OPEN VOICE NETWORK
Voice for everyone

THE FUTURE OF MEDIA AND ENTERTAINMENT INFORMED BY VOICE

OPEN VOICE NETWORK

Whitepaper

June 2021

By Donald Buckley and Janice K. Mandel

About the Future of Media and Entertainment Informed by Voice

This paper is a high-level view of how some western media and entertainment firms are beginning to use voice technology to expand the reach and value of their content assets. We cite nearly 20 different examples of current use, each a great idea with significant promise for monetization, improved customer experiences, customer acquisition and retention. Voice, in this context, is showing its promise as a critical engagement platform that will, in many cases, become the primary means of connection between business and consumer.

But as I noted above, this is just the beginning. There are dozens of experiments, use cases, and experiences in media and entertainment still to be reported, studied, and documented. We're at a new growth phase of the voice technology revolution that will lead us to an open, interoperable world of billions of voice-enabled end points. Your customers will be talking to bots, talking to websites, talking to smart environments, talking to humans. They, and you, will be living in an AI-driven world, accessible through speech.

New creative approaches to advertising lie ahead, using the power of voice. New KPIs and voice strategies will form around a growing ecosystem of voice media. New forms of entertainment will emerge and extensions of entertainment IP will begin to populate, indeed, have already begun to populate the developing channels of voice.

That is the world that's before us.

We wrote this paper to not only show some of what's now in deployment but to raise the curtain for the media and entertainment industries on a future of conversational AI. And to invite you to join us, the Open Voice Network, and The Linux Foundation in ensuring that such a future works for you.

Donald Buckley
Open Voice Network Ambassador
https://www.linkedin.com/in/donald-buckley-583b193?

Table of Contents

> “An invasion of armies can be resisted, but not an idea whose time has come.” – Victor Hugo

The Future of Media and Entertainment Informed By Voice

It turns out you actually *can* change the world by thinking differently, but timing and execution are everything. Voice technology, which enables public connection, conversation, and commerce through various digital devices, has grown as a significant tool for innovation in media, entertainment, advertising, and other industries. The breadth and depth of voice-driven experiences—from sonic extension of products and programs to sentiment analysis and community engagement—offers the potential to create value through innovative engagement and save on costs by maximizing the value of existing assets.

We’ve seen the borders between digital video, traditional TV, streaming, podcasting, and social audio blend as creatives explore new ways to understand and engage with their audiences. And, market transitions predicted to shape the next three-to-five years—and the pitfalls to avoid—are expected to accelerate as voice technology grows more powerful and is incorporated into new products and services.

It’s OK if you haven’t done much with voice technology yet, but it’s time to start exploring.

Voice technology use is spreading across certain devices and platforms

The statistics on voice adoption across mobile apps, websites, and smart speakers, such as Amazon Alexa and Google Home, tell us that people are interested in the option of engaging with their devices through conversation. According to dedicated voice technology resource, Voicebot.ai, in January 2021 the installed base of smart speakers in the U.S. reached 90.7 million, one-third of the U.S. Adult population. Adoption of smart displays used with voice jumped more than 50% in 2020 to one in four users (25.5%). Voice assistant users on smartphones also rose 11% between 2018-2020, and daily active users climbed 23%.

The public's adoption of voice assistance products and services has grown at a faster rate than smartphone ownership, the very thing that introduced most of us to voice with the early iPhone. By 2023, voice commerce is predicted to reach $80 billion.

Voice technology and its use cases are advancing rapidly. This paper is a snapshot of selected developments, trends, and opportunities for media and entertainment companies.

The back story: Get out of my dreams and into my phone

How we interact with technology colors our experience. There were other smartphones on the market when Apple introduced the first stylized smartphone with touchscreen technology in June 2007, but the iPhone let people swipe and tap with their fingers to interact with their phones. Building on its hip image as the creator of the iPod, the company sold its first 100 million iPhones in its first four years and established the interface as the category standard.

But it wasn't until the iPhone 4S acquired and integrated Siri, the "intelligent assistant," that Apple brought voice-activated technology to the public at scale. Not only could you ask Siri to perform tasks, through artificial intelligence and machine learning your intelligent assistant knew what you wanted and how. Above all, a master of marketing, Steve Jobs, created signature brand experiences and integrated Siri across all Apple operating systems, from iPhone to computer giving all Apple products one signature voice.

And it wasn't until filmmaker Spike Jonze saw over 100 million+ people "falling in love with their iPhones" that he was inspired to write the award-winning romantic sci fi movie *Her*. Released by Warner Bros. Pictures in 2013, its complex, soulful hero falls in love with the intelligent assistant in his operating system designed to meet his needs.

Then in 2016 the US had another taste of voice technology at scale. Amazon introduced the general public to Amazon Alexa and the Amazon Echo. Google Home's launched later that year. The Asian market was quick to adopt top smart speaker brands released there: Baidu, Alibaba, and Xiaomi.

Otto Söderlund, CEO and co-founder of Speechly, which produces a fully streaming Natural Language Understanding API, predicts,

> Consumer expectations will change very rapidly. We remember how quickly swiping was adopted. In the beginning, that was new, a touch screen had never existed before. Adoption was extremely fast. Voice adoption will happen as fast as, or faster than, touch.

