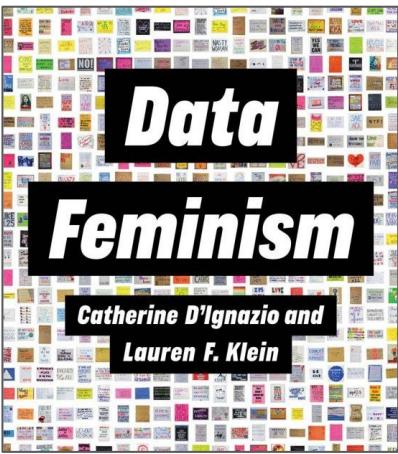


## **Computational Journalism**

Tuesday and Thursday Evenings, Spring of 2023  
Instructors, Mark Hansen and Wayne Lee

# **JOURNALIST'S BRIEFING: QUANTIFICATION**





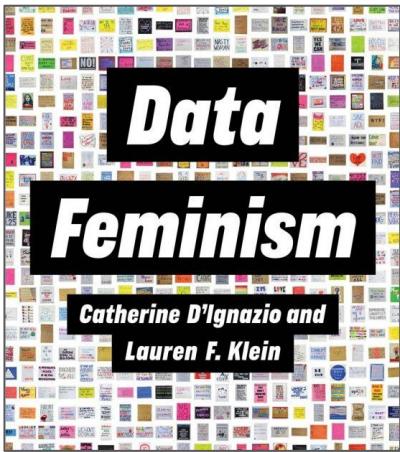
## Analyze and Expose

This is one of the challenges of using data about people as an input into a system: the data are never “raw.” Data are always the product of unequal social relations—relations affected by centuries of history. As computer scientist Ben Green states, “Although most people talk about machine learning’s ability to predict the future, what it really does is predict the past” ... But machine learning algorithms don’t just predict the past; they also reflect current social inequities.

The field of journalism has long prided itself on “speaking truth to power.” But today, the location of that power has shifted from people and corporations to the datasets and models that they create and employ. These datasets and models require new methods of interrogation

Table 2.1: From data ethics to data justice

Concepts That Secure Power	Concepts That Challenge Power
Because they locate the source of the problem in individuals or technical systems	Because they acknowledge structural power differentials and work toward dismantling them
Ethics	Justice
Bias	Oppression
Fairness	Equity
Accountability	Co-liberation
Transparency	Reflexivity
Understanding algorithms	Understanding history, culture, and context



## Action

Taking action can itself take many forms, and in this chapter we offer four starting points:

1. Collect: Compiling counterdata—in the face of missing data or institutional neglect—offers a powerful starting point...
2. Analyze: Challenging power often requires demonstrating inequitable outcomes across groups, and new computational methods are being developed to audit opaque algorithms and hold institutions accountable.
3. Imagine: We cannot only focus on inequitable outcomes, because then we will never get to the root cause of injustice. In order to truly dismantle power, we have to imagine our end point not as “fairness,” but as co-liberation.
4. Teach: The identities of data scientists matter, so how might we engage and empower newcomers to the field in order to shift the demographics and cultivate the next generation of data feminists?



## Ubiquity

In general, the trick is to question every apparently natural easiness in the world around us and look for the work involved in making it easy.

## Materiality and Texture

A good operations researcher could describe how and whether things would work together, often purposefully blurring the physical and conventional boundaries in making the analysis. But what is missing is a sense of the landscape of work as experienced by those within it. It gives no sense of something as important as the texture of an organization: Is it smooth or rough? Bare or knotty? What is needed is a sense of the topography of all of the arrangements: Are they colliding, coextensive, gappy, or orthogonal One way to get at these questions is to take quite literally the kinds of metaphors that people use when describing their experience of organizations, bureaucracies and information systems...

## The Past

There is no way of ever getting access to the past except through classification systems of one sort or another... Take the apparently unproblematic statement: "In 1640, the English revolution occurred; this led to a twenty-year period in which the English had no monarchy." The classifications involved here, all problematic, include...

## Politics

There are two processes associated with these politics: arriving at categories and standards, and, along the way, deciding what will be visible or invisible within the system.

Playground - OpenAI API

beta.openai.com/playground

Overview Documentation Examples Playground

Help Columbia University

Playground

Load a preset... Save View code Share ...

Maximum length 797

Stop sequences Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

Inject start text

Inject restart text

Show probabilities

Here is a paragraph summarizing the latest monkeypox statistics for New York state:  
As of July 13 2022, a total of 359 confirmed orthopoxvirus/monkeypox cases - a designation established by the Centers for Disease Control and Prevention (CDC) - have been identified with 336 in New York City, 12 in Westchester County, 4 in Suffolk County, 2 in Nassau County, 1 in Sullivan County, 1 in Chemung County, 1 in Rockland County, 1 in Erie County, and 1 in St. Lawrence County.

Create a CSV where each row contains information on a place, recording bot the place and the case count.  
Place, Case Count

Submit

140 Off

Playground - OpenAI API

beta.openai.com/playground

Overview Documentation Examples Playground

Help Columbia University

Playground

Load a preset... Save View code Share ...

Maximum length 797

Stop sequences Enter sequence and press Tab

Top P 1

Frequency penalty 0

Presence penalty 0

Best of 1

Inject start text

Inject restart text

Show probabilities

Here is a paragraph summarizing the latest monkeypox statistics for New York state:  
As of July 13 2022, a total of 359 confirmed orthopoxvirus/monkeypox cases - a designation established by the Centers for Disease Control and Prevention (CDC) - have been identified with 336 in New York City, 12 in Westchester County, 4 in Suffolk County, 2 in Nassau County, 1 in Sullivan County, 1 in Chemung County, 1 in Rockland County, 1 in Erie County, and 1 in St. Lawrence County.

Create a CSV where each row contains information on a place, recording bot the place and the case count.  
Place, Case Count  
New York City, 336  
Westchester County, 12  
Suffolk County, 4  
Nassau County, 2  
Sullivan County, 1  
Chemung County, 1  
Rockland County, 1  
Erie County, 1  
St. Lawrence County, 1

Submit

140 Off

Y BuzzFeed to Use ChatGPT Cre x +

news.ycombinator.com/item?id=34534921

Hacker News new | past | comments | ask | show | jobs | submit login

▲ BuzzFeed to Use ChatGPT Creator OpenAI to Help Create Some of Its Content (wsj.com)  
7 points by jackallis 46 minutes ago | hide | past | favorite | 3 comments

[add comment](#)

▲ add-sub-mul-div 3 minutes ago | next [-]  
This is really the dangerous part of where we're going with this. Our culture will end up with a net increase of Buzzfeed type content and decrease of substantive content.  
Substantive content is going to become the new vinyl. Sure, it will still exist. As a niche. You won't be able to find much of it.  
It's like the indignity of Google replacing customer service with algorithms knowing that it's less effective, but not caring because it's cheaper and it scales. But now it's about to be applied to our entire society.  
For every artist using generative AI to find a new way to be creative there will be 50 companies using it to give us a worse experience that saves them money, and with it becoming the new norm they don't have to worry about losing business from it.  
[reply](#)

▲ insane\_dreamer 6 minutes ago | prev | next [-]  
ChatGPT is a boon to click-bait content generators like BuzzFeed. There's a ton of low-quality media content out there that's about to become fully automated. A lot of it is of course, like sports articles, and that content is probably going to improve with ChatGPT and its successors.  
[reply](#)

▲ wetpaws 10 minutes ago | prev [-]  
I was waiting for it eagerly and here we are. Dont think much will change for buzzfeed content-wise, it was always pretty indistinguishable from GPT.  
[reply](#)

---

[Guidelines](#) | [FAQ](#) | [Lists](#) | [API](#) | [Security](#) | [Legal](#) | [Apply to YC](#) | [Contact](#)

# NL2INTERFACE: Interactive Visualization Interface Generation from Natural Language Queries

Yiru Chen\*  
Columbia University      Ryan Lit†  
University of Washington      Austin Mac ‡  
University of California, Santa Barbara      Tianbao Xie§  
University of Hong Kong      Tao Yu‡  
University of Hong Kong      Eugene Wu†  
Columbia University

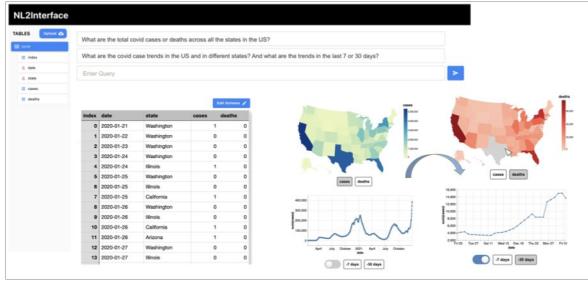


Figure 1: This demonstrates the NL2INTERFACE user interface. A user uploads the COVID-19 dataset using the left panel and the table contents are displayed in the bottom left of the interface. Then, this user types two natural language queries in the top textbox. Our NL2INTERFACE will return an interactive interface consisting of two visualizations, one toggle, two button sets, and a click interaction over the map visualization. The user can interact with the interface to study the covid death trend within the last 30 days in Texas by clicking the state of “Texas” on the map, clicking the “deaths” button, and then clicking the “-30 days” button. The interface after these interactions is shown on the right.

