REPRESENTING TWITTER USERS' ENGAGEMENT BY SONIFICATION

Robert Wheeler

Columbia College Chicago,
Audio Arts and Acoustics Department
600 S Michigan Ave, Chicago, IL 60605
USA
robert.wheeler@loop.colum.edu

David Worrall

Columbia College Chicago,
Audio Arts and Acoustics Department
600 S Michigan Ave, Chicago, IL 60605
USA
dworrall@colum.edu

ABSTRACT

Social Media has become one of the main ways that people engage with each other and companies engage with their customers and potential customers, so the statistical analysis of this engagement has become common, and a useful tool to track social trends and preferences. This project's aim was to create a tool to sonify and analyze Twitter users' engagement with a user-defined topic over time.

1. INTRODUCTION

This project was undertaken for the purpose of quantifying users' engagement with a topic or keyword on Twitter [1] over time, and presenting this data using sonification so that that it can be absorbed in a compressed amount of time. There have been several previous reports of the sonification of Twitter data, one notable example being a 2012 report by Thomas Hermann et al.[2]. The majority of previous work seems to focus on real-time data.

This project ("Mike Twypyson (MT)") takes each tweet (a post by a user of Twitter), converts it into a discrete data-point and sonifies the data-points individually along a user-scalable timeline. Unlike other approaches, this project deals with historical not realtime data in the hope that this would be more useful for observing longer term trends in engagement.

2. DATA COLLECTION

A data set is drawn by the user from Twitter in non-real time. When the program starts, the user is prompted to enter

(a) Search term (a name, keyword, or hashtag).
(b) The length of time they wish the sonification to take place over.
(c) How many tweets should be sonified, and
(d) A "temporal resolution threshold" (see later).

In order to access their database, Twitter requires access credentials which it grants upon registration. In order to access Twitter via Twython [3], a Twitter API, requires access credentials from Twitter, and so the application "Mike Twypyson" (MT) was registered through the site's developer portal to obtain these credentials.. For the Twitter database access, only the search term and the number of tweets to sonify are required. MT retrieves chunks of tweets from the Twitter database for the given user query until the tweets collected meets the number specified by the user, or Twitter runs out of tweets to return. If the number of user-requested tweets is large, accumulating the data may take a considerable time, and/or the search for tweets for a given search-term is exhausted. A list of retrieved tweets is parsed for useful information and turned into `Tweetts`, a Python Class object, for ease of processing.

3. DATA PROCESSING

Once the collation of `Tweetts` is complete, their timestamps are used to order them temporally in preparation for the time-warping necessary for sonification according to the user's requirement for a sonification of a particular duration. A scalable temporal gamut in which all of the sonified tweets will be placed is then calculated. Due to the potentially high density of tweets per minute, however, it proved necessary for some periods of time to be further sub-divided. If, for example, in one minute of "tweet-time" (i.e. real time), there was more than one tweet, we sub-divide the portion of the temporal gamut devoted to that minute into an equal number of divisions, according to the number of tweets in that minute. This creates an arpeggio-like effect. The `Tweetts` in this arpeggio's timbres are increasingly "brightened" by the addition of more harmonics based on their position in the arpeggio.

With the temporal division of the sonification settled, and mechanisms to sonify the density of `Tweetts` in place, the `Tweetts` are analyzed for user engagement, which is quantified using four main criteria:

(a) Retweets,
(b) Favorites,
(c) User Mentions, and
(d) The Follower Count of the user that authored the tweet.

The number of retweets, favorites, and mentions of other users (through authors tagging their handles) are summed to obtain one "engagement value", with each element of this value being weighted based on how important these criteria are when judging engagement. For example, retweets are weighted x4, user mentions are weighted x2, and favorites are weighted x1). This value is converted to the pitch of the tone that will be created from this tweet. If the tweet's author has 10,000 or more followers, this is seen as a potentially influential tweet in its own way, and so, in order to distinguish it, the resulting sonified data-point's fundamental frequency is lowered by three octaves.

When the density of the data over the user-specified sonification time was very high, a user-defined "temporal resolution threshold" mechanism was implemented to make the sonification more listenable. This mechanism lets the user decide how quickly they are able to absorb the data points, combs through the times that these sonified points are set to play, and if the time between points is smaller than what the user has specified for their temporal resolution threshold the `Tweetts` are consolidated into a single sonification event; the timbre brightness adjusted and the duration increased accordingly. In this way, the sonification of tweet density over the allotted playback time is not lost, and the user is made aware that temporal separation of events have been lost. This separation loss is also indicated in the statistics provided to that the user before and after the playing the sonification.

