion and Future Work

From this exploration into data sonification, we have found that there have been multiple attempts to sonify different types of visual data representations or graphs. However, given the interdisciplinary body of literature, there has not been a comparative review of past efforts, which we hope to fill. We explored papers based on what type of visual data representation the authors were trying to sonify, and found that attempts to sonify everything from choropleths to box plots have been made. When analyzing these papers by methodology, we found that some papers proposed designs with limited or no testing, while other papers did rigorous experimentation. There was also wide array of types of stimuli used, from those that were from fabricated datasets to datasets with meaning (i.e., weather patterns). The stimuli that were created ranged from varying the timbre, pitch, rhythm, brightness of the tones, sometimes having underlying meaning that matched the data it was representing (i.e., staccato piano to match rainfall) and other times it was just abstraction into pure tones or sine waves. Additionally, many of these auditory data representations were either explored passively, or in conjunction with another device (visual GUI, keyboard, mouse, joystick, etc).

It was difficult to assess how well each of these designs truly worked, since we were unable to obtain many of the stimuli from the original papers. We thus decided to recreate some of the stimuli, applying approaches used in certain papers to data from the 2010 U.S. Census and 2015 American Community Survey – and presenting them in a contemporary, conversational interface, Amazon's Echo device. We chose to sonify the U.S. population data by state, according to Zhao et al's sonified choropleth, as well as employment trends by age (drawing from Brewster and Brown's timbre line graphs) and by education (like the Franklin and Roberts Morse code pie chart). Out of these three designs, there was no clear favorite, although 3 out of 5 of the participants most enjoyed the sonified choropleth of the U.S. population data by. All participants remarked that this was a first-time experience to listen to data, and, while they were unable to precisely recall the data in the designs, they enjoyed the experience and found it fun and exciting.

It was clear from our user research that listening to sonified data in conversational interface is a new and novel task, and all of our participants saw potential for future interventions. However, similar to the literature we read, it is not clear how effective our designs were at actually communicating the data. For future work, we propose a longitudinal approach like Tomlinson et al – participants would use various sonified datasets over time and be evaluated upon task completion, or areas of improvement in speed and comprehension.

Another point that became clear from our research was the user's expectation to easily ask the smart speaker to query only a few results and aurally display them. Given the limitations of past designs in the literature not being incorporated into smart speaker devices, it is clear that the medium of a virtual assistant allows for new possible interactions. Furthermore, it is becoming ever easier to create new auditory prototypes using speech, given that (a) natural language processing is increasingly prevalent and (2) creators of smart speakers are opening up and offering more support for their platforms, enabling greater access to new audio-application developers. Even recent authors are still grappling on this very subject of data sonification. Just last month at CHI 2018, there was paper on "Investigating Perceptual Congruence Between Data and Display Dimensions in Sonification;" it highlights that "listeners often have a mental model of how a data variable should sound during sonification and this model is not considered in most data:sound mappings."[52]

[52] Ferguson, Jamie, and Stephen A. Brewster. "Investigating Perceptual Congruence Between Data and Display Dimensions in Sonification." Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. ACM, 2018.

While this may be a fun and exciting new approach to data interaction design, it may have practical implications as well. If we are able to design auditory data representations in a meaningful and scalable way, we can help make data accessible to people with limited numeracy, much like the Culturally Situated Design Tools (CSDT)[53] which strive to improve numeracy education through the examples from everyday life. It could also help create better tools for those who are visually impaired; when these tools become used by many people, they get more attention and resources. We hope that this project helps expands the possibilities of voice-enabled interfaces and furthers the advancement of novel data exploration tools.

[53] https://csdt.rpi.edu/about/

Appendices

Appendix 1. Interaction Script for Prototype

URL for final design: https://getstoryline.com/projects/14208

Welcome

Welcome to Tally Ho, a skill to explore U.S. Census Data! You can explore population by state, or employment by age or education level. If you'd like me to repeat anything at any time, just say, "Repeat." Which would you like to explore: population or employment?

