



Five ways data is transforming music

Digital technology turned the music industry on its head. But the revolution is not yet over. **Robert Langkær-Bain** explains how data is changing the way music is created, shared, heard and understood



One day last summer, thousands of users of the music streaming service Spotify spontaneously started listening to the same song. The track was Bonnie Tyler's *Total Eclipse of the Heart*. The date was 21 August – the day of the solar eclipse.

This topical song saw a 3521% spike in popularity as listeners across the United States hit "play" to coincide with the eclipse. Spotify made a visualisation to show this happening, with peak streaming of the song seen moving from coast to coast, roughly following the path of the moon's shadow (bit.ly/spotifyeclipse). It was a vivid illustration of the immediacy of digital music, and of the vast amount of information on listening habits that is now up for grabs.

Data can tell us how listening habits are changing. But these changes also may be influencing the structure of the music we listen to. For example, long instrumental intros are fast

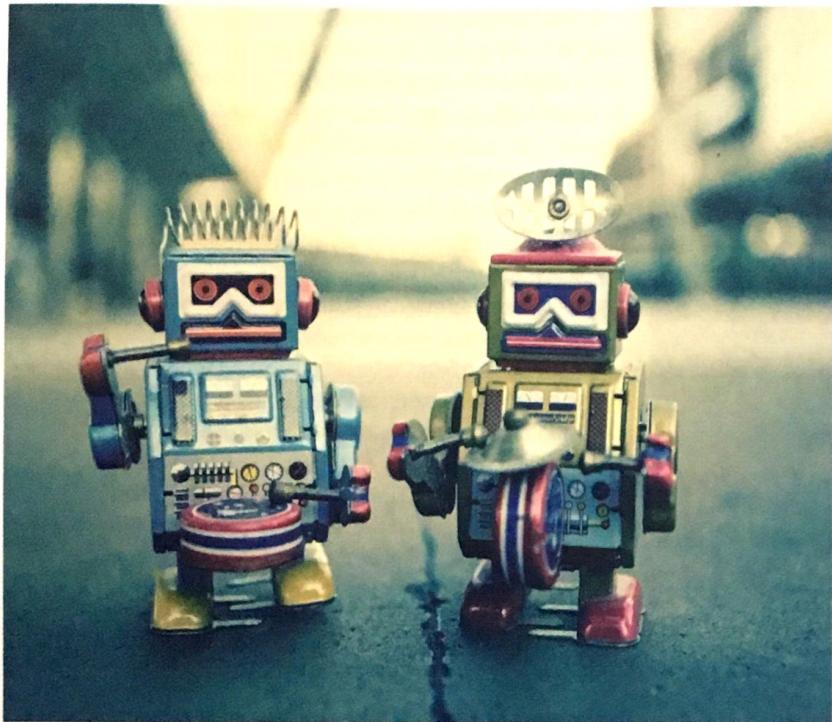
becoming a thing of the past, according to Hubert Léveillé Gauvin of Ohio State University, who analysed hundreds of top-10 hits from 1986 to 2015. He found that the average intro has become 78% shorter in that time, falling from more than 20 seconds to about five.¹

Léveillé Gauvin puts this down – at least partly – to digital music and the meteoric rise of streaming services like Spotify, Deezer and Apple Music. "If people can skip so easily and at no cost, you have to do something to grab their attention," he says (bit.ly/2z7fSJK).

Armies of data scientists and analysts are engaged in this fight for listener attention. Their goal is to figure out how to write songs people want to hear, to identify up and coming acts and match them with potential fans, and to track all the buzz being generated along the way. These are just some of the ways that data is transforming music.



Robert Langkær-Bain
is a freelance journalist based in the UK, and a regular contributor to *Significance*. He was previously editor of *Lux* magazine and deputy editor of *Research* magazine.



Writing a hit tune

Can data help songwriters – or even replace them?

Many would say that there is something about music that is indefinable – magical, even – meaning that it cannot, or perhaps just should not, be submitted to cold statistical analysis.

Data scientists are working hard to prove this theory wrong. London-based start-up Jukedeck has trained an artificial intelligence program to compose and produce music, using neural networks and machine learning to understand composition. The results (available to hear at soundcloud.com/jukedeck) are pretty convincing, and the company hopes it will take off as a way for companies and individuals to create royalty-free background music for videos and games.