4. PLAYBACK

After the main processing is complete, each `Tweett` is converted into a Csound event and appended to the sonification event list which is then passed to Csound for sound for rendering in real-time or to soundfile. The user is provided with some basic statistics about the sonification they are about to hear, and then prompted to initiate the sound synthesis. The Csound instrument used for the sonification is a simple fixed-waveform wave-table tone generator, which, as discussed earlier, adds additional harmonics based on the density of tweets within time.

5. EXAMPLES

Examples of the project's sonifications are publicly available on Soundcloud [4] are provided in order to illustrate what the project does well, and what the project fails to do well, as well as show the range of results the program can yield:

(a) *monster energy* [5] is the sonification produced by the program with the settings "monster energy" for query term, "120" for the sonification length in seconds, 5,000 tweets being sonified, and 5ms set as the resolution threshold. The example shows a change in users' engagement over time. The sonification begins with sparse tweets, many with little engagement, but with small bursts of both tweet density (sonified with brighter timbres), and engagement with those tweets (sonified by pitch), and with the occasional tweet by a user with over 10,000 followers (sonified by a note 3 octaves lower than usual). As the sonification proceeds, we observe more frequent bursts of engagement, suggesting a trend of increased Twitter user engagement with the keywords over the period of the tweets sonified, which, in this case, was 20 hours prior to the enquiry (and as reported by the program's user interface).

(b) *switch* [6] is the sonification produced by the program with the settings "switch" for query term, "240" seconds for the desired sonification length, 5,000 tweets to be sonified, and 30 milliseconds as the resolution threshold. This sonification highlights the resolution threshold in action well. Around the 6 second mark, we hear the density of sonified data points decrease, but the length and timbre of these points increases. The increased length is the signifier that points have been consolidated, and from the bright timbre we get an indication of the density of the consolidated points (the point we hear takes on the timbre of that of the "brightest" note absorbed).

6. EVALUATION

While this project does provide the user with accurate information about users' engagement with keywords on Twitter, there are many possible improvements. For example, one of the main aims for several of the features of the sonification was to make it more listenable and easily absorbed by the user. Unfortunately, the data sonified can be so temporally dense that it becomes hard to get a sense of trends in the sonified data. High density of tweets over a short period of time can also be a problem in drawing a large enough data set to be able to discern gradual trends in the data over time. Twitter's publicly searchable database of tweets may only extend a day or two into the past when querying for even thousands of tweets, but if the query is extended into the ten-thousands of tweets the time it takes to process them all becomes significant. The lack of information about how many tweets on a specific keyword the public database contains makes the data acquisition process cumbersome: Queries for thousands of tweets often fall short, and more testing is required to determine if this is because of a lack of tweets by users or if the database simply doesn't return some search results. Also, it is difficult for the listener to discern how much time has passed in the sonification in regards to the relation of the timestamp of the current tweet being sonified to the timestamp of any other tweet being sonified. A system to differentiate the different kinds of engagement with the tweets (eg. retweets, favorites, user mentions, etc.) as opposed to combining all of them into one value for general engagement might improve the user's understanding of the complexity of the tweetscape: High numbers of followers for the author of a tweet is identified, but having each type of engagement control different parameters of the sonification's timbre may yield a result with more richness in conveyed data. The use of a more complex sound-synthesis model would make listening to the flurry of sine tones generated by the project easier.

7. REFERENCES

[1] "Twitter. It's What's Happening." *Twitter*, Twitter, http://www.twitter.com/. Accessed May 2, 2018.

[2] Hermann, T., A. V. Nehls, F. Eitel, T. Barri, and M. Gammel, "Tweetscapes: Real-time sonification of twitter data streams for radio broadcasting" in *Proceedings of the 18th International Conference on Auditory Display, Atlanta, GA, USA. June 18-21, 2012*

[3] "Twython." *Twython-Twython 3.6.0 Documentation* http://twython.readthedocs.io/en/latest/ . Accessed May 2, 2018.

[4] *Mike Twipyson* SoundCloud. http://soundcloud.com/twipyson Accessed May 2, 2018.

[5] *monster energy* SoundCloud, http://soundcloud.com/twipyson/monster-energy. Accessed May 2, 2018.

[6] *switch* SoundCloud. http://soundcloud.com/twipyson/switch. Accessed May 2, 2018.