Soundbeam: a Platform for Sonifying Web Tracking

Charles Celeste Hutchins
University of Kent
celesteh@gmail.com

Holger Ballweg
BiLE
holger.ballweg@gmail.com

Shelly Knotts
Durham University
michelle.knotts@durham.ac.uk

Jonas Hummel
Manchester Metropolitan University
mail@jonashummel.de

Antonio Roberts
BiLE
antonio@hellocatfood.com

ABSTRACT

Government spying on internet traffic has become ubiquitous. Not to be left out, the private sector tracks our online footprint via web beacons and cookies. Web services such as Google track our progress as we surf the net and click on links. The Mozilla plugin, Lightbeam (formerly Collusion), shows the user a visual map of every site to which a surfer's data is sent. An interconnected web of advertisers and other otherwise invisible data-gatherers quickly builds during normal usage. We have modified this plugin so that as the graph builds, its state is broadcast via OSC.

We will act as translational agents in a process of live data sonification. The collected data is the material with which we will develop a set of musical gestures based on patterns we may discover. The findings of our data collection and the developed music will be presented in the form of an audiovisual live performance. Snippets of collected text and URLs will both form the basis of our audio interpretation and also be projected onto a screen, so an audience can voyeuristically experience the activities of governments and advertisers.

Keywords

big data, Lightbeam, privacy, sonfication

1. BIG DATA

Our original goal in setting out was to create something that was based on Open Source Software that engaged the issue of surveillance. We became specifically interested in big data because of the "public/private surveillance partnership"[9] between big data and governmental bodies and the collaboration this requires of the most popular websites. The Mozilla Foundation, who do activism around issues of privacy and the open internet, publish a Firefox plugin called Lightbeam. This tool, formerly called Collusion, tracks every server that a user visits. This includes both the URLs that they purposefully click on or type into the address bar of their browser and also websites that their intended destination refers to, but that the user does not connect to directly[8].

The additional connections embedded in a page can be entirely benign. For example, Facebook uses a dedicated domain name (fbcdn.net) for pictures. However, some third party sites can and do track users around the web, with the collaboration of website owners. Sites such as ScoreCard Research embed invisible images into third party web pages and use cookies to track browsing habits[1]. These data are used for advertising and audience analytics[3]. ComScore, which owns ScoreCard, tracks two million users at a time[1].

Even websites that do not have advertising help gather information for big data. Many websites are concerned about Search Engine Optimisation: they want to know which search terms brought users to their website, so they install Google Analytics[5]. It is within Google's power just to share that information with website owners. However, with Google Analytics, website owners get a nice interface and can easily buy things from Google to increase traffic, such as search keywords[5] or to track repeat users and serve adverts to them based on past visits[4]. This data-gathering appears on many websites that do not serve adverts. Both IRCAM and the University of Kent have Google tracking on their websites[10, 6].

By design, this tracking is all invisible. The Lightbeam plugin aims to make this visible by building a table that users can view, or, more strikingly, a graph showing the interconnected network of trackers and sites that updates as the user surfs (see Figure 1). While these data are educational to users and useful for researchers, users must divert their attention to the plugin to see the data. Users do not get an immediate sense of the constant growth of the graph.

2. SONIFICATION

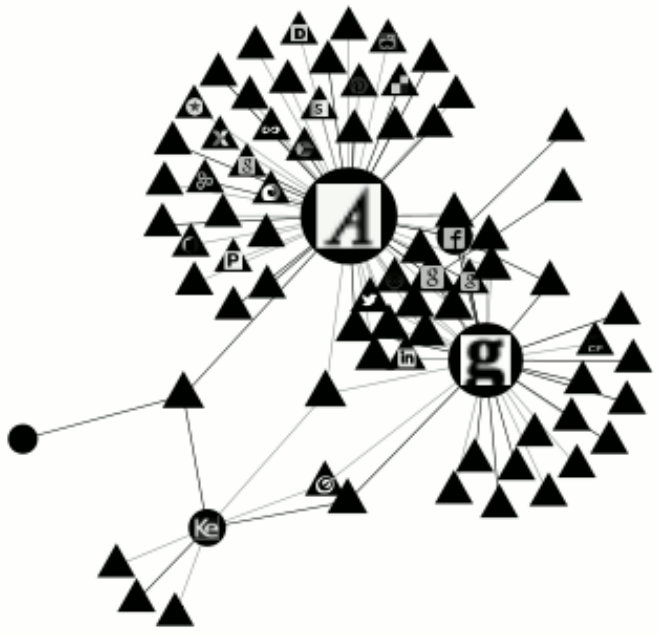
Our project builds on Lightbeam to provide a more powerful illustration of tracking in real time. By sonifying ongoing data, we give users a stronger sense of just how much tracking is going on. Even people who are concerned about the privacy issues surrounding big data do not tend to have a sense of just how many calls to third parties are made when they click an advertising-supported website. With sonification, calls to tracking websites can be brought to the user's attention in real time.