But companies like Jukedeck now face competition from Google, which in 2016 announced a project called Magenta (magenta.tensorflow.org), aiming to apply machine intelligence to “music and art generation”. It published a 90-second piece of piano music, written by a trained neural network.

A survey of music generation systems, published last year by researchers in Singapore, the UK and the USA, identified numerous methods being used to automatically generate music, including Markov models, neural networks and evolutionary optimisation algorithms.²

A key challenge, they found, was creating music with “long-term structure” – which often manifests as recurring themes, leitmotifs and other patterns – and capturing “higher-level content such as emotion and tension”. Another challenge is getting computers to “play” a piece of music using synthesised instruments in a way that sounds real and satisfying.

Crafting the right words

Can songwriters use data to hone their lyrics?

A pro tip for heavy metal songsmiths: try to include the word “burns” in your song, but avoid the word “particularly”.

Those, according to data scientist Iain Barr, are the words in the English language with the highest and lowest scores for “metalness”. To create this metric, Barr analysed the lyrics of more than 200 000 songs from more than 7000 bands, looking at the frequency with which words appear (bit.ly/2BjgZbp).

Barr went on to use Markov chains and neural networks to generate metal lyrics automatically (try it out for yourself at deepmetal.io). While this is an entertaining exploration of the tropes of the genre, the results suggest that computers are still some way from replacing human lyricists. Here’s something a machine prepared earlier: “I want to find the cross the dark and here the path, take it all the flame of the time / the sea the struggle the life / we light sacred from the light and leave the black the wind”.

One genre where lyrical dexterity is particularly prized is rap – and data journalist Matt Daniels last year tried to settle the question of which rapper has the largest vocabulary (bit.ly/2BnJvs). Daniels analysed the first 35 000 words of the oeuvres of 85 hip hop artists (thus putting new acts on a level footing with established stars). The winner, with 7392 unique words, was Aesop Rock, while GZA of the Wu-Tang Clan took second place with 6426. Global superstar Jay-Z only managed 4506.

Daniels’s analysis suggests verbosity is no shortcut to commercial success: Aesop Rock remains relatively obscure, while Jay-Z has sold more than 100 million records and won 21 Grammys. As Jay-Z puts it in one of his own songs: “I dumbed down for my audience to double my dollars / They criticized me for it, yet they all yell ‘holla’”.





Spotting the stars of the future

Which songs and artists are destined to set the charts on fire?

Once you have written a song, can statistics predict if it will be a hit?

A team at the University of Antwerp thinks so. They mined data from Spotify on audio features of dance songs released between 1985 and 2013, looking at variables such as tempo, time signature and timbre, as well as fuzzier attributes like "energy". Then they used various classifiers to come up with models for predicting whether a given song in the top 30 or 40 would reach the top 10.³ (Try it out by uploading your own song at bit.ly/predictahit.)

It worked "pretty well", says Dorien Herremans, one of the researchers behind the project. When the same model was applied to the top songs of 2015, it gave them all a 65% or higher chance of being a hit. Herremans thinks part of the tool's success may have come from limiting its scope to dance music, meaning it was studying songs with similar attributes (the study also excluded lyrics, a topic that Herremans says is "on my to-do list").

Companies such as New York-based Music Xray already claim to use statistical analysis of audio to spot potential hits. Music Xray also has a service to connect musicians with record labels and companies seeking specific types of music for ads, films and TV shows; possible matches are determined by comparing the "acoustic properties and mathematical patterns" in "seed" songs supplied by companies against those submitted by artists.

Another company that prides itself on its ability to spot hits early is Shazam, whose app allows users to identify any song they hear when out and about, using their mobile phone.

Scott Holechek, senior director of research at Shazam, says: "We've come up with algorithms that help us identify early signs of growth for a given artist, so we've trained our systems to identify when a given track experiences a very unique spike in activity for a given reason, which could be anything from their song airing for the first time on a commercial, or being discovered in an underground scene in the clubs of Germany. That will often catch artists very early in their careers, and predict whether they will experience sustained growth. Many times when we spot an artist, their career will grow and – boom – their music is on the radio."