## ABSTRACT

We develop NL2INTERFACE to explore the potential of generating usable interactive multi-visualization interfaces from natural language queries. With NL2INTERFACE, users can directly write natural language queries to automatically generate a fully interactive multi-visualization interface without any extra effort of learning a tool or programming language. Furthermore, users can interact with the interfaces to easily transform data and quickly see results in the visualizations.

## 1 INTRODUCTION

Interactive visualization interfaces (or simply *interfaces*) play a critical role in nearly every stage of data management—including data

cleaning [32], wrangling [15], modeling [12], exploration [4][19], and visualization [6][14]. A specific domain-oriented [17] is a series of data transformations commonly performed with SQL that gradually converts data into an ideal format. The interface design primarily focuses on visualizations and providing interactions. Interactions including widgets and visualization interactions such as brushing and clicking help users easily specify the desired transformations and let users quickly see results in the visualizations.

Such interfaces require considerable expertise and trial-and-error to design and implement. This is because end users have to determine the data management task, then choose visualizations, interactions, and layout to express the underlying data transformations [18].

Existing works [1][7][11][13][20][29] allow users to express data transformations in natural language queries and return data visualizations. Using natural language queries for data transformation helps users express the transformation directly without having to learn a tool or a programming language. This is desirable for normal users, especially non-experts. However, existing works only create individual visualizations instead of generating multi-visualization interfaces. Unlike an individual visualization, creating an interactive multi-visualization interface requires much more consideration on widgets choices, layout design, number of visualizations, whether or not to allow interactions on visualizations, types of interactions, etc.

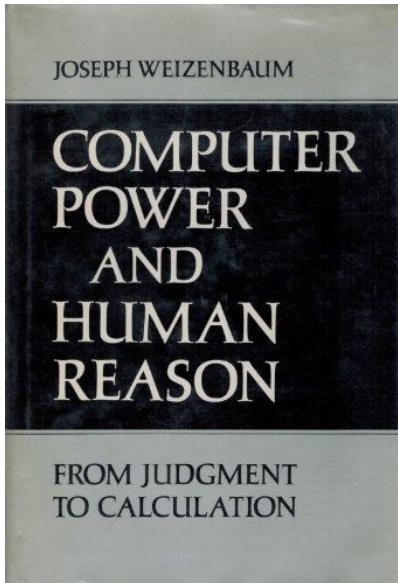
\*e-mail: yiru.chen@columbia.edu  
†e-mail: lansong@cs.washington.edu  
‡e-mail: austinnmac@cs.ubc.ca  
§e-mail: tianbao@cs.hku.hk  
‡e-mail: tyu@cs.hku.hk  
\*e-mail: ewu@cs.columbia.edu

## **Lived experience into data**

*Sometimes when my children were still little, my wife and I would stand over them as they lay sleeping in their beds. We spoke to each other in silence, rehearsing a scene as old as mankind itself. It is as Ionesco told his journal: "Not everything is unsayable in words, only the living truth."*

## Lived experience into data

*Sometimes when my children were still little, my wife and I would stand over them as they lay sleeping in their beds. We spoke to each other in silence, rehearsing a scene as old as mankind itself. It is as Ionesco told his journal: "Not everything is unsayable in words, only the living truth."*



Welcome to

EEEEEE	LL	III	ZZZZZ	AAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LL	II	ZZZ	AAAAAAA
EE	LL	II	ZZ	AA AA
EEEEEE	LLLLL	III	ZZZZZ	AA AA

Eliza is a mock Rogerian psychotherapist.  
The original program was described by Joseph Weizenbaum in 1966.  
This implementation by Norbert Landsteiner 2005.

ELIZA: Is something troubling you ?  
YOU: Men are all alike.  
ELIZA: What is the connection, do you suppose ?  
YOU: They're always bugging us about something or other.  
ELIZA: Can you think of a specific example ?  
YOU: Well, my boyfriend made me come here.  
ELIZA: Is it important to you that your boyfriend made you come here ?  
YOU: He says I'm depressed much of the time.  
ELIZA: I am sorry to hear that you are depressed.  
YOU: It's true. I am unhappy.  
ELIZA: Can you explain what made you unhappy ?  
YOU:

## THE NIGHT SKY

This is what you see at night, in early July, if you're in Boylan Heights and you look up at the sky . . . if you can get out from under the trees. At the top of the hill, in the middle of Boylan Avenue, we lay on our backs to make this map of the stars above the neighborhood.

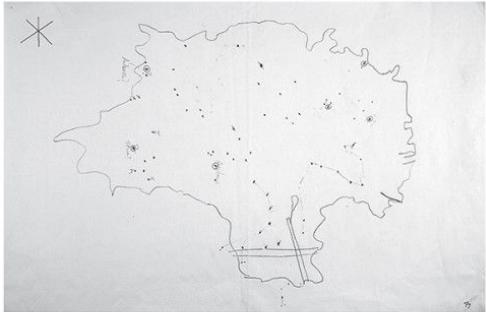
It was about ten o'clock and the asphalt was still warm with the day's heat. We had a star finder, a flashlight to read it with, paper and pencil. We made a sketch of the horizon and roughed in the stars we could see and returned the next day to make dozens of detailed drawings. Afterwards we linked these together for a 360° view and used charts to make sure of our stars. With a shrunk-down copy, we went back to the street at night and fiddled with it until we got it right.

During summer in Boylan Heights, when you look up, you mostly see trees. At the right, where the horizon dips toward the

north, you can see across the cement factory toward downtown and the cylindrical bulk of what was then the Holiday Inn. The white rectangles at top are the lit windows of a house on the east side of Boylan Avenue. The streetlight's on the west side. The mass of foliage to the left leads south down Boylan.

Above? Vega—one of the night sky's brightest stars—is nearly overhead in the Lyre of Orpheus that the Muses placed in the sky after he died. Zeus put the Ursas in the sky. With the glare and summer humidity, we couldn't see Ursa Minor at all, and all we could see of Ursa Major was its tail, our Big Dipper.

Where is Boylan Heights? It's in the United States and North Carolina and Wake County and Raleigh, but first and last it's in the universe. As William Saroyan said, "Birth is into the world, not into a town."

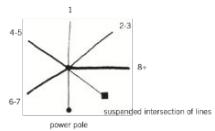


## SQUIRREL HIGHWAYS

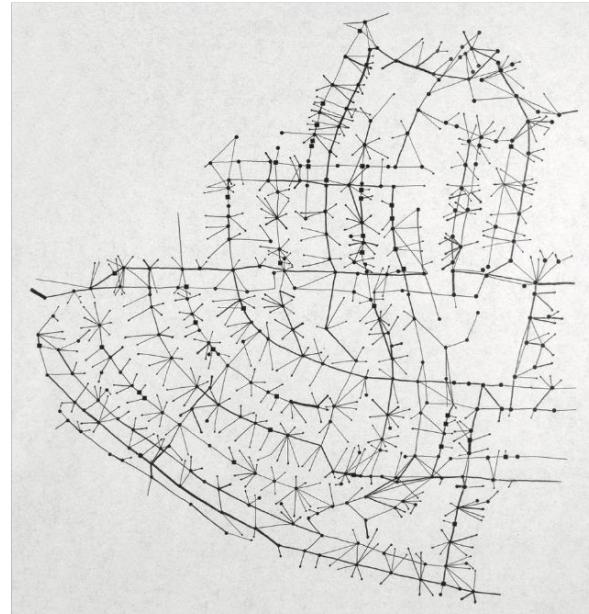
Just as pipes vein the underground, overhead lines innervate the neighborhood, electricity and information flowing out of and back into national and international grids.

Each powerpole has a birthmark, a metal tag that says who owns it—mostly Carolina Power & Light. The other utilities rent space. When all the utility lines are attached to a single pole, they stack up like this naked electrical lines on top, copper phone lines (bundled in black, shielded cable) next, then the cablevision lines, fat, shiny, and silver.

Nervous squirrels, afraid of an attack on the ground, use the phone and television cables as highways wherever the tree canopy's broken. Birds rest on the power lines. "Man and nature, not separate . . ."