Today, Apple's Siri continues to be the #1 voice assistant used on smartphones in the U.S. and conversational agents and voice assistant systems have found their way into our homes and cars. And the places we expect to find branded voice assistants are multiplying. The question is: *How can media and entertainment companies use their assets to engage audiences through voice-enabled experiences and platforms?*

Why it's finally time to tap into voice technology

Insights on Media and Entertainment by Deloitte, PwC, and Accenture point to a growing opportunity for businesses to better understand what content their audiences want and to help them attract and retain customers by delivering it in ways they prefer.

In Deloitte's *2021 Outlook for the US Telecommunications, Media, and Entertainment industry*, Kevin Westcott, Deloitte's US Tech, Media, and Telecom leader describes a convergence of factors affecting creative content providers and distributors:

> ...As consumption of streaming content rises, we've seen growth not just in the number of subscription services, but also in ad-supported models designed to satisfy increasingly cost-conscious consumers. What's more, customer retention (versus acquisition) has become top of mind—making it important that providers offer a broad range of content: video, music, games, and even podcasts. This new reality places a premium on understanding consumer behavior patterns and developing a more nuanced approach to engaging with customers. As consumers experiment with their entertainment options, we strongly encourage

> providers to adopt new strategies and agile approaches for content development, aggregation, and delivery.

And, as streaming choices multiply, broadcasting and cable companies are using voice commands and new voice-controlled experiences to help customers find what they need and do things that they like.

The five-year projection of consumer and advertiser spending data, across 14 segments and 53 territories, captured in PwC's *Global Entertainment & Media Outlook 2020-2024* supports these ideas and suggests the need for policies that respect user privacy and right to choose with whom they share data and why. The PwC report states:

> More consumers are interacting with artificial intelligence (AI) technology through AI assistants on mobile phones, but privacy concerns have limited uptake in other AI consumer technologies.

Nevertheless, PwC forecasts that

> ...by 2024 there will be 543 million smart speakers alone owned across the 20 countries it covers, with growth driven primarily by the Asia Pacific region, which is set to account for 43.8% of global smart speaker ownership in 2024.

Distribution is maturing beyond smartphones and smart speakers

Brandon Kaplan, CEO of voice-first agency Skilled Creative, helps the world's biggest brands build voice programs on Amazon Alexa, Google Assistant, Samsung Bixby, and other platforms. He has observed an interesting duality in voice right now:

> Distribution is maturing. There are hundreds of millions of smart devices, billions of phones, cars, hearables, and wearables. We see an interesting tension as that continues and it is begging for creative innovation. Every car company, every speaker company, every refrigerator company is working on deploying some kind of voice function or assistant within their product, either custom or for one of the platforms. There is some data out there that voice is the fastest growing channel in history. The content hasn't caught up with the platform/distribution. It is a unique opportunity.

Accessibility through universal design: Who's in control now?

Before Smart TV, there was Comcast's groundbreaking X1 remote, originally released with voice control in 2015 as an accessibility feature.

Jennifer Musser Metz, Comcast's Executive Director, Product Management, Voice NLP/AI Platform leads the team working on the Comcast/Xfinity X1 remote, which won an Emmy in 2017 for Contextual Voice Navigation. Comcast's Voice/NLP platform powers the voice experience at a large scale in the U.S. and overseas. Says Jennifer:

> In 2015, we began integrating voice control into X1 as both a convenience and an accessibility feature and it has had side effects we never imagined. It has unlocked TV for the low-vision community and those with dexterity issues. The ability to use your voice to find what you are looking for is the ultimate shortcut.

For the past five years, Jennifer's team worked to make the TV experience easier for all users by building a content-first voice platform that searches across the entire catalog of programming available on X1, including a growing amount of streaming services. When they launched Flex, a 4K streaming device offered free to broadband customers in 2019, they brought the same voice platform and unique cross-app search functionality to a new segment of streaming-first customers. More recently, they scaled this technology out to Sky in Europe as well. Jennifer observes:

> We believe the voice remote is part of a compelling entertainment package for all Xfinity customers, whether they take X1 as a pay-TV customer, or use Flex as an Internet-only customer. We know our customers love the remote, and they talk about it to their friends. It's recognized as a great free feature of our platform.

A recent study by Los Angeles market research company, Guts+Data, conducted between October, 2020 and April, 2021 showed that among the 1500 frequent entertainment consumers surveyed, 14% always used voice commands to help find and view movies and series on streaming services, up from 11% six months earlier. Those who reported that they never used voice commands dropped from 53% to 45%. Those who occasionally use voice commands on remotes number 41%, up from 36% earlier.

Today, we're seeing voice functions becoming a standard for other TV remotes, such as those by Roku and Amazon Fire TV. Catch-up is never easy.

Jennifer says: "We're going to continue looking for ways to enhance our platforms with interactive experiences that make viewing fun and engaging for our customers."

Beyond smartphones and speakers: voice + smart TV

In 2018, Juniper Research predicted that "the fastest-growing category for voice over the next several years will not be smart speakers. It will be smart TVs." By the end of 2020, Smart TVs were already #1 in the smart home adoption category and in 37.9% of U.S. households.

Kirill Petrov, CEO and Founder of Just AI, an international vendor of professional conversational AI tools and technologies, observed in a Voicebot.ai year-end round up:

> Voice assistants in smart TVs is an obvious placement for a voice assistant. With a smart assistant on your TV, you can easily browse the channels, search for the content, launch apps, change the sound mode, look for information, and many more, depending on the TV model.

In today's increasingly crowded market, having plenty of high-quality content is a competitive advantage, but only if audiences can find it. Streaming video, voice experiences, music, and podcasting platforms help retain customers through positive customer journeys and an easy-to-use search and discovery experience.