Help

This is a new way to explore the US Census Data through audio graphs, or visualizations translated into sounds. <break time='0.2s' /> There are two datasets to explore: population and employment. To explore population by state, say "Population." To explore employment by age or education level, say "Employment." And if you'd like me to repeat anything at any time, just say, "Repeat."

Stop

- Thanks for exploring the U.S. Census Data with Tally Ho! Good-bye.
- Thanks for exploring with Tally Ho! To learn more about the upcoming 2020 Census, visit www dot census dot gov.

Population - Pitch Overview

Let's explore the population of each state, translated into sound. The higher the pitch, the larger the state population. For instance, this is the pitch of Wyoming, which is the state with the smallest population: [WY.mp3] <break time='0.2s' /> The small population means a lower pitch. <break time='0.2s' />

By contrast, here's the sound of California, which is the state with the highest population: [CA.mp3] <break time='0.2s' /> The larger population means a higher pitch. <break time='0.2s' /> So lower pitches mean smaller populations, while higher pitches mean bigger populations. If you would like me to repeat, say "repeat" at any time. Are you ready to continue?

[User options: Yes/It makes sense, No/Repeat]

Population - Order Overview

Great! Let's keep going. The order of state population sounds will go North to South, then West to East. I will group states by region of the United States, starting in the West, then the Southwest, Midwest, Northeast, and Southeast regions. I won't say the name of the state, but I will say the name of the region to make sure you don't get lost. <break time='0.2s' /> If you'd like me to repeat, let me know. Are you ready to hear the population map?

[user options: Yes/Ready, No/Repeat]

Population Sounds
All right! Here we go, starting with the west:

[West.mp3] <break time='0.2s' />
Southwest [Southwest.mp3] <break time='0.2s' />
Midwest [Midwest.mp3] <break time='0.2s' />
Northeast [Northeast.mp3] <break time='0.2s' />
Southeast [Southeast.mp3] <break time='0.2s' />

Would you like me to play this again? Just say "Play All", or the name of the specific region to explore further. The region options are West, Southwest, Midwest, Northeast, and Southeast. <break time='0.2s' /> Otherwise, you can just say "I'm done."

[user options: Play All, West, Southwest, Midwest, Northeast, Southeast, Done - Population]

Population Sounds - West
All right, let's explore the population of states in the West. The order of sounds will go North to South, then West to East again. <break time='0.2s' /> Here we go, starting with Washington:

[WA.mp3] Oregon [OR.mp3] California [CA.mp3] Idaho [ID.mp3] Nevada [NV.mp3] Montana [MT.mp3] Wyoming [WY.mp3] Utah [UT.mp3] Colorado [CO.mp3]

Would you like me to play this again? Just say "Play Again", or the name of another region to explore. The other region options are Southwest, Midwest, Northeast, and Southeast. <break time='0.2s' /> Otherwise, you can just say "I'm done."

Population Sounds - Southwest
All right, let's explore the population of states in the Southwest. The order of sounds will go North to South, then West to East again. <break time='0.2s' /> Here we go, starting with Arizona:

[AZ.mp3] New Mexico [NM.mp3] Texas [TX.mp3], Oklahoma [OK.mp3]

Would you like me to play this again? Just say "Play Again", or the name of another region to explore. The other region options are West, Midwest, Northeast, and Southeast. <break time='0.2s' /> Otherwise, you can just say "I'm done."

Population Sounds - Midwest

All right, let’s explore the population of states in the Midwest. The order of sounds will go North to South, then West to East again. <break time='0.2s' /> Here we go, starting with North Dakota:

[ND.mp3] South Dakota [SD.mp3] Nebraska [ND.mp3] Kansas [KS.mp3] Minnesota [MN.mp3] Iowa [IA.mp3] Missouri [MO.mp3] Wisconsin [WI.mp3] Michigan [MI.mp3] Indiana [IN.mp3] Ohio [OH.mp3]

Would you like me to play this again? Just say “Play Again”, or the name of another region to explore. The other region options are West, Southwest, Northeast, and Southeast. <break time='0.2s' /> Otherwise, you can just say “I’m done.”

Population Sounds - Northeast

All right, let’s explore the population of states in the Northeast. The order of sounds will go North to South, then West to East again. <break time='0.2s' /> Here we go, starting with Maine:

New Hampshire Vermont New York Massachusetts Rhode Island Connecticut Pennsylvania New Jersey Delaware Maryland

Would you like me to play this again? Just say “Play Again”, or the name of another region to explore. The other region options are West, Southwest, Midwest, and Southeast. <break time='0.2s' /> Otherwise, you can just say “I’m done.”

Population Sounds - Southeast

All right, let’s explore the population of states in the Southeast. The order of sounds will go North to South, then West to East again. <break time='0.2s' /> Here we go, starting with West Virginia:

Virginia Virginia Arkansas Tennessee North Carolina South Carolina Mississippi Alabama Georgia Louisiana Florida

Would you like me to play this again? Just say “Play Again”, or the name of another region to explore. The other region options are West, Southwest, Midwest, and Northeast. <break time='0.2s' /> Otherwise, you can just say “I’m done.”

Population - Done

We hope you enjoyed exploring the U.S. Census Bureau's population data! You can say "Employment" to explore employment data, or you can say “Stop” to quit Tally Ho.

[user options: Employment, Stop/Quit]

Employment - Instrument Overview

Let’s explore employment and unemployment data from the past five years, translated into sound. Employment will always be played with a piano, while unemployment will be played with a trumpet. For instance, this is the sound of employment: [Piano.mp3] <break time='0.2s' /> By contrast, here's the sound of unemployment: [Trumpet.mp3] <break time='0.2s' /> Are we good to continue, or would you like me to repeat this?

Employment - Variables Overview
Great! Let's keep going. We can get a better understanding of employment data by filtering according to age or education level. Which do you want to explore: age or education?

[User options: Age, Education]

Employment - Age
All right, let's explore employment by age groups, translated into sound. Lower pitches mean a lower percentage, while higher pitches mean a higher percentage. Remember that the sound of employment will be played with the piano, while unemployment will be played with a trumpet. Here is the employment percentage, starting with the 16 to 24 year old age group, then ending with the oldest age group, 65 years and over:

[employment-piano.mp3] Employment-age: 16 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 years and over

Here is the unemployment rate, starting with 16 to 24 year old age group, then ending with the oldest age group, 65 years and older:
[unemployment-trumpet.mp3] Unemployment-age: 16 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 years and over

If you'd like me to play this overview again, say "Play Overview", or I can play the specific subset - just say "Employment", or "Unemployment" to explore in more detail. <break time='0.2s' /> You can also say "Education" to explore a different aspect of this data. <break time='0.2s' /> "Otherwise, you can just say "I'm done."

Employment - Age Employment Repeat
Here is the employment by age groups, from 16 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 years old and older.
[employment-piano.mp3] Employment-age: 16 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 years and over

Employment - Age Unemployment Repeat
Here is the unemployment by age groups, from 16 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, and 65 years old and older.

[unemployment-trumpet.mp3] Unemployment-age: 16 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 years and over

Employment - Education

All right, let's explore employment rates by education level, translated into sound. We'll represent a value of 10% with a longer tone, versus a value of 1% with a short tone. For example, 4 long tones and 3 short tones would add up to represent 43%. Don't worry, we'll start slow so you get used to counting. Each education level will play at a progressively faster speed.

Remember that the sound of employment will be with the piano, while unemployment will be played with a trumpet. Here we go, starting with people below the high school graduation level, then ending with the most educated people, with a bachelor's degree or higher:

Employment-Education.mp3: Less than high school graduate, High school graduate or equivalent, Some college or associate's degree, Bachelor's degree or higher

Unemployment-Age.mp3: Less than high school graduate, High school graduate or equivalent, Some college or associate's degree, Bachelor's degree or higher

Would you like me to play this again? Just say "Play Again", "Employment" or "Unemployment" to explore in more detail. <break time='0.2s' /> You can also say "Age" to explore a different aspect of this data. <break time='0.2s' /> "Otherwise, you can just say "I'm done."

Employment - Done
We hope you enjoyed exploring Tally Ho! You can say "Population" to explore population data, or you can say "Stop" to quit.

[user options: Population, Stop/Quit]

Appendix 2. Usability Testing Screener Survey

1. Do you own a smart speaker? (Google Home, Amazon Echo/Alexa, Apple HomePod)
    a. Yes, I own one
    b. No, I do not own one but I have used one before
    c. No, I do not own one and I have never used one
2. How often do you use your smart speaker?
    a. Every day
    b. Once or twice a week
    c. Monthly
    d. Almost never
    e. I don't own a smart speaker
3. What do you use your smart speakers for?
    a. Listening to music
    b. Listening to podcasts
    c. Ordering things (delivery food, items from Amazon, etc)
    d. Listening to the news
    e. Controlling devices around the home
    f. Other: ____________________
4. How many years of music training or experience do you have? [11-point scale: No music training or experience – 10+ years of music training or experience]
5. How would you rate your musical ability? [5-point scale: No musical ability – Expert musician]
6. How would you rate your familiarity with the US Geography? [5-point scale: Not at all familiar (could place <3 states) – Very familiar (could place all 50 states on a map)]
7. How familiar are you with the US population and census data? [5-point scale: No familiarity on metrics used – Very familiar on common metrics]
8. Is there anything else you would like us to know? [Optional text response]

Appendix 3. Usability Testing Script

Intro

Thank you so much for taking the time to chat with us today! The purpose of today's testing sessions will be to evaluate the usability of a new way to represent US census data - using sound. We will be asking you to evaluate a few different designs on how to represent these data - I want to clear that we are evaluating the usability of the designs, not you! This is a prototype so the interface may not behave exactly as you would expect, but after each task we will ask you to walk through what you thought, since we are unable to "think aloud" and pause the interaction as we go through this voice interface, so try to remember anything that you might expect or that surprised you.

We are doing this usability study as our final project for our Needs and Usability class. Would it be ok if we recorded this session and your image for our notes and possibly for our final deliverable? We will not identify you by name, but it might not be anonymous since we are using your voice and image. If this is not ok with you, we can focus the camera only on the Alexa device. Do I have your permission to continue?

Warm-Up

Let's get started. Have you used an Alexa before? [wait for user response]

Okay, this is a little bit different than normal Alexa experience. We have made a prototype "skill," or Alexa app, that allows you to explore US Census data through sound. How it will work is that we will say "Open Tally Ho," and Alexa will start walking you through the data that is presented in a new exploratory way. Once we get started, you don't need to push the button, and you can say "repeat" at any time. Also, if the skill breaks, I will read out what Alexa should say, followed by the stimuli. Does that sound ok?

Task 1: Population (Choropleth)

Great, let's get started. For this first task, could you please explore the "population" option in the skill? Whenever you are done exploring, you can say "I'm done" and I have a few questions for you. Let's launch Tally Ho - and then you will be the one talking to Alexa.

- What were your initial impressions? (was it easy or hard)
- Was there anything that surprised you or was something that you weren't expecting?
- What do each of the sounds mean?

Task 2: Employment - Age

Great, so let's explore another subset of the data - employment. You will be prompted to explore employment rates by age or education - for this, choose age. Whenever you are done exploring, you can say "I'm done" and I have a few questions for you. Let's launch Tally Ho - and then you will be the one talking to Alexa.

- What were your initial impressions? (was it easy or hard)
- Was there anything that surprised you or was something that you weren't expecting?
- What do each of the sounds mean?

Task 3: Employment - Education

Okay we have one more for you, so let's explore the last subset of the data. You will be prompted to explore employment rates by age or education - for this, choose education. Whenever you are done exploring, you can say "I'm done" and I have a few questions for you. Let's launch Tally Ho - and then you will be the one talking to Alexa.

- What were your initial impressions? (was it easy or hard)
- Was there anything that surprised you or was something that you weren't expecting?
- What do each of the sounds mean?

Overall Assessment

Great job! Just as a reminder, you explored three different conditions - Population, Employment by Age, and Employment by Education.

- Did you have a favorite? (why or why not)
- Do you feel like you noticed anything different from using sounds?
- What would be one thing you would do in order to improve these examples?