36

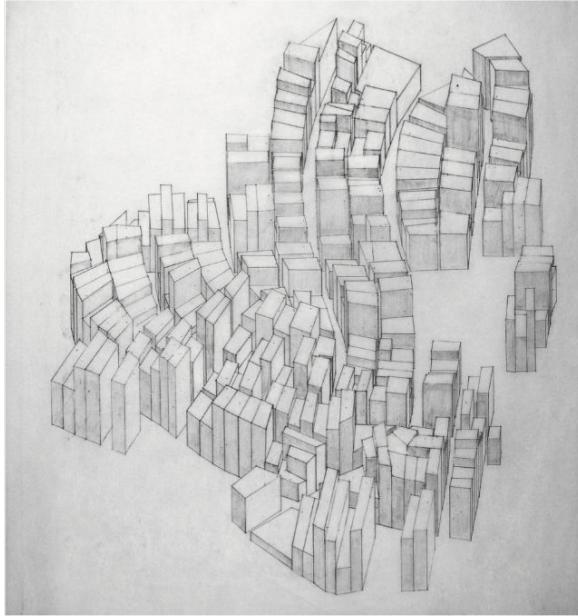
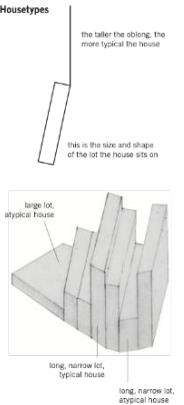


## SHOTGUN, BUNGALOW, MANSION

The characteristic Boylan Heights house is a mutt. Some shotgun, some bungalow, it's a white, wooden, one (or one-and-a-half) story house on a red brick foundation with a front porch and wooden railing. While there are variations and differences between the houses at the top of the hill and those at the bottom, there is also an organic coherence among them.

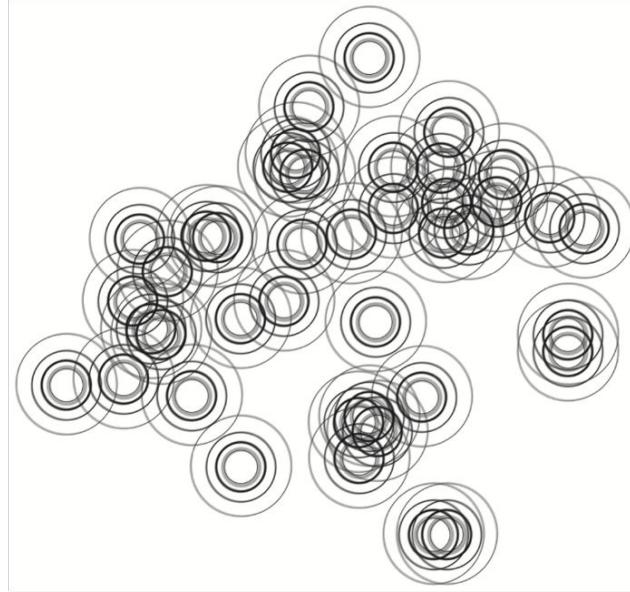
Kelsey and Guild's restrictive covenants endowed the neighborhood with a deep structure in the way the houses *mantle* the hillside: the mansion on top, its finial the final to the whole hill; the stately houses around it, *sold*, but free of the extravagances of the mansion; then, as you begin to slip down the hill, houses with less *heft* until, finally, it's rows of shotgun houses on long narrow lots at the bottom. The subtle modulation between the houses along this gradient and their quiet conversation don't eliminate the distance between them, but do allow you to imagine bridging it.

The map displays the typical Boylan Heights house which we determined by analyzing a mass of data we collected on every house in the neighborhood. The tighter a house hews to this norm, the taller the oblong (whose plan indicates the size and shape of the lot the house sits on). It's easy to see that large lots precipitated distinctive houses that deviated far from the norm. The shallowness of the oblong on the mansion site marks the distance between its architectured form—designed by the English architect William Percival—and those of the tall oblongs lining South Street innocent even of an architect's passing glance.



## WIND CHIMES

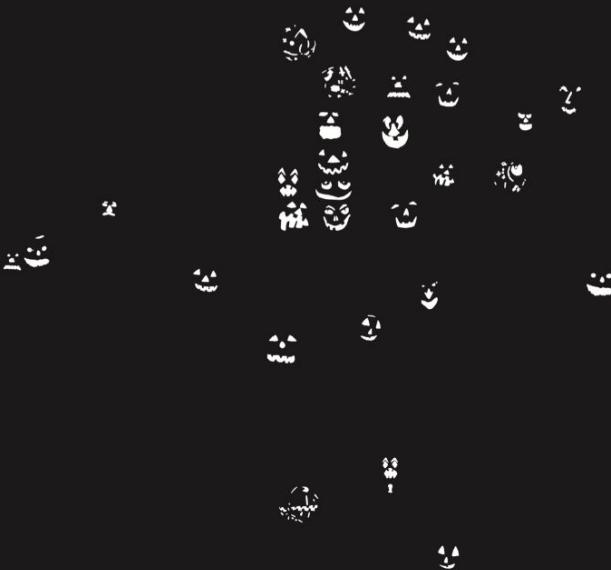
When we did the house types survey, we also paid attention to the presence of wind chimes. They were all over—bamboo, glass, shell, metal tubes. Depending on where you stood, the force of the wind, and the time of day, you could hear several chiming, turning the neighborhood into a carillon.



## JACK-O'-LANTERNS



I rode through the neighborhood on my bicycle—it was 1982—and took pictures of all the jack-o'-lanterns. On the map, there's a jack-o'-lantern at every address where there was one or more pumpkins on the porch, and most of those porches were at addresses that were frequently mentioned in the newsletter.



Miley Cyrus - Flowers (Official) +

youtube.com/watch?v=G7KNmW9a75Y

YouTube Search M

A screenshot of a YouTube video player window. The video frame shows a close-up of Miley Cyrus with blonde hair and sunglasses, looking thoughtfully at the camera. She is wearing a dark jacket. The background is a scenic view of a city or town nestled in a valley under a clear sky. A lyrics overlay at the bottom of the frame reads "♪ I can buy myself flowers ♪". In the bottom right corner of the video frame, the VEVO logo is visible. Below the video player, there is a control bar with a play button, volume controls, and a progress bar indicating the video is at 0:34 of 3:21. To the right of the progress bar are various video controls: a pause button, closed captioning (CC), settings (HD), and other sharing options.

#1 on Trending for music

**Miley Cyrus - Flowers (Official Video)**

Miley Cyrus 17M subscribers [Subscribe](#)

112M views 13 days ago  
Official Video for "Flowers" by Miley Cyrus

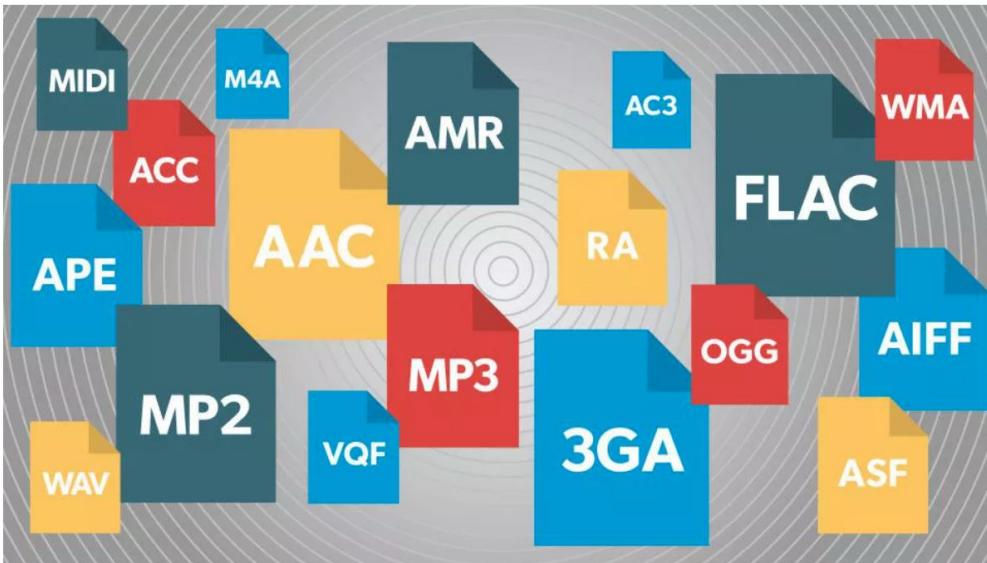
Listen to & Download "Flowers" out now: <https://mileycyrus.lnk.to/Flowers> Show more

# MP3, AAC, WAV, FLAC: all the audio file formats explained

By [Becky Scarrott](#) published July 24, 2022

Do you know your OGG from your ALAC?