For entrepreneurs like Jason and Danny Cohen, the brothers who built MyBundle.TV to help make the streaming experience simpler for all, voice can be a powerful tool to find specific content quickly. They are especially interested in how smarter and deeper voice control will allow their device-agnostic platform to extend from phones, tablets, and computers to the TV screen.

Finding their voice: Early M&E industry experiments

As the public first learned to access voice skills and apps through conversational agents, marketing teams at some media and entertainment companies experimented.

Warner Bros. Pictures launched a 2016 collaboration to create the first Amazon Alexa skill to combine voice-first technology with produced audio assets for music and sound effects. Warner Bros., head writers at DC Comics, and Unusual Films collaborated with Amazon to create a choose-your-own interactive skill called *The Wayne Investigation* to promote the feature film, *Batman v Superman: Dawn of Justice*.

During its first week, *The Wayne Investigation* engaged seven times more (per weekly average) than all other skills combined; earning the top spot for both total time spent engaging and average time spent per user.

John Limpert, former Warner Bros.' VP of Emerging Marketing Technologies observes:

> *The Wayne Investigation* was our—and Amazon's—first foray into creating an interactive and immersive audio experience. We were pleased with the results as it was one of the top Alexa skills for months even after the theatrical release of *Batman v Superman: Dawn of Justice* had ended. Bringing audiences into the world of our films was always a top goal for us and *The Wayne Investigation* proved that voice skills were a powerful new way to reach them.

Warner experimented with multimodal the following year when it launched an Alexa skill called *Destination Dunkirk* to promote the Christopher Nolan film *Dunkirk*. Limpert notes:

> The new twist was the ability to show images that corresponded to the audio on Amazon devices that had a screen. I'm confident that as voice skills continue to evolve and allow for more complex voice interactions, their popularity will grow.

Combining practical information with voices of the stars

In 2017, Showtime Networks engaged the Rain agency to build an Amazon Alexa skill that provided program scheduling, additional information about the show, and featured audio clips of the stars from the series, *Billions*, *Shameless*, and *Homeland*. It was a simple question-and-answer tree that allowed users to ask about scheduling for the network's movies, documentaries, sports and series. The skill concludes with the question: "Would you like to hear more about the show?" Those answering "yes" hear an audio clip stripped out of existing promotional trailers and interviews. It was a cost-effective way to reuse Showtime's audio assets and engage the audience with the limited functionality available at that time.

Early experiments evolved beyond the once-and-done approach to pushing out a single project. Instead, media companies integrated voice as a sonic extension of their brand's products and programs. For example, Warner Bros. for its *Shazam* superhero movie, gave fans another taste of the multimodal experience in 2019 with the first voice-activated augmented reality lens on Snapchat, a filter that responded to the voice command, "Shazam!," and allowed users to see themselves as the superhero.

Voice adds that *sonic something* to brand identity

Done well, voice and other sound experiences connect audiences at a more emotional level. Audrey Arbeeny, founder and CEO of Audiobrain, is among the pioneers in this space. Her company has provided sonic identity and projects for over 20 years for iconic brands, such as the Olympics, and others:

> Adding voice to any media or entertainment initiative makes it a richer experience. Voice brings an emotional human connection that exceeds those of visuals. It enables the listener to feel closer to the artist. It is also more memorable, and can clearly communicate with the audience. That's why many brands—pre-pandemic and I'm sure post-pandemic—who formerly had expensive meet and greets now have free or inexpensive personal fan experiences. There are so many places it can resonate: interactive advertising, promotions—nothing like hearing the right voice to enrich the experience.

Engaging over voice adds an information-rich layer to the experience that can be analyzed and looked to for audience insights. Customer conversations reveal not just what has been said, but the tone, vigor, and vitality with which it is stated. NASA analyzes the voices of its in-flight astronauts as part of its physical and mental health screenings. This is among the reasons that transparent disclosure about privacy policies must protect the privacy and security of user voice data.

A SVP of the Worldwide Consumer Insights & Audience Analytics division of an American film production and distribution company says that she loves to hear the excitement in the fans' voices when they engage in a skill where they can leave comments about their favorite parts of a movie. She also appreciates the need for analytics coming back from any voice experiences. “I’d necessarily want to look through an “audience lens” – and it will be very helpful,” she observes:

> I want to open the aperture and think about voice in a holistic way—thinking about podcasts, devices, platforms, social audio. I want to know everything about how audiences are engaging. What are they responding to? What else are they doing in the voice ecosystem? How can we paint a better picture of the moviegoer?

Entertainment companions or extensions

Brandon Kaplan from Skilled Creative observes that there is a movement to entertainment companions or extensions, including interactive podcasts, voice-activated podcasts, and a long list of voice-activated games like *Jeopardy!* and *Who Wants to be a Millionaire?* He says:

> There’s a great opportunity for media and entertainment to leverage their distributed channels, but there is almost a responsibility on their part to start creating awareness and demand. *Minesweeper* was originally deployed as a training mechanism to teach people how to use a mouse.

> We need that kind of training now. Media companies have the scale and audience to do that.

Comcast integrates subtle on-screen suggestions for using its voice remote. Elijah Vargas, who leads Voice and Sensory Design at Comcast, helps provide subscribers with voice feature awareness and educate them on its use. He says:

> These suggestions drive user engagement with features we promote in context, and often result in increased overall voice usage. We view the user experience through a multisensory lens and consider how we can leverage sound, touch, and lighting to provide a richer, more accessible experience.