For those who fear that overreliance on data will make music samey and bland, tools like Shazam offer hope that it might do the opposite, says Brian Moon, assistant professor of music at the University of Arizona. "Pop music has always got somebody saying, 'This is the end of everything,'" says Moon. "But the idea that the industry homogenises in order to maximise chances of success doesn't play out for pop music because there's always going to be that kid in a basement somewhere making a sound and getting a small audience that can amplify and become viral on the internet."

Finding your next favourite band

If you liked that, you'll love...

No matter how individual and rarefied your taste in music, digital music providers are making it their business to work out what you like so they can keep you listening – all by analysing data.

US online radio service Pandora employs a team of musicologists to run its Music Genome Project, which uses an algorithm to identify up to 450 attributes (known as "genes") for every song. A matching algorithm then finds similarities



between songs, allowing Pandora to choose music that a particular listener may enjoy.

Streaming giant Spotify uses deep learning and neural networks to come up with personalised playlists, drawing on technology from a company called The Echo Nest, which Spotify acquired in 2014 after it was spun off from the MIT Media Lab.

Spotify's recommendations are based partly on collaborative filtering – looking for similarities between what users listen to, to find new things they are likely to enjoy. But it also organises songs into around 1500 genres and subgenres (things like "Russian alternative", "Aussietronica" and "dirty Texas rap") and analyses the acoustic properties of songs for similarities – which is particularly useful for songs that have not been played much, and for new users who have not yet listened to many songs.

The data-based playlists have been a huge hit: within a year of Spotify launching them, listeners had used them to stream five billion tracks (bit.ly/2AkhKNv).

Tracking buzz

How data is taking us beyond the charts

Before everything went digital, data on music consumption was based on weekly store sales and radio airplay. Now, streaming services and other tools such as Shazam provide a granular, real-time view of what audiences are listening to. And the results are not always what we expect.

The UK's official chart has incorporated streaming since 2014, but that change is now being blamed for making the countdown sluggish and dull. Initially the charts used a formula treating 100 streams as equivalent to one sale, but this had to be increased to 150 as streams quickly became dominant. In March last year, all 16 tracks from Ed Sheeran's album *Divide* were in the UK top 20 at once, prompting the Official Charts Company to limit the number of songs from any given artist to just three, and to kick songs out of the charts that had lingered past their peak.

Meanwhile, data is coming from new sources to give us an even more vivid picture of how music is discovered, shared and enjoyed. With 20 million people a day using Shazam to identify songs, the data operation is "massive", says Scott Holecek. It tracks what songs people try to identify, where they are, and what they do next (for instance, listen to the song, share it with a friend, or add it to a playlist).

The abundance of information might even disabuse us of some of the things we thought we knew about music, says Brian Moon of the University of Arizona. Data, he says, can help us confront established notions of what makes music valuable or culturally significant, and even what genre an artist belongs in.

For example, in the age of streaming, hip hop and R&B have together emerged as the biggest music genre in the USA, despite radio stations, chart compilers and award ceremonies often treating it as a separate, lesser category to mainstream pop and rock (bit.ly/2AkQVbV).



A. Parva Bigstock.com

Data is coming from new sources to give us an even more vivid picture of how music is discovered, shared and enjoyed

Similarly, the most played songs of particular artists or genres on streaming platforms "often fly in the face of canonical understanding", says Moon. For instance, Louis Armstrong is revered in the jazz world for his pioneering trumpet style. But his most played songs on Spotify – including *What a Wonderful World* – all have him on vocals, not trumpet. Similarly, the contemporary musician Ludovico Einaudi is categorised by Spotify under classical, where he massively outperforms the likes of Mozart, Beethoven and Brahms. Moon says: "The notion of the canon is being uprooted by individual taste, which may or may not conform to what we think of as important." ■

References

1. Léveillé Gauvin, H. (2017) Drawing listener attention in popular music: Testing five musical features arising from the theory of attention economy. *Musicae Scientiae*, to appear. doi: 10.1177/1029864917698010.
2. Herremans, D., Chuan, C.-H. and Chew, E. (2017) A functional taxonomy of music generation systems. *ACM Computing Surveys*, 50(5), 69.
3. Herremans, D., Martens, D. and Sørensen, K. (2014) Dance hit song prediction. *Journal of New Music Research*, 43, 291–302.