[!\[\]\(1ac7c971e7df5bf204fbb84fd617a50a\_img.jpg\)](#) [!\[\]\(9b29a11d4b428981cb461cf79c956bed\_img.jpg\)](#) [!\[\]\(f4f17a61bbc5b86d96dc143d5a21dfe3\_img.jpg\)](#) [!\[\]\(3f2912468bfcc46ea3f149d193f211ca\_img.jpg\)](#) [!\[\]\(7293cf0f068825e901e7fa8e2402051c\_img.jpg\)](#) [Comments \(10\)](#)



## Lyrics

We were good, we were gold  
Kinda dream that can't be sold  
We were right 'til we weren't  
Built a home and watched it burn

Mm, I didn't wanna leave you  
I didn't wanna lie  
Started to cry but then remembered I

I can buy myself flowers  
Write my name in the sand  
Talk to myself for hours  
Say things you don't understand  
I can take myself dancing  
And I can hold my own hand  
Yeah, I can love me better than you can

Can love me better  
I can love me better, baby  
Can love me better  
I can love me better, baby

Paint my nails, cherry red  
Match the roses that you left  
No remorse, no regret  
I forgive every word you said

Ooh, I did not wanna leave you, baby  
I didn't wanna fight  
Started to cry but then remembered I

I can buy myself flowers  
Write my name in the sand  
Talk to myself for hours, yeah  
Say things you don't understand  
I can take myself dancing, yeah  
I can hold my own hand  
Yeah, I can love me better than you can

Can love me better  
I can love me better, baby  
Can love me better  
I can love me better, baby  
Can love me better  
I can love me better, baby  
Can love me better  
Oh, I

I did not wanna leave you  
I didn't wanna fight  
Started to cry but then remembered I

I can buy myself flowers (oh)  
Write my name in the sand (mmh)  
Talk to myself for hours (yeah)  
Say things you don't understand  
I can take myself dancing, yeah  
I can hold my own hand  
Yeah, I can love me better than  
Yeah, I can love me better than you can

Can love me better  
I can love me better, baby (oh)  
Can love me better  
I can love me better (than you can), baby  
Can love me better  
I can love me better, baby  
Can love me better

# FLOWERS

Words and Music by MILEY RAY CYRUS,  
MICHAEL POLLACK and GREGORY HEIN

Disco Pop Am Dm G

We were good, — we were gold, — Kind of dream — that can't be sold.  
cher - ry red, — match the ros - es that can't be left.

mp

C Am Dm G

— We were right — 'till we weren't, — built a home — and watched it burn.  
No re - morse, — no re - gret, — I for - give — ev - 'ry word — you said..

C Am Dm

Mmm, I did - n't want to leave — you, I did - n't want to lie, — Start-  
Ooh, I did - n't want to leave, — babe, I did - n't want to fight, —

How did music notation actually begin? + ↻

classicfm.com/discover-music/how-music-notation-began/

Home Radio News Discover Music Composers Artists Lifestyle Win Charity Podcasts Events Dating

## Discover Music

Discover Music Home Periods & Genres Occasions Mood Instruments Music Theory Learning Humour Women in Music

# How did music notation actually begin?

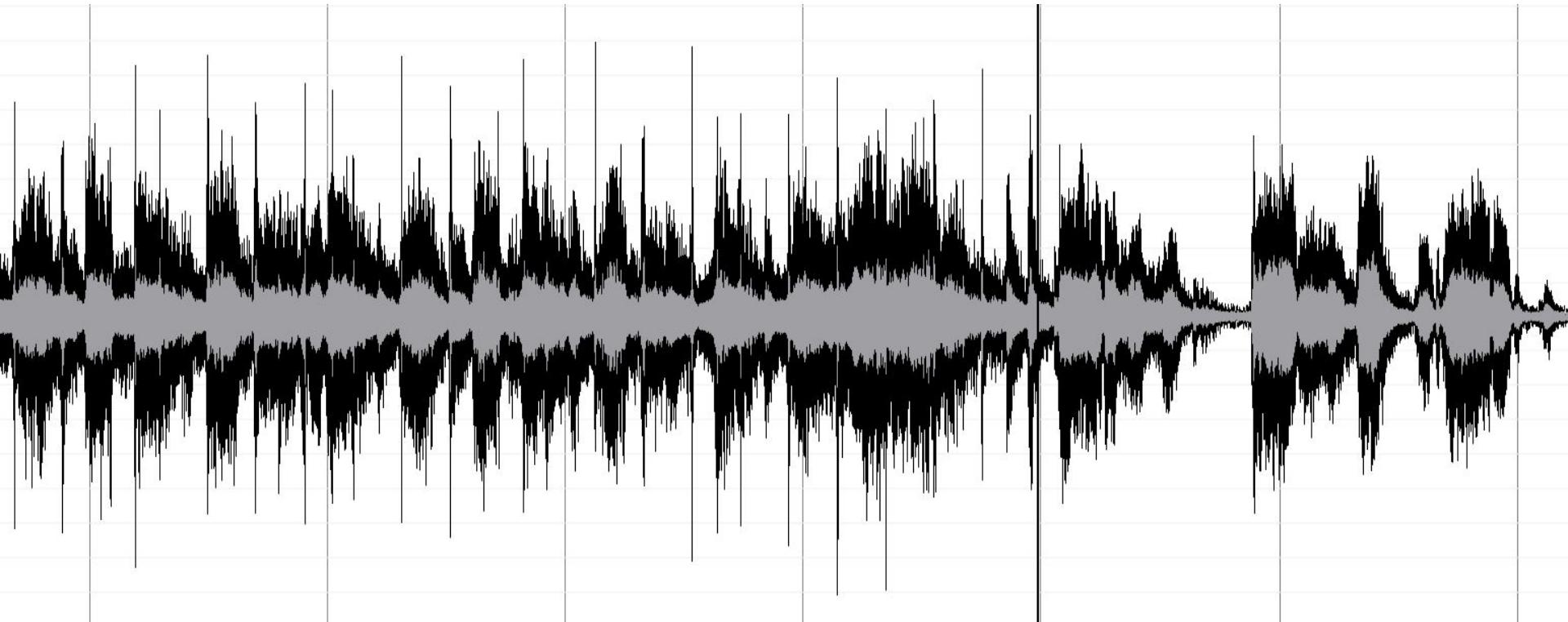
17 March 2018, 19:56 | Updated: 22 March 2018, 10:50

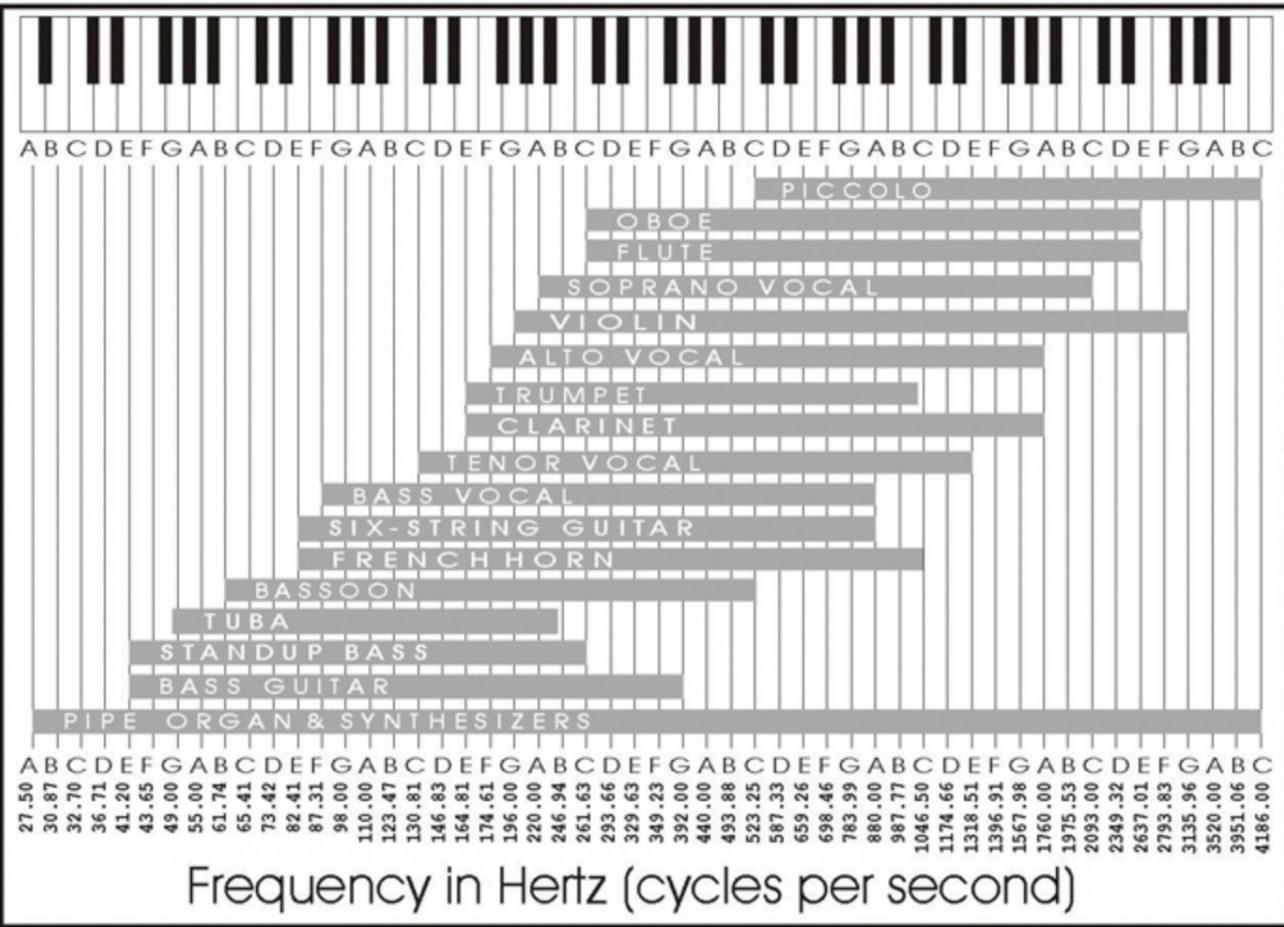
Promoted by Honda Jazz Sport

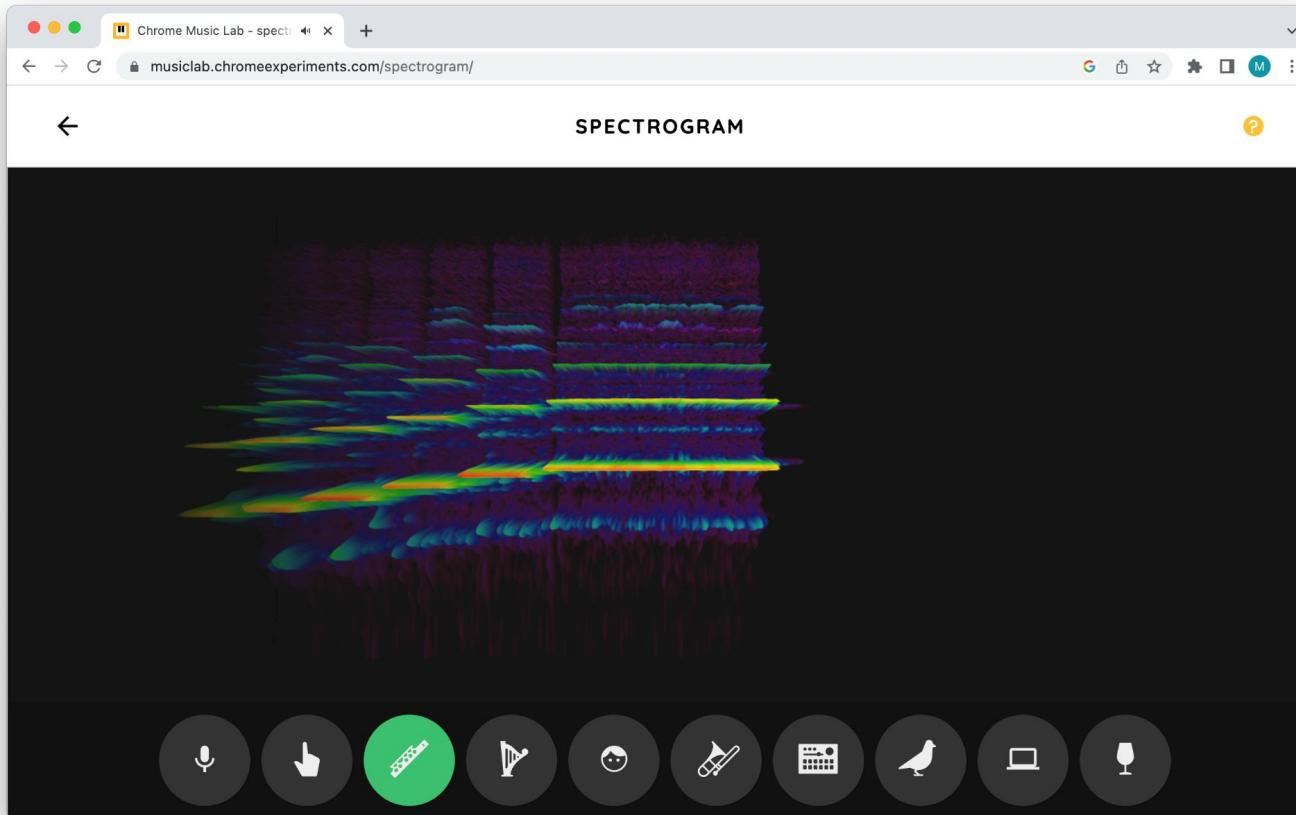
Staves next, then symbols to indicate note duration...

In 650 AD, St Isidore developed a new system of writing music, using a notation called 'neumes'. Vocal chants would be written on parchment with the text, above which neumes would be notated, indicating the contour of the melody.

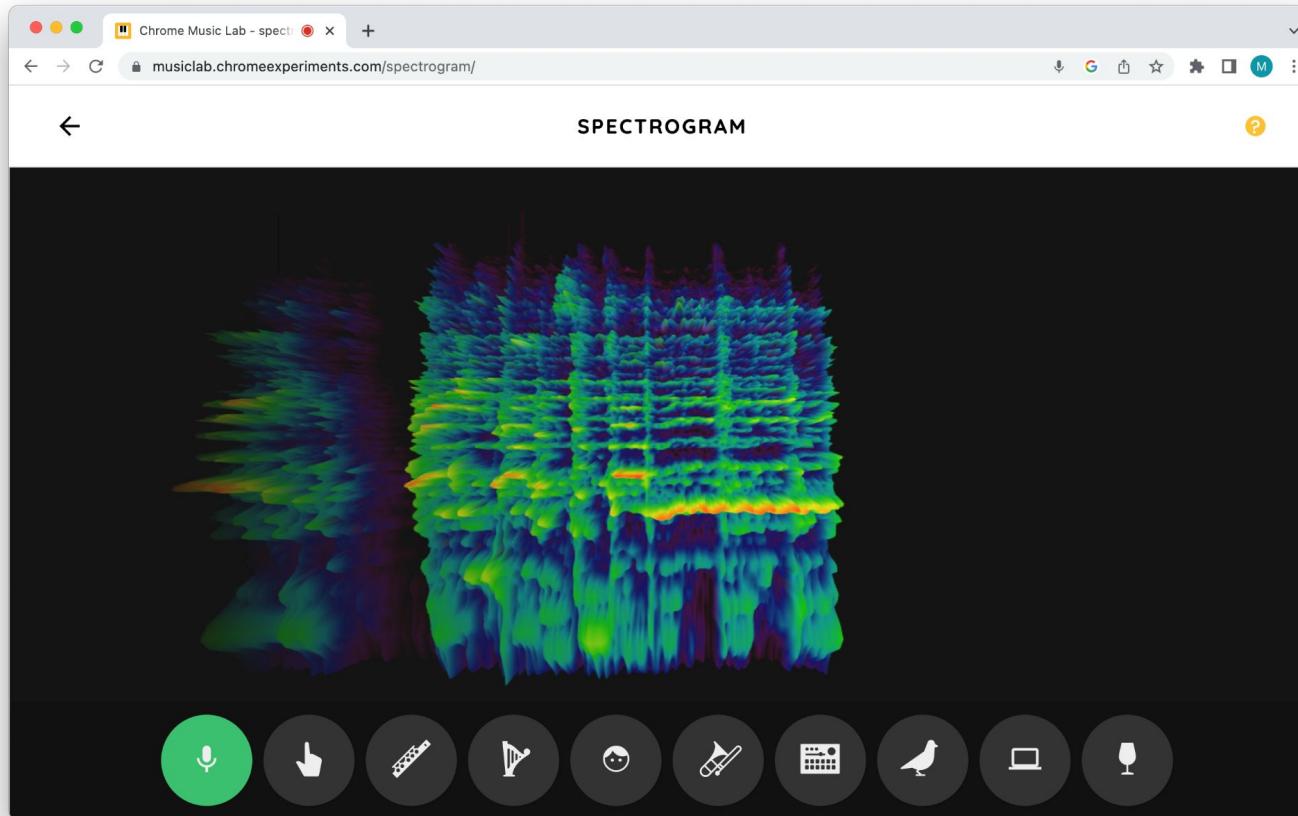
Neumes... had one major flaw: the singers didn't know exactly which note to sing – only whether to sing higher and lower than the last note.







<https://musiclab.chromeexperiments.com/spectrogram/>



<https://musiclab.chromeexperiments.com/spectrogram/>

# An Industrial-Strength Audio Search Algorithm

Avery Li-Chun Wang  
[avery@shazamteam.com](mailto:avery@shazamteam.com)  
Shazam Entertainment, Ltd.

USA:  
2925 Ross Road  
Palo Alto, CA 94303

United Kingdom:  
375 Kensington High Street  
4th Floor Block F  
London W14 8Q

*We have developed and commercially deployed a flexible audio search engine. The algorithm is noise and distortion resistant, computationally efficient, and massively scalable, capable of quickly identifying a short segment of music captured through a cellphone microphone in the presence of foreground voices and other dominant noise, and through voice codec compression, out of a database of over a million tracks. The algorithm uses a combinatorially hashed time-frequency constellation analysis of the audio, yielding unusual properties such as transparency, in which multiple tracks mixed together may each be identified. Furthermore, for applications such as radio monitoring, search times on the order of a few milliseconds per query are attained, even on a massive music database.*

## 1 Introduction

Shazam Entertainment, Ltd., was started in 2000 with the idea of providing a service that could connect people to music by recognizing music in the environment by using their mobile phones to recognize the music directly. The algorithm had to be able to recognize a short audio sample of music that had been broadcast, mixed with heavy ambient noise, subject to reverb and other processing, captured by a little cellphone microphone, subjected to voice codec compression, and network dropouts, all before arriving at our servers. The algorithm also had to perform the recognition quickly over a large database of music with nearly 2M tracks, and furthermore have a low number of false positives while having a high recognition rate.

This was a hard problem, and at the time there were no algorithms known to us that could satisfy all these constraints. We eventually developed our own technique that met all the operational constraints [1].

We have deployed the algorithm to scale in our commercial music recognition service, with over 1.8M tracks in the database. The service is currently live in Germany, Finland, and the UK, with over a half million users, and will soon be available in additional countries in Europe, Asia, and the USA. The user experience is as follows: A user plays a song on their mobile phone and samples 10 to 15 seconds of audio. An identification is performed on the sample at our server, then the track title and artist are sent back to the user via SMS text messaging. The information is also made available on a web site, where the user may register and log in with her mobile phone number and password. At the web site, or on a smart phone, the user may view her tagged track list and buy the CD. The user may also download the ringtone corresponding to the tagged track, if it is available. The user may also send a

30-second clip of the song to a friend. Other services, such as purchasing an MP3 download may become available soon.

A variety of similar consumer services has sprung up recently. Musiwave has deployed a similar mobile-phone music identification service on the Spanish mobile carrier Amena using Philips' robust hashing algorithm [2-4]. Using the algorithm from Relatabile, Neuros has included a sampling feature on their MP3 player which allows a user to collect a 30-second sample from the built-in radio, then later plug into an online server to identify the music [5,6]. Audible Magic uses the Muscle Fish algorithm to offer the Clango service for identifying audio streaming from an internet radio station [7-9].

The Shazam algorithm can be used in many applications besides just music recognition over a mobile phone. Due to the ability to dig deep into noise we can identify music hidden behind a loud voiceover, such as in a radio advert. On the other hand, the algorithm is also very fast and can be used for copyright monitoring at a search speed of over 1000 times realtime, thus enabling a modest server to monitor significantly many media streams. The algorithm is also suitable for content-based cueing and indexing for library and archival uses.

## 2 Basic principle of operation

Each audio file is "fingerprinted," a process in which reproducible hash tokens are extracted. Both "database" and "sample" audio files are subjected to the same analysis. The fingerprints from the unknown sample are matched against a large set of fingerprints derived from the music database. The candidate matches are subsequently evaluated for correctness of match. Some guiding principles for the attributes to use as fingerprints are that they should be temporally localized, translation-invariant, robust, and sufficiently entropic. The temporal locality

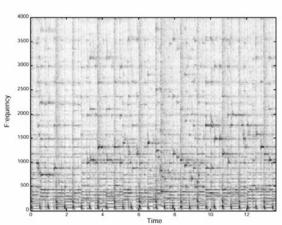


Fig. 1A - Spectrogram

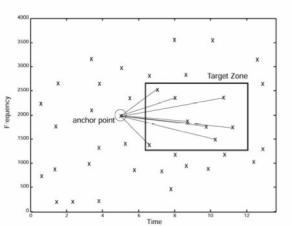


Fig. 1C - Combinatorial Hash Generation

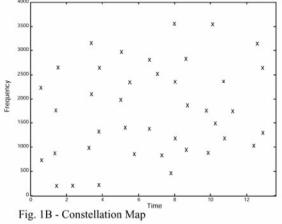


Fig. 1B - Constellation Map

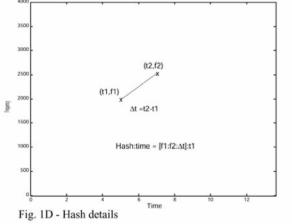


Fig. 1D - Hash details

guideline suggests that each fingerprint hash is calculated using audio samples near a corresponding point in time, so that distorts events do not affect the hash. The translation-invariant aspect means that fingerprint hashes derived from corresponding matching content are reproducible independent of position within an audio file, as long as the temporal locality containing the data from which the hash is computed is contained within the file. This makes sense, as an unknown sample could come from any portion of the original audio track. Robustness means that the hashes generated from an original clean database should be reproducible from a degraded copy of the audio. Furthermore, the fingerprint tokens should have sufficiently high entropy in order to minimize the probability of false token matches at non-corresponding locations between the unknown sample and tracks within the database. Insufficient entropy leads to excessive and spurious matches at non-corresponding locations, requiring more processing power to cull the results, and too much entropy usually leads to fragility and non-reproducibility of fingerprint tokens in the presence of noise and distortion.

There are 3 main components, presented in the next sections.

## 2.1 Robust Constellations

In order to address the problem of robust identification in the presence of highly nonstationary noise and distortion, we experimented with a variety of candidate features that could survive GSM encoding in the presence of noise. We settled on spectrogram peaks, due to their robustness in the presence of noise and approximate linear superposability [1]. A time-frequency point is a candidate peak if it has a higher energy content than all its neighbors in a region centered around the point. Candidate peaks are chosen according to a density criterion in order to assure that the time-frequency strip for the audio file has reasonably uniform coverage. The peaks in each time-frequency locality are also chosen according amplitude, with the justification that the highest amplitude peaks are most likely to survive the distortions listed above.

Thus, a complicated spectrogram, as illustrated in Figure 1A may be reduced to a sparse set of coordinates, as illustrated in Figure 1B. Notice that at this point the amplitude component has been eliminated. This reduction has the advantage of being fairly insensitive to EQ, as

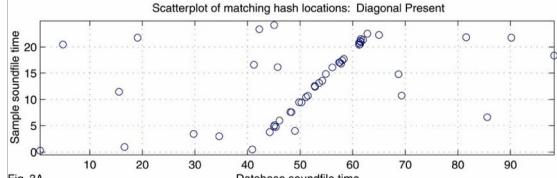
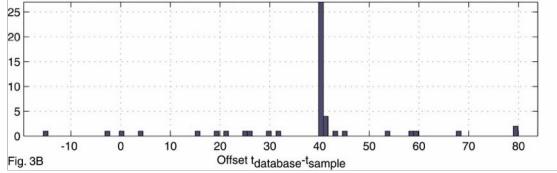


Fig. 3A

Histogram of differences of time offsets: signals match



aggregate 64-bit struct, 32 bits for the hash and 32 bits for the time offset and track ID. To facilitate fast processing, the 64-bit structs are sorted according to hash token value.

speedup is a factor of about  $100000/F^2$  or about 10000, over token searches based on single constellation points.

Note that the combinatorial hashing squares the probability of point survival, i.e. if  $p$  is the probability of a spectrogram peak surviving the journey from the original source material to the captured sample recording, then the probability of a hash from a pair of points surviving is approximately  $p^2$ . This reduction in hash survivability is a tradeoff against the tremendous amount of speedup provided. The reduced probability of individual hash survival is mitigated by the combinatorial generation of a greater number of hashes than original constellation points. For example, if  $F=10$ , then the probability of at least one hash surviving for a given anchor point would be the joint probability of the anchor point and at least one target point in its target zone surviving. If we simplistically assume IID probability  $p$  of survival for all points involved, then the probability of at least one hash surviving per anchor point is  $p*(1-(1-p)^F)$ . For reasonably large values of  $F$ , e.g.  $F>10$ , and reasonable values of  $p$ , e.g.  $p>0.1$ , we have approximately

$$p \approx p^* [1-(1-p)^F]$$

so we are actually not much worse off than before.

We see that by using combinatorial hashing, we have traded off approximately 10 times the storage space for approximately 10000 times improvement in speed, and a small loss in probability of signal detection.

Miley Cyrus' 'Flowers' Debuts

billboard.com/music/chart-beat/miley-cyrus-flowers-number-one-debut-global-charts-1235204125/

billboard

LOGIN SUBSCRIBE

CHARTS MUSIC CULTURE MEDIA BUSINESS

Year-End Charts Chart Beat Honda Music Samsung NXT Hot Trending Songs Global Español

**CHART BEAT**

## 'Flowers' Power: Miley Cyrus Soars In at No. 1 on Both Billboard Global Charts

Plus, TAEYANG's "Vibe," featuring Jimin, of BTS, debuts on Global Excl. U.S. at No. 9.

By Gary Trust +  
01/24/2023

f t p +

ADVERTISEMENT

Waiting for collector.sheknows.com...



Miley Cyrus 'Flowers' Review - X +

esquire.com/entertainment/music/a42489563/miley-cyrus-flowers-lyrics-review/

ENTERTAINMENT STYLE LIFESTYLE SHOPPING NEWS & POLITICS

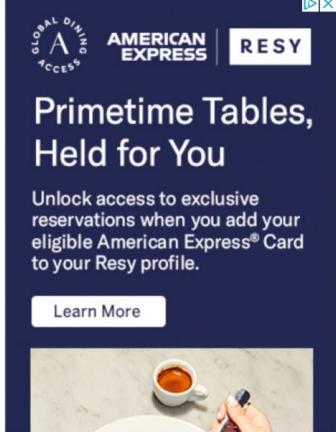
SUBSCRIBE SIGN IN

Entertainment > Music

# With 'Flowers,' Miley Cyrus Ushers In an Assured New Era

The singer's first track off her anticipated 2023 album is all about healing, renewal, and perhaps, a bit of revenge.

By Bria McNeal PUBLISHED: JAN 13, 2023



**T**hen it comes to Miley Cyrus, you should always expect the

12:07

Search

Q miley cyrus flowers

Top Users Videos Sounds Hashtags

reasons why she is the QUEEN

1. she filmed this in the house that her ex husband cheated on her with 14 different women

2. she released this on his birthday

1-15

...should 🤡 😞  
#miley Cyrus #flowers

kveselino... 1.1M

Sponsored

...#miley Cyrus #flowers  
#fyp #viral #tiktok #sin...

UniGracia\_ 162

Top liked AND so subtle hint?

Donuts

This is such a masterpiece 😔

1-18

...#miley Cyrus #flowers  
#fun #lyric

12:08

Q miley cyrus flowers

Top Users Videos Sounds Hashtags

WE WERE GOOD

Not Miley made a banger and recorded it in the house he cheated in

And then basically used the song by Bruno Mars and made it her own 😊

1-14

...#miley Cyrus and Liam Hemsworth

UniGracia\_ 58.0K

1-13

...#miley Cyrus #flowers  
#miley #fyp

raeshorn101 161.4K

Liam Hemsworth dedicated "When I Was Your Man" by Bruno Mars to Miley Cyrus. This is her response.

I should have bought you flowers

Miley Cyrus flowers out

1-13

#miley Cyrusflowers

12:08

Q miley cyrus flowers

Top Users Videos Sounds Hashtags

Pop Tingz @ThePopTingz

The house where the music video for Miley Cyrus' "Flowers" was recorded was previously used by Liam Hemsworth to cheat on Miley with more than 14 women while they were married.

Miley Cyrus & Jennifer Lawrence feud

5d ago

9:09 AM - 1/13/23

1-13 Reactions 2,145

...#miley Cyrus flowers  
#jennifer lawrence #lia...

thezenblon... 3146

...#miley Cyrus flowers  
#miley Cyrus #music vid...

\_brattybr... 2,302

Others searched for

Miley Cyrus

Miley Cyrus Twitching Her Eye

Miley Cyrus Performances

Miley Cyrus Best Covers

Malibu By Miley Cyrus

1-14

#miley Cyrus flowers

Twitter (7) Miley Cyrus (@MileyCyrus) +

twitter.com/MileyCyrus?ref\_src=twsrctwsrc%5Egoogle%7Ctwcamp%5Eserp%7Ctwgr%5Eauthor

Miley Cyrus

12.1K Tweets

**Miley Cyrus**

@MileyCyrus

ENDLESS SUMMER VACATION. THE ALBUM. MARCH 10. FLOWERS OUT NOW.

[mileycyrus.com](http://mileycyrus.com) Joined March 2011

73 Following 46.8M Followers

Followed by Hendrik Hinzel, Jason D. Myles, and 18 others you follow

**Tweets** [Tweets & replies](#) [Media](#) [Likes](#)

Pinned Tweet

**Miley Cyrus**  @MileyCyrus · Jan 12  
I can love me better than you can...

FLOWERS SINGLE + VIDEO OUT NOW.

[mileycyrus.lnk.to/Flowers](http://mileycyrus.lnk.to/Flowers)

Follow Search Twitter

**Taylor Swift**  @taylorswift13

**Rihanna**  @rihanna

**Kim Kardashian**  @KimKardashian

Show more

**What's happening**

Tennis Tournament · LIVE  
**Australian Open 2023**

#InfinityPool  
See the electrifying thriller Infinity Pool in theaters – if you dare  
Promoted by NEON

Sports · Trending  
**#RipLegend**

**Messages**

Have you heard the new  
miley cyrus?

Nope

The economy was resilient in 2022

npr.org/2023/01/26/1151320042/us-economy-recession-inflation-housing-spending-slowdown

**npr** GBH 89.7

SIGN IN NPR SHOP DONATE

NEWS CULTURE MUSIC PODCASTS & SHOWS SEARCH

ECONOMY

## The U.S. economy ended 2022 on a high note. This year is looking different

January 26, 2023 · 8:33 AM ET

SCOTT HORSLEY



On Air Now

HOURLY NEWS LISTEN LIVE PLAYLIST

American Cancer Society Every cancer. Every life.

LEARN MORE

The Commerce Department said Thursday that the nation's gross domestic product — the broadest measure of economic activity — grew at an annual rate of 2.9% between October and December.

That was down slightly from the previous three months, when the economy grew at an annual rate of 3.2%, but a marked improvement from the first half of the year, when GDP shrank...

---

"Everyday stuff is just costing more," says Nikki Moore, a married mother of two in Apopka, Florida. "The four of us going to the movies — nothing super, not 3D — just the four of us for tickets and concessions, that's like \$100, just for a movie night."

---

Mortgage rates have declined in recent weeks, but they remain well above 6%, much higher than they were a year ago. That's weighing on both new home construction and the sale of existing homes.

"The housing sector is in a recession," says Doug Duncan, chief economist for the mortgage giant Fannie Mae. "It will be a difficult year."

Computational Thinking — *Jeannette Wing (DSI)*

**Decomposition** — Break down complex problems into smaller and more manageable parts. This will make even complicated tasks easier to understand. It will also help with “debugging” when applied to actual code-based projects.

**Pattern Recognition** — Finding patterns can simplify tasks because you can use the same code, say, for a different problem. Our journalism work relies heavily on identifying patterns, giving us a critical approach to this step as well.

**Abstraction** — This is at the core of much of our data work. We will create quantifications, data abstractions, of lived experiences and then are able to compare the abstractions. It’s important to keep in mind that abstractions both hide as well as reveal. When it comes to coding, abstraction will let us work with “classes of objects” like abstractions of web pages for scraping, or maps for analyzing spatial information.

**Algorithm Design** — Here we break things down into steps to follow. An algorithm is just a recipe, a series of actions to take.

An alternative

**Computation** — Ultimately, we are learning how to form some computation in the world. This will be through writing code or using some online tool. The first step is often acquiring data, a quantification step, a reduction. Keep that in mind.

**Collaboration** — Computing is a team sport. You might consult people in your newsroom, outside experts, and, increasingly, community members or stakeholders to help evaluate the computation you've performed and the inferences you are drawing. Context is everything. Community members can be both consulted for data but also involved in the modelling process. You should also share the artifacts and methods of your computation, perhaps on a public repository.

**Creativity** — Everything we do in this class will be intensely creative. As we noted earlier, journalists can't completely eliminate their biases, but they can adopt practices that minimize their effects. The same is true for computation. Statistics is shot through with subjective choices. The data you collect, the analysis you perform, the way things are presented. It is all subject to choice, it is subject to your creativity.

Have you heard the new  
miley cyrus?

Nope

Ok then what do you know  
about objects in Python?

Some Far-Right Accounts | Use of N-Word on Twitter | Twitter trolls are back with a vengeance

washingtonpost.com/technology/2022/10/28/musk-twitter-racist-posts/

The Washington Post Democracy Dies in Darkness mark

TECHNOLOGY

## Racist tweets quickly surface after Musk closes Twitter deal

Anonymous Twitter accounts celebrated Musk's takeover and argued it meant the old rules against bigotry no longer applied

By Drew Harwell, Taylor Lorenz and Cat Zakrzewski  
Updated October 28, 2022 at 6:36 p.m. EDT Published October 28, 2022 at 9:27 a.m. EDT



Processing request...

Some Far-Right Accounts | Use of N-Word on Twitter | Twitter trolls are back with a vengeance

businessinsider.com/elon-musk-twitter-takeover-sparked-n-word-use-jum...

INSIDER

HOME > NEWS

## Use of N-word on Twitter jumped by almost 500% after Elon Musk's takeover as trolls test limits on free speech, report says

Joshua Zitser and Sam Tabashri Oct 29, 2022, 8:20 AM



Processing request...

Latest Twitter Problem: Its API Has Been Down And No Info Has Been Provided

PODCAST Techdirt - In Defense Of The Global, Open Internet

SOUNDCLLOUD

Follow Techdirt

Search Techdirt

### The Techdirt Greenhouse

Read the latest posts:

- Writing to Our Latest Greenhouse Panel: The Lessons Learned From SOPA/PIPA
- From The Revolt Against SOPA To The EU's Upload Filters
- Did We Miss Our Best Chance At Regulating The Internet?

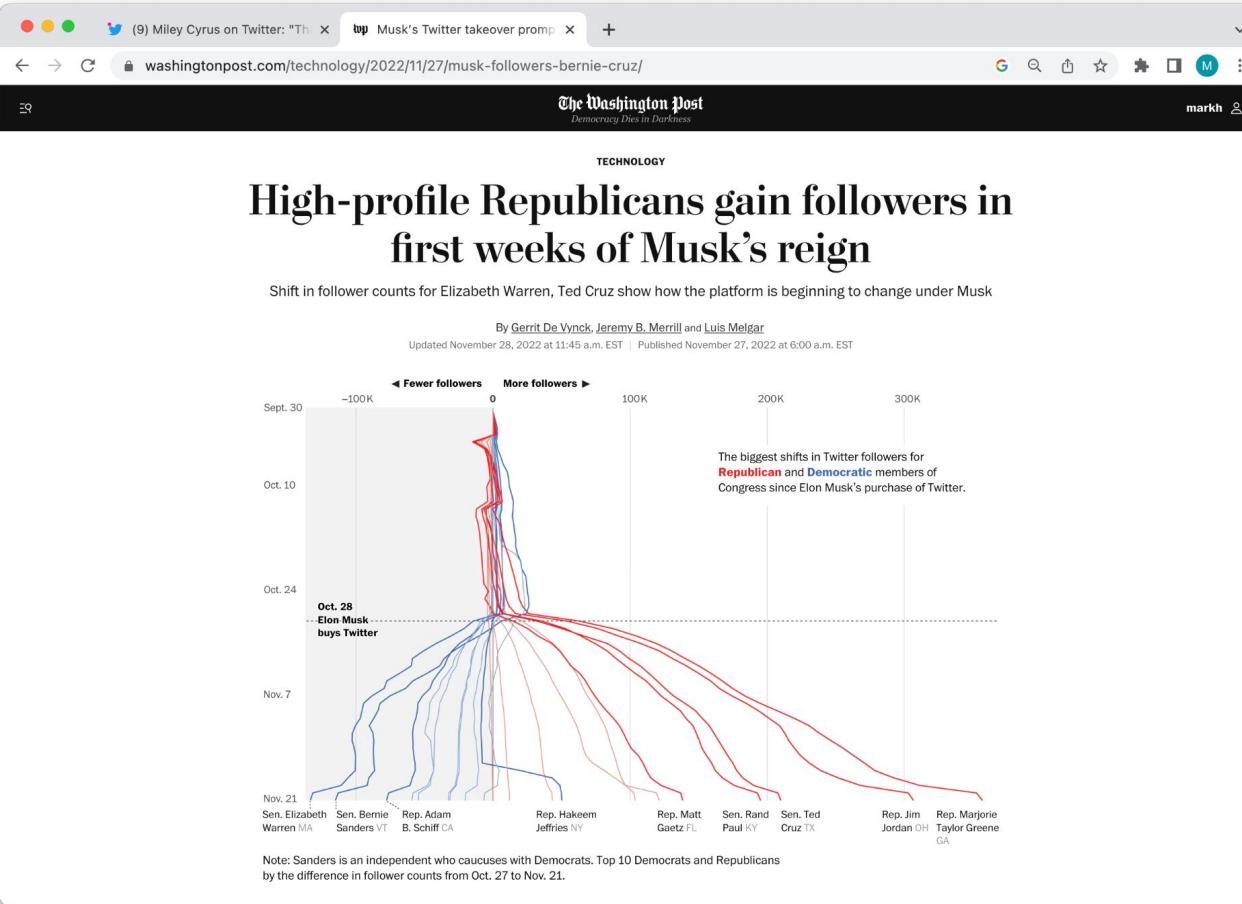
**Read All ▾**

### Trending Posts

- Court Laughs Off QAN Conspiracy Network's Claim It Was 'Censored' By DirecTV
- Pittsburgh PD Decides It Can Ignore Ordinance Banning Traffic Stops For Minor Plate Violations
- UK's Online Safety Bill Gets Ridiculous: Includes Jail Time For Tech CEOs

Techdirt Deals





Republican senators gained almost 160,000 Twitter followers after Elon Musk's deal

Graphic detail | Daily chart

# Republican senators gained almost 160,000 Twitter followers after Elon Musk's deal

News of the takeover could be shifting the platform's demographics

Apr 27th 2022

Share

You win some, you lose some

United States senators, Twitter followers  
% change on previous day, 2022

— Republicans  
— Democrats  
— Independents  
— Average  
--- Individual accounts

Twitter accepts Elon Musk's offer

Marsha Blackburn 3  
Ted Cruz 1  
Bernie Sanders -0.5  
Jon Ossoff -0.5

18 19 20 21 22 23 24 25 26 April

Sources: Twitter; SocialBlade; *The Economist*

Date	Republicans (%)	Democrats (%)	Independents (%)	Average (%)	Individual accounts (%)
Apr 18	0	0	0	0	0
Apr 19	0	0	0	0	0
Apr 20	0	0	0	0	0
Apr 21	0	0	0	0	0
Apr 22	0	0	0	0	0
Apr 23	0	0	0	0	0
Apr 24	0	0	0	0	0
Apr 25	0	0	0	0	0
Apr 26	3	-0.5	-0.5	0	0
Apr 27	0	0	0	0	0

Miley Cyrus's Twitter Stats Summary

[socialblade.com/twitter/user/mileycyrus](https://socialblade.com/twitter/user/mileycyrus)



**Miley Cyrus @MileyCyrus**

FOLLOWERS: 46,892,039 FOLLOWING: 73 LIKES: 1,379 TWEETS: 12,166 USER CREATED: Mar 18th, 2011

User Summary Future Projections Detailed Statistics Live Follower Count

**A+** TOTAL GRADE

**34th** FOLLOWERS RANK **8,220,451st** FOLLOWING RANK **2,069,017th** TWEETS RANK

**155.611K** ▲1,041.6% FOLLOWERS FOR THE LAST 30 DAYS

**3** ▲200.0% FOLLOWING FOR THE LAST 30 DAYS

**42** ▲425.0% TWEETS FOR THE LAST 30 DAYS

FOLLOWERS GAINED (WEEKLY) FOLLOWING GAINED (WEEKLY) TWEETS POSTED (WEEKLY)

Success Academy Careers  
Hiring 2022 Grads to Teach

APPLY NOW

TWITTER STATUS DATE

 Save up to 40% with the CRM Starter Pack.

Get the tools and support you need to succeed now.

 Learn more

Miley Cyrus's Twitter Stats Summary

[socialblade.com/twitter/user/mileycyrus](https://socialblade.com/twitter/user/mileycyrus)

DATE	FOLLOWERS	FOLLOWING	TWEETS
2023-01-13 Fri	—	46,804,672	— 12,169
2023-01-14 Sat	+12,175	46,816,847	— 12,165
2023-01-15 Sun	+3,588	46,820,435	— 12,165
2023-01-16 Mon	+11,388	46,831,823	— 12,165
2023-01-17 Tue	+2,361	46,834,184	— 12,165
2023-01-18 Wed	+12,650	46,846,834	+1 12,165
2023-01-19 Thu	+2,393	46,849,227	+1 12,165
2023-01-20 Fri	+16,474	46,865,701	— 12,163
2023-01-21 Sat	+3,445	46,869,146	— 12,163
2023-01-22 Sun	+8,647	46,877,793	+1 12,164
2023-01-23 Mon	+2,691	46,880,484	— 12,164
2023-01-24 Tue	+2,760	46,883,244	+1 12,165
2023-01-25 Wed	+5,656	46,888,900	+1 12,166
2023-01-26 Thu	+3,139	46,892,039 <small>● LIVE</small>	— 12,166
Daily Averages	+5,187	—	+1
Last 30 days	+155,611	+3	+42

Share on Facebook Tweet This

**TWITTER ANALYTICAL HISTORY FOR MILEYCYRUS**

 Save up to 40% with the CRM Starter Pack.

Get the tools and support you need to succeed now.

 Learn more

Republican senators gained almost 160,000 Twitter followers after Elon Musk's deal

News of the takeover could be shifting the platform's demographics

Apr 27th 2022

Share

You win some, you lose some

United States senators, Twitter followers  
% change on previous day, 2022

— Republicans  
— Democrats  
— Independents  
— Average  
--- Individual accounts

Twitter accepts Elon Musk's offer

Marsha Blackburn 3  
Ted Cruz 1  
Bernie Sanders -0.5  
Jon Ossoff -0.5

18 19 20 21 22 23 24 25 26 April

Sources: Twitter; SocialBlade; *The Economist*

