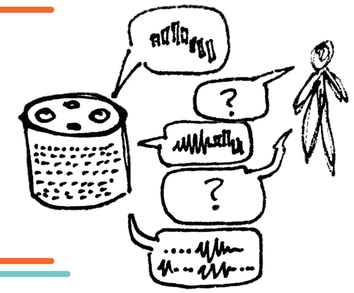
Conversational Sonification: Exploring the intersection of data sonification, voice interfaces, and interactivity

Jordan Wirfs-Brock, University of Colorado Boulder

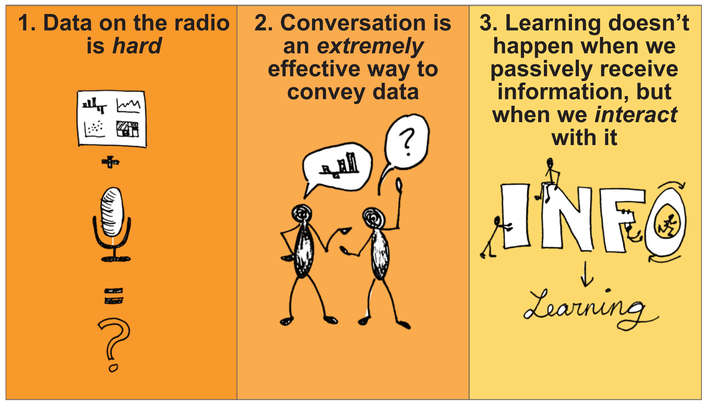


What would an interactive sonification that harnesses the unique possibilities of voice assistants *sound* like?

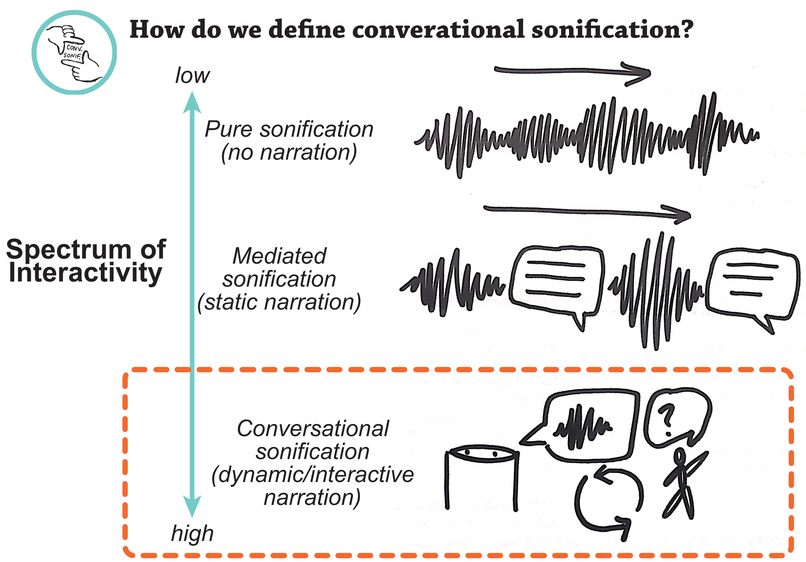
In a world saturated with information, many people still find it difficult to uncover meaning in data. Further, data representations are predominantly visual and we typically experience them alone. This approach neglects vast swaths of our human capabilities. Smart speakers enabled with voice assistants like Amazon's Alexa and Google Assistant present new opportunities for interacting with audio, yet they rely almost exclusively on computer-generated speech to communicate quantitative data. To respond to these challenges, I'm exploring what happens when conversational interfaces, data sonification, and interactivity collide.

This design concept, ***conversational sonification***, combines the give and take of interpersonal communication with dynamic sonification that responds to natural language queries. Conversational sonification is a potential bridge between gallery-installation-based sonification and our homes: We can inject everyday activities (chopping vegetables, watering plants, folding laundry) with rich, multi-modal quantitative data. My research explores the role interactive sonification could play in the context of voice assistants in four sub-areas: aesthetics/best-practices, technology needs, situational context and evaluation.

Inspired by 3 lessons I learned from many years as an educator, data journalist and radio producer



How do we define converational sonification?



Conversational Sonification Design Fictions: Prototyping the Experience

Sonification in the News Feed

While folding laundry, you ask Alexa for the latest news. She reads a few headlines then says, "Here's an audio graphic of DACA program demographics." First, you hear a series of descending musical tones. Alexa tells you that you're hearing data about the number of DACA beneficiaries by state. Next, you hear a chorus of voices, each saying a country's name. Mexico is by far the loudest, but you can also make out whispers from El Salvador, South Korea and India. Alexa says, "The loudness of the voices corresponds to the number of DACA beneficiaries from that country." You ask, "How many are from South Korea?" Alexa says 7,250. You can ask to zoom, filter and compare, and Alexa updates the audio graphic as you do.

Dinner Party Debate

You're hosting a small dinner party, enjoying beers while the enchiladas finish baking in the oven.

The conversation turns to the economy and one of your guests, Roger, says, "What's there to complain about? We all make enough to afford to drink craft beer that costs $13 for a six-pack. And it's not just us—unemployment, right now, is lower than it's ever been."

Another guest, Monica, responds, "Yeah, but that's because people can't file for unemployment because they are perpetual contract or gig economy workers. Something's gotta give."

As the debate thickens, you point out that we don't just have to argue ad nauseum—we can actually pull up some data that will help us understand. You say, "Alexa, what's the unemployment rate?"

Alexa says, "The current unemployment rate in Colorado is 3%. Would you like to know how that's changed over time?"

You say yes.

Alexa now plays you a sonification that is a blend of chattering voices and musical tones. She explains that when the voices get louder, and the pitch rises, this represents higher unemployment levels. She also explains that you'll hear the past 50 years of data, at a rate of one year a second. Every ten seconds, she pops in saying, "...1980...1990...2000...2010…" etc.

After the entire sound clip plays, Alexa asks if you'd like to repeat it, learn more about a specific year or decade, switch from Colorado to the entire U.S. or a different state, or learn how this sonic graph was made.

As she waits for a response, the discussion picks up again. Monica wants to know what is actually being represented by unemployment data. So she asks, "Alexa, tell me more about the unemployment data."

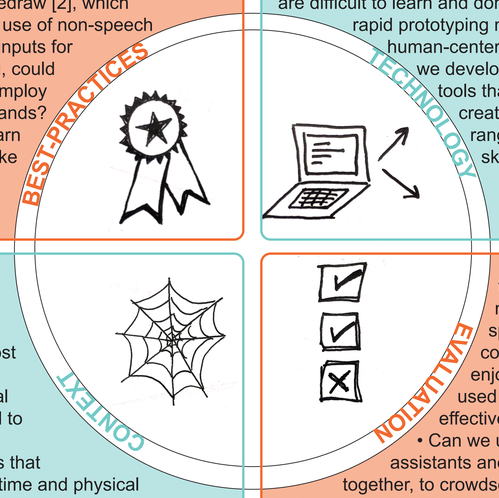
Alexa tells you the data is from the Bureau of Labor Statistics, and starts reading information about how the U.S. federal government calculates unemployment. Roger interrupts Alexa and says, "Wait, so is there data on underemployment?"

Alexa responds that the BLS does collect alternative data on the underutilization of the labor force, and asks if you'd like to hear a sonic graph of that data. In unison, you all say, "Yes."

Immersive Audio-Tactile Lunch

One of your favorite longform investigative news organizations has been doing a series on agricultural water use. They've represented data about how much water we use in the U.S. as a guided recipe for a salad: The amount of each ingredient in the salad is proportional to the water required to grow it (so you had better like almonds). As your smart speaker walks you through the recipe for your lunch salad, telling you to dice half a cup of tomatoes and thinly slice a few grams of cucumber, you hear the voices of farmers who grow each ingredient, telling you stories about their lives. This creates a multi-sensory experience where the story, data and sound combine with tactile immediateness and personal relevance.

Next Steps: Research questions in four key areas



- What best practices can we adapt from radio and podcast production, interactive museum exhibits, and accessible graphics?
- What techniques—parameter mapping, audification, model-based sonification [1] —are most appropriate in a conversational setting?
- What types of interaction work best?
- Inspired by Voicedraw [2], which demonstrates the use of non-speech vowel-sounds as inputs for computer drawing, could voice interfaces employ non-verbal commands?
- Could people learn to use data-karaoke [3] utterances to interact with a smart speaker?

- Which elements of a conversational sonification platform are mature and which require further research and development? Perhaps existing technologies like Tsuchiya et al's Data-to-music API [4] could be extended for voice assistants.
- Authoring tools for conversational interfaces are difficult to learn and don't support the rapid prototyping necessary for human-centered design. Can we develop prototyping tools that support content creators with a broad range of technical skills?

- In what domains—news, education, collaborative data analysis—is this most promising?
- Can conversational sonification be used to facilitate immersive multimodal activities that incorporate sound, time and physical involvement?
- How does the type of content or interaction style change as users move from cooking dinner, driving to work, gardening in the yard, or hosting a board game night?

- What tasks and metrics (*i.e.* time spent, comprehension, enjoyability) can be used to evaluate effectiveness?
- Can we use voice assistants and sonification, together, to crowdsource perceptual studies and collect data about effectiveness?
- Can conversational sonification support and improve the process by which users learn how to interpret sonifications?

References

[1] B. Walker, M. Nees, "Theory of sonification," in The sonification handbook, 2011.
[2] S. Harada, J.O. Wobbrock, and J.A. Landay, "Voicedraw: a hands-free voice-driven drawing application for people with motor impairments," in Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility, 2007, pp. 27-34.
[3] A. Supper, "Data karaoke: sensory and bodily skills in conference presentations," in Science as Culture, vol. 24, no. 4, pp. 436-457, 2015.
[4] T. Tsuchiya, J. Freeman, and L.W. Lerner, (2015). "Data-to-music api: Real-time data-agnostic sonification with musical structure models," Georgia Institute of Technology, 2015.

Contact Information

Jordan Wirfs-Brock
jordan.wirfs-brock@colorado.edu
Twitter: @jordanwb
jordanwb.com
303-884-7591

University of Colorado **Boulder**

Presented at ICAD 2018 with support from the ThinkTank program