For the past decade, Comcast's X1 and Flex have integrated a variety of interactive features, made possible by the platform's ability to use deep metadata to know what's going on in a given scene at a given time, and trigger an interactive experience on top of it. Examples of the type of interactive experiences enabled include live voting for reality shows like *The Voice* and *America's Got Talent.*

Studios continue to experiment, mindful of the opportunities and challenges in a rapidly developing industry. Says one senior digital marketing executive:

> We are thinking about a lot of voice activations and partnerships this year that will be fun and effective, but the early efforts feel a little like marketing stunts. There's no consistent way we are currently working with Amazon, Google, Apple or any of the others.

Until the industry matures a little, he wonders if media companies might be better off testing out ideas by hitching a ride with others who have built a voice audience.

Partnering for distribution

The number of platforms cultivating and reaching specific audiences is growing with voice-control features for apps, products, platforms and digital channels.

As PwC's media and entertainment industry predictions for 2021 noted: "'Epic Games' *Fortnite* can now be credibly described as the world's largest event space, capable of hosting major live music, cinematic, and other experiences."

Examples continue to emerge as the early 2020's shapes up to be an era in which media and entertainment companies develop voice experiences related to the content they produce or feature. Some platforms, such as Spotify, are enabling content creators—who tend to attract fans, such as musicians or professional sports figures—to be discovered more widely and engage with their communities. In Spotify's case. Examples include:

- Drivetime is the first entertainment company founded and funded to develop interactive voice games for drivers. In September 2019, it launched the *Jeopardy!* channel inside the Drivetime mobile app and announced that Drivetime Premium subscribers will get to play episodes of *Jeopardy!* for the first time as an interactive voice experience in their cars. Paul Joffee, Vice president of Games at Sony Pictures Television, has said his team looks forward to "building on *Jeopardy!*'s mobile offerings and entering the in-car entertainment space."

- Spotify announced in June 2020 its multi-year deal to produce and distribute an original slate of narrative scripted podcasts featuring various DC Comics superheroes and villains. One entertainment industry executive close to the deal says: "Spotify gets a first look at original scripted narrative of DC Comics podcasts starring recognizable characters from the rich history of comics. The expanding audio streaming platform will also work with Warner Bros. and DC to create new programming from original intellectual property."

- AMC Networks, home of *The Walking Dead*, became the first premium content partner with the launch of The Walking Dead Universe Twitch Channel to engage viewers on Twitch, the American video live-streaming service that Amazon acquired and rebranded in 2014. Later, McDonald's became the channel's first sponsor. Active integrations included host shout-outs and customized "chatbot" language that was triggered when fans used specific words in the chat.

- Sony and popular communication service, Discord, announced a partnership in May 2021 that will integrate Discord's gaming-focused chat app with PlayStation's own built-in social tools. Sony out-negotiated Microsoft and took a minority stake in the company, which will use its own branded option to start a voice chat.

Technology vendors assist with specialized analytics tools

Winner of this year's annual Industry Star Award for Media Excellence, Veritone has been a trusted technology partner for many leading media and entertainment clients. Media networks, broadcasters, podcasters, movie studios, and professional sports teams use Veritone's operating system for artificial intelligence, aiWARE™. Veritone's President, Ryan Steelburg, explains:

> From Veritone's perspective, we continue to see an ever-increasing appetite for personalized and engaging content, either independently or in conjunction with another application or service, such as Fantasy Sport. Also, the format and nature of content demands continue to expand,

> including video, on-demand audio, including podcasts and voice assistants, imagery, and data, such as stats, predictions, polls, and more.
>
> For our Audio and Video customers, we currently ingest and process over 50K hours/day of content. We run these disparate feeds, streams, files, through various cognitive models, including NLP, Object Detection, OCR, Face Detection and Identification. In addition to this baseline model optimization, we are also looking at multi-variant queues, to better inform or train models. For example, if we are trying to validate humorous sentiment from a person's speech, we can analyze the face, we can reference the video context and look for confirming indicators, such as smiling or laughter.

Veritone is transparent about the privacy policy that governs how it processes and protects information used to serve clients. The creator of the first AI operating system was the first to join the Open Voice Network's new "Supporter" category of sponsors. Veritone user analytics depends on access to personal customer data. Their privacy policy is published on their website and outlines principles that make their clients comfortable working with them, such as:

1. The right to be informed of the specifics regarding your personal data, including what data is being processed, the purposes for such processing, the third parties with which we share your data, the sources from which we obtain your data, and the period of time for which your personal data will be stored;

2. The right to receive a copy of your information that we process; and

3. The right to ask that we provide your data in an easily readable format to another company.

Veritone also offers the option not to use their services if a potential client doesn't agree with its privacy policies.

The "audio natives" lead the way

Audio natives, such as National Public Radio (NPR), Pandora, and Spotify, are expanding quickly through voice-driven innovation, acquisitions, and partnerships. They know what works, and what needs to work to function on a voice equivalent of the internet and how to position themselves to dominate in new categories. These, like the BBC, are early adopters on the front lines for public feedback, positive and negative. They're still figuring it out, but they share insight into how to capture opportunity and the risk of public controversy over certain uses of machine learning and voice recognition technologies.

NPR extends its quality content across major new technologies

On April 9, 2014 NPR announced that when the first Amazon Echo hit the market users who asked Alexa to "play the news" would hear the most recent NPR hourly newscast. Mic drop... It was a natural extension of the NPR.org brand to deliver what they called "the best nationally and locally produced public radio content on smart speakers."

The terrestrial radio public broadcaster is accustomed to exploring emerging technology options for sharing their rich content more widely and in ways that listeners like to explore. Along the way they also produced mobile apps and podcasts. Voice assistance technology has been a natural option that has helped engage more listeners with high-quality content. Despite the earliest conversational agent's limited natural language understanding and processing issues that prevented interaction with their brand experience, NPR continues to explore new voice options.

Since 2015, NPR's principal product manager, Ha-Hoa Hamano and her team of engineers and Voice Experience designers (VX) have been delivering innovative news formats, stories, and games creating voice skills and apps to complement regular programming. Their aim is to provide options for listeners to further engage and respond by speaking to a conversational endpoint, such as a mobile device, smart speaker or personal computer. NPR has since branched out to voice apps and skills accessible on Siri and Google as well as Alexa.

NPR's former CTO wrote in a memo to local stations:

> [Voice assistant platforms are] a huge part of our shared future, and when you ask Alexa, Siri, or Google to play your local Member station, or NPR, or the headlines, or your favorite podcast or show, we want to ensure the experience is as direct, relevant, and compelling as possible.

Die-hard NPR fans never have to be without their favorite NPR station. They listen in their cars, homes, offices, smartphone apps, or other conversational endpoints, on smart speaker apps and skills, mobile devices, and personal computers. That's a holistic brand experience.

More on audio natives after this word about voice ads

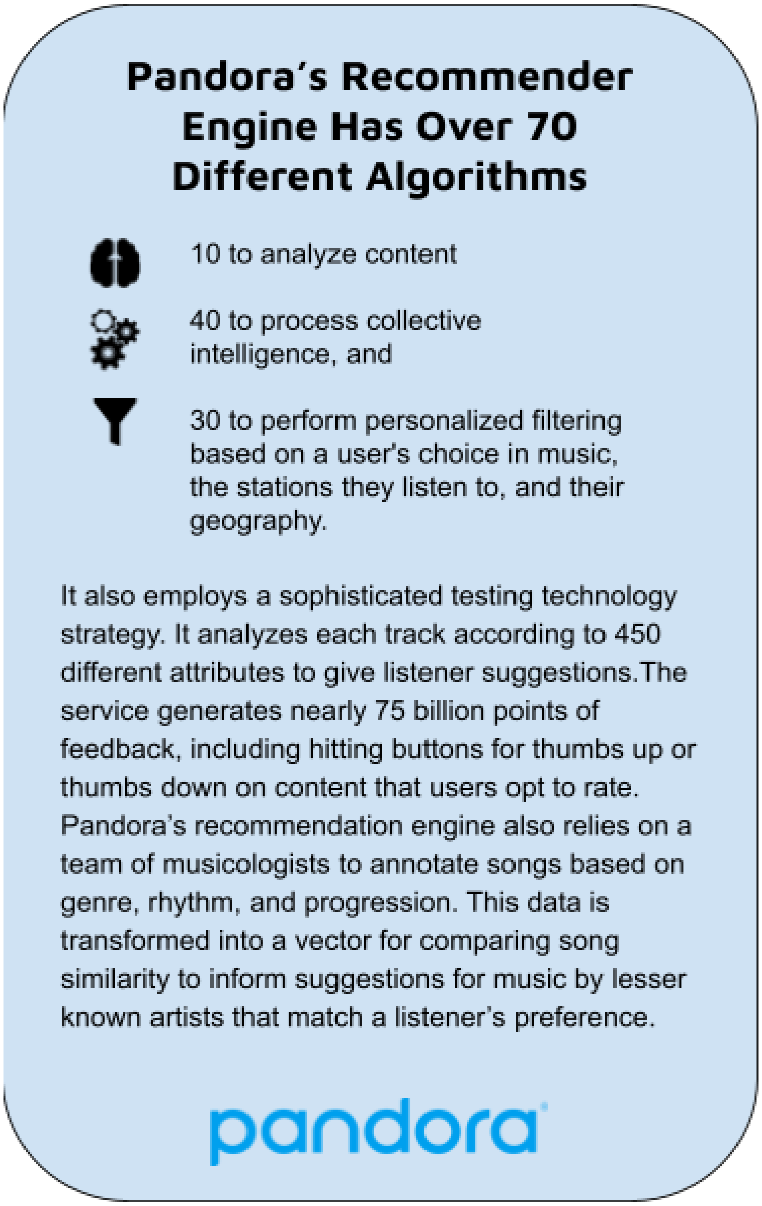
Pandora, founded as a subscription-based music streaming service in 2000, built its reputation on a familiar concept: customized radio stations—but, delivered over the internet. Five years later, to remain competitive, Pandora research created Music Project Genome, a better music recommender engine that includes 60 million labeled songs.

In retrospect, Pandora's trajectory to be among the first to deliver interactive voice advertising took a logical route. Launched during tough economic times, the subscription-based music streaming service found it had more serial "free trial" users, those using multiple email addresses to get free music instead of paying for a subscription. In response, Pandora created a free version of its music streaming service supported by ad revenue, i.e., running about seven 15- or 30-second ads per hour. At its peak, with 81.5 million listeners in the final quarter of 2014, Pandora's competition from Apple Music, Amazon, and new streaming rival that year, TIDAL, pushed music streaming into the commodity category. Pandora expanded its recommender engine service know-how to podcasts with its Podcast Genome Project in July 2018.

At the end of the following year, Pandora was among the first to start testing interactive voice ads with early adopters, such as Doritos, Comcast, Unilever, and Turner Broadcasting. It introduced its new advertising formats into wider public testing where people would be likely to be engaging in activities that occupy their hands, such as cooking, driving, or house cleaning.

Pandora's advertising business is still growing in 2021. The first metrics for voice ads show promising results. Ad revenue was up 29% year-on-year to $312 million in Q1. Instead of using the online ad metric of "click rates," Pandora uses a standard first-party metric for measuring verbal engagement with voice ads called: "say-through" rates. To obtain this, the company uses software from Instreamatic, which provides a voice AI platform for managing, measuring, and monetizing conversations between brands and consumers in mass marketing channels which in April 2021 received $6.1 Series A funding to scale.

Inside Pandora's Recommender Engine



Pandora's Recommender Engine Has Over 70 Different Algorithms

- 10 to analyze content
- 40 to process collective intelligence, and
- 30 to perform personalized filtering based on a user's choice in music, the stations they listen to, and their geography.

It also employs a sophisticated testing technology strategy. It analyzes each track according to 450 different attributes to give listener suggestions.The service generates nearly 75 billion points of feedback, including hitting buttons for thumbs up or thumbs down on content that users opt to rate. Pandora's recommendation engine also relies on a team of musicologists to annotate songs based on genre, rhythm, and progression. This data is transformed into a vector for comparing song similarity to inform suggestions for music by lesser known artists that match a listener's preference.

pandora

Source: "How Pandora built a better recommendation engine," *TheServerSide*, by George Lawton https://www.theserverside.com/feature/How-Pandora-built-a-better-recommendation-engine

Molly Mitchell, Sr. Manager, Ad Innovation Strategy at SiriusXM has seen voice ads outperform traditional text click-through rates. She said:

"During closed beta, voice ads had up to 10x higher say-through rates over click-through rates, suggesting that voice is 1) a more native way to engage with audio and 2) enabling a way to engage while listeners are otherwise hands-free.

Using Pandora Surveys, we also found that voice ads scored 7.6 points higher in voice ad recall over the control. This makes sense: voice ads are not only a memorable new ad experience, but since users are engaging with these ads at high rates, they're more likely to remember what they're engaging with.

> Finally, our Veritonic tests showed that voice ads had 27% higher purchase intent than the audio ad benchmarks for those who engaged the ad. They also had consistently high scores for attributes such as Relevant to Me, Interesting and Trustworthy."

If you ask Sarah Andrew Wilson, Chief Content Officer of voice-first game creator, Matchbox.io, if her company would welcome ads for its growing portfolio of 20+ voice experiences on Amazon Alexa, Google Assistant, and Samsung Bixby, you'll get a resounding "Yes!" Sarah says: "We have been waiting for four years for advertising to be available on Smart Speaker platforms. It will be a game changer. That's when global brands will be truly invested."

She understands the hesitancy of the platforms to introduce advertising, noting: "We have learned that many people who buy smart speakers regard them as something they paid for to use without commercialization and advertising." This was not the case with radio and TV. As we move towards platform ads, Sarah wonders if the platforms will assert control of the content of advertising as they have controlled customer data to date. Will the new generation of platforms that provide audiences for content creators also share their customer data?

Spotify gears up to be the Netflix of audio, the world's audio browser

Like NPR and Pandora, Spotify grew up as an audio native. Founded in 2006 as a music streaming service provider, Spotify embraced multiple uses of voice technology, superior sound, and an extensive range of audio content. As of Q1 2021 the company reports over 356 million monthly active users, up 24% year over year including 158 million subscribers, up 21% year over year. Key to its future growth is global distribution—Spotify recently added 80 new markets and has offices in 17 different countries to help provide maximized revenue, a wider reach, and discoverability to its platform users.

Spotify began expanding as an all-audio platform strategy in February 2019 when it acquired podcast producers Gimlet Media and Anchor, which provides podcast creation technology. The company's March 2021 acquisition of Betty Labs, the creators of sports-focused, live social audio app, *Locker Room*, signaled Spotify's shift as a community facilitator for anyone with fans—from musicians to sports teams. The platform provides options to process subscriptions, sell merchandise, or pay per view. It's where Barack Obama and Bruce Springsteen join forces in their original podcast, Renegades: Born in the USA.

In the company's first Quarter 2021 earnings remarks, co-founder and CEO, Daniel Ek, outlined the strategy to move from a streaming service to an audio platform that enables access to paid audio products on top of Spotify. He observed in his Q1 2021 report to investors:

> I continue to believe that the opportunity in audio is still largely untapped with tremendous growth potential—far beyond what most of us can imagine today. Take music as an example. The latest numbers from IFPI (International Federation of the Phonographic Industry), just out this quarter, reinforce the strength of the industry, which has seen an increase of 54% in global recorded music revenues since the 2014 low. And streaming revenues, with a growth rate of more than 600% over this same time period, continue to propel the industry forward. Industry sources recently forecasted that the streaming market will triple, reaching $79 billion in revenues by 2030 and Spotify continues to be the primary driver that is pushing global music revenues to record highs.

In April 2021, Spotify announced a new voice-controlled experience, “Hey Spotify,” to make it easier for users to start, navigate, search, and customize discovery for music and podcasts by speaking to a conversational agent. Spotify’s users took this as an opportunity to protest a recently granted voice recognition patent.

Voice insiders have a clearer view of the path to successful adoption and growth

Successful skills are part of programs, not standalone projects.

From early 2018 through 2020, voice designer and strategist, Nina Neuf, developed and promoted a series of voice-driven experiences for the German office of a global music company at which she’d worked for 21 years. The voice experiments ranged from “catchy tune of the day” and trivia questions to radio plays and music mixed with news about releases, artists, and upcoming performances. In the end, she says, the most successful experiences were part of a broader program to promote music discovery.

For example, her favorite experience was *Let’s Discover Classical Music*, an Alexa skill that facilitates a dialog between a brand voice or a classical music artist.

> First you hear a story, fun trivia about a piece of music or a composer. Then, you get a music suggestion, and if a music streaming service is connected to the Alexa account, we can play the music.

> The key to success for these voice experiences is to make them 'visible' by marketing them in social media and other corporate channels. It worked best when the artists had a large social media following with which to discuss the experiences.
>
> Music content is attractive when listeners find what they are looking for. The artist name, song, and album title must be recognized and found as well as playlist titles or keywords. For example, in order to 'play dinner music,' data needs to be identified as such. Dinner music sounds different in the summer compared to around the holidays. It can be more attractive when listeners feel the content has been personalized for them.

Test, tweak, repeat…monetize

Customer acquisition costs are significantly higher than retention costs. As part of its trial and error on growth, Matchbox.io not only focuses on how to attract audiences, the voice-first question and answer game maker doubles down on customer retention.

Sarah Andrew Wilson observes:

> Question of the Day, our seemingly simple first question/response game on Amazon Alexa has been around for four years and we're still experimenting and looking for opportunities.
>
> Part of our efforts in retention for Question of the Day included an expansion to a website, a podcast and now, a mobile app. What works for us is new, high-quality content every single day.

Question of the Day has grown to 11,000,000+ total users since its launch. As Sarah explains:

> We keep iterating, tweaking, A/B testing. At the beginning, we launched three skills in January of 2017. *Question of the Day*, *Three Questions* and *Daily Poll*. We were trying to figure out what kind of model would work. One question? Three questions? Or do people want to be polled?
>
> With its success, we shifted more focus on retention. How do we keep people coming back day after day?
>
> We want users to be surprised, never knowing what's coming up next. With *Kids Quiz* or *Question of the Day* we try to vary the questions, something you may have seen on the news a while ago, or a moment in history you may or may not remember from school. Our hope is that it is always fresh, always fun.
>
> We expanded to include French, Spanish and Brazilian Portuguese, but we hired interpreters and translators, who look for issues like cultural relevance and of course, pronunciation. They even write new content specific to their regions and then audio test after translation to make sure Alexa and Google pronounce correctly.

Sometimes small things really add up

Matchbox.io wasn't getting the traction they wanted with subscriptions so they tested and tweaked.

> We did some A/B testing and found that by changing one word, we saw 17% more people sign up for Trivia Club than with our previous messaging. We are running those kinds of tests all the time.
>
> Amazon allows subscriptions, one-time purchases and consumables. We have experimented with all three. An example of a consumable might be an extra 'life,' which when used, is gone. One-time purchases are things like game-packs, which can be used over and over, like owning a book. And subscriptions, which we offer.
>
> *Guess My Name* is one of our very popular games, which has all three of these. But the subscription model on *Question of the Day* really took off. People around the world really enjoy trivia, and if you give them one question a day that is high quality, many will be willing to pay more for a subscription at $2.99 a month to get 4 questions per day, plus unlimited game packs, leaderboards and other features. It has really driven the growth of the business.

Tools for automating faster, less-costly audio content creation.

Rumble Studio co-founder and CEO, Carl Robinson, hosts the *Voice Tech Podcast*. His recently funded company uses conversational AI to automate interviewing and recording guests asynchronously and publishing the final interview. More and more consumers are now accessing content on-demand via voice search, so an increasing proportion of all content consumed will be in audio form. Carl observes:

> Audio content marketing via voice will soon be considered a primary channel for many organizations. In response, brands will produce more bite-sized, specialized content to answer questions posed by current and prospective customers via voice, as well as longer content for entertainment purposes. Interactive content and advertisements will also grow in popularity, as the lines between the two become ever more blurred.

Veritone draws on advanced artificial technology to provide the San Francisco Giants with the ability to maximize the utilization and monetization of its rich media assets. Ryan Steelberg explains:

> By ingesting their archive content into our Cloud-based DMH (Digital Media Hub), powered by aiWARE cognition, we provide an indexed, searchable and actional media library. Now, this content is immediately available for Marketing campaigns and sponsorships, consumer experience, and social media engagement.

Everyone needs trustworthy functionality

Jordan Mirrer has built a career on digital production and the design and development of games, apps, and experiences for voice, mobile, TV, and other digital platforms. Starting with the release of *Jeopardy!* on Alexa in 2015, he’s been making voice games with Sony Pictures Entertainment, currently as director of design and production for interactive voice. The experience has taught them what works and what needs work. Among the top needs, he says, include improved functionality and transparency on data use. Says Jordan:

> All of the hopes that everyone had five years ago, with the hype of voice assistants, those hopes haven't really materialized. And that is one of the biggest problems holding the platforms back. It even goes beyond trust of data. Trust of functionality is an issue. That’s the most common complaint we have on user reviews for *Jeopardy!* and *Millionaire*—misunderstandings and failures of interpretation of answers and words.
>
> Even basic data transactions, like signing up for a newsletter, are difficult. User friction for the email permission is just too high. You have to go to the Alexa App to do these things and nobody wants to do that.

Among the concerns cited are smart speakers and other platforms that collect and use revealing personal information from users, such as biometric voice data, to customize responses. Matchbox.io makes a point of collecting only the data that platforms require to serve users, such as email address and zip code.

“Personally it makes me uncomfortable,” says Sarah Andrew Wilson. “Matchbox.io has nothing planned for that kind of information gathering. Maybe we would be aided by knowing when a user is getting frustrated by their tone of voice, but we can already tell if they shut us off early. There are other ways of determining pleasure or displeasure in the voice experience.”

"I'd like to be optimistic. Where we are moving to, there is the potential for self-sustaining products," Jordan adds. "No one has had a breakout hit yet like when the App Store first opened up. The closest thing to *Angry Birds* or *Fruit Ninja* I think may be *Jeopardy!*, but we are still not seeing that level of success yet."

Market transitions shaping the next three-to-five years

Can we reasonably use what we know about the past five years of voice to help us understand what's going to happen in the next five? Will it all change dramatically? Is it worth pondering?

Accenture Innovation and Insights manager, Oxana Gouliaeva, thinks it's just the beginning. She predicted in Voicebot.ai's most recent industry roundup:

> The convergence of software agents, voice and conversational technologies, 3D, VR/AR, supported by the 5G, will lead to a further development of virtual characters acting as contextualized interactive storytellers and influencers on behalf of brands. This accelerated blending of the physical and digital worlds will be particularly interesting to see in the gaming, entertainment, and educational fields for a better engagement with Gen Z.

The TV and digital video industry appears likely to continue to flatten among traditional TV, streaming, and social video companies.

The first three days of the 2021 Interactive Advertising Bureau's NewFronts demonstrated how the TV and digital video industry is flattening. The event's fourth and final day featured a social video platform, publishers and a TV network group sharing the virtual stage. TikTok called its platform TV-like, while Meredith [Corporation] talked up People TV and NBCUniversal introduced a streaming ad format that looked a lot like linear TV among traditional TV, streaming, and social video companies.

Wearables will play a significant role

Worldwide end-user spending on wearable devices will total $81.5 billion in 2021, an 18.1% increase from $69 billion in 2020, according to the latest forecast from Gartner, Inc. The rise in remote work and increased interest in health monitoring during the COVID-19 pandemic was a significant factor driving market growth. According to Ranjit Atwal, Senior Research Director at Gartner:

> The introduction of health measures to self-track COVID-19 symptoms, along with increasing interest from consumers in their personal health and wellness during global lockdowns, presented a significant opportunity for the wearables market. Ear-worn devices and smartwatches are seeing particularly robust growth as consumers rely on these devices for remote work, fitness activities, health tracking and more.

Spending on ear-worn devices rose 124% in 2020, totaling $32.7 billion and is forecast to reach $39.2 billion in 2021 This growth has been largely attributed to remote workers upgrading their headphones for video calling and consumers purchasing headphones to use with their smartphone devices.

Machine learning used with voice recognition continues to emerge as core to new products and services and be a lightning rod for privacy and security concerns.

When Spotify announced its “Hey Spotify,” voice app in April 2021, it came on the heels of the company having been granted a patent that would enable its voice recognition technology to identify the sentiment, gender, age range, and health reflected in user voices. Combined with listening to ambient sound, Spotify determines if a playlist is being created for use in a crowd or a quieter setting. Spotify underestimated the shift that had occurred in terms of concerns for user privacy.

That same month, nearly 200 musicians and allies from civil rights groups signed an open letter asking Spotify’s CEO never to develop the patent that would enable the company to serve up a choice of songs for a customized playlist based on mood, gender, or age that might manipulate feelings or, at best, lead to awkward recommendations based on stereotyping. Amazon itself is among others with patents for similar functionality. It is common industry practice to secure patents relevant to strategic opportunities.

The extent to which variations of controversial capabilities are being developed and patented throughout various industries and how they can be managed responsibly is not well understood by the public. This is among the reasons the Open Voice Network is reviewing existing policy and regulations to suggest new guidelines and help organize industry infrastructure to manage destination registration, the voice equivalent of web addresses for brands and standards for greater interoperability, and making voice assistance accessible to more users and worthy of user trust.

Ensuring data security and privacy of the supply chain will be good for business.

Looking ahead, organizations will do well to take a risk-based approach to evaluating partners and vendors, and establish agreements about topics such as data breach notification obligations and cooperation in fulfilling data subject requests. According to Gartner, ”brands that put in place user-level control of marketing data in 2023 will reduce customer churn by 40% and increase lifetime value by 25%.”

Companies need to ensure that their partners, suppliers, resellers, and service providers are protecting user data properly. The GDPR requires working only with third parties that demonstrate they have measures in place to protect personal data. The Open Voice Network supports taking a risk-based approach to ensure that user privacy and security protect and serve users first.

Experienced developers look to a future with more, not fewer assistants

Jordan Mirrer takes an expansive look at a future with many voice assistants:

> My hope is that things get opened up. As you can choose what browser to use, you should be able to choose the assistant on your device. My concern is not that tech companies are getting too big and need to be split up, necessarily, but it is that they need to be more transparent. If I want to move all of my data from Facebook to another company