We forked the Lightbeam plugin[2] to add support for OSC messaging. Our version of the plugin uses node.js to send OSC messages to the localhost on port 57120, which is the default port for SuperCollider. The OSC message tags are `/visited` and `/unvisited`, the latter of which are potentially tracking the user. Additional arguments are the site's url, the number of other sites it's connected to, the number of times visited, and the number of cookies attached.

The piece we are developing with this, called Soundbeam, is therefore based on web surfing. Thus, in accord with the old computer-musician joke, we **are** checking our email on stage, but also Facebook, Buzzfeed or whatever other site

NIME'14, June 30 – July 03, 2014, Goldsmiths, University of London, UK.

Figure 1: A Lightbeam graph generated by visiting some of this article's references.



captures our fancy. While we surf, we sonify the third party sites we hit. Additionally, to highlight both the ubiquitousness and interconnectedness of tracking, each user also announces to the LAN the third party sites they encounter. When a user receives such a notification, if they have also been tracked by the same site, they also sonify a match. The score for the piece calls for this sound to be more in the foreground than third party sites hit individually. The presence of cookies, the number of connections and number of times hit may also impact the timbre of the sound event.

In creating this piece, we have researched big data tracking companies and are building a table of regular expressions to match URLs. Matches are associated with known big data companies.

Listing 1: Regular Expressions that match Google

```
google[^ \t\r\n\v\f\.]*\.com
google[^ \t\r\n\v\f\.]*\.co\.uk
gstatic\.com
[^ \t\r\n\v\f\.]*\.google[:word:]*\.com
[^ \t\r\n\v\f\.]*\.googleapis.com
doubleclick.net
```

This is used to determine whether two players are being tracked by the same company. Also, depending on how benign we deem a third party site to be (e.g. Facebook's image serving site has no nefarious intentions that we know of), this may change associated sound events.

As with all BiLE pieces, how sound is generated and played is left up to each player[7]. In order to avoid interfering with web browsing, the score specifies that players should have only minimal interaction with any musical GUI. If players want more control over their gestures, therefore, they will be empowered to use devices such as MIDI faders to control aspects of their sounds.

3. VISUALISATIONS

Graphical projections will greatly increase the clarity of the piece to audiences. This part of the piece is very much still under development, but possible projections could include the Lightbeam graphs generated by some of the players, URLs of tracking sites, and text from websites we intentionally visit.

4. OTHER APPLICATIONS

In addition to live performance, this technology could be well-suited to to interactive installations. Users could be encouraged to browse the web via a public kiosk. As they browsed, sound events would alert them to big data companies tracking them. Additionally, the installation could be fitted with several additional monitors which could display information about the tracking sites they encounter. If a user hit a site with five trackers, such as the University of Kent homepage[10], five extra monitors mounted around the kiosk could display information about each tracker including the owner, their corporate logo, market penetration and types of tracking employed. While this would be less musical than a performance, it could potentially be more educational and make a greater impression on the audience.

5. CONCLUSION

Surreptitious data collection, including via tracking websites is one of the major privacy issues affecting web users today. We hope that the Soundbeam piece described here addresses these issues in a way that is both educational and musical and that the code we develop finds other artistic uses that engage these issues.

6. REFERENCES

[1] T. Beleaga and J. Geary. Scorecardresearch (comscore): What is it and what does it do? *The Guardian*, April 2012. `http://gu.com/p/374xn/tw` [Access date: 30 April 2014].

[2] BiLE. Bilensemble/lightbeam, February 2014. `https://github.com/BiLEnsemble/lightbeam` [Access date: 30 April 2014].

[3] ComScore. Products, 2014. `http://www.comscore.com/Products` [Access date: 30 April 2014].

[4] J. Geary. Tracking the trackers: What are cookies? an introduction to web tracking. *The Guardian*, April 2012. `http://www.theguardian.com/technology/2012/apr/23/cookies-and-web-tracking-intro` [Access date: 30 April 2014].

[5] Google. Features: Google analytics, n.d. `https://www.google.com/analytics/features/index.html` [Access date 30 April 2014].

[6] IRCAM. Www ircam, n.d. `http://www.ircam.fr/` [Access date 30 April 2014].

[7] S. Knotts. Bile manifesto, 2011. `http://www.bilensemble.co.uk/manifesto/` [Access date: 1 April 2014].

[8] Mozilla Foundation. Lighbeam for firefox, n.d. `https://www.mozilla.org/en-US/lightbeam/` [Access date: 30 April 2014].

[9] B. Schneier. Don't listen to google and facebook: The public-private surveillance partnership is still going strong. *The Atlantic*, April 2014. `http://bit.ly/1rutQdN` [Access date: 28 April 2014].

[10] The University of Kent. The university of kent, n.d. `http://www.kent.ac.uk/` [Access date: 30 April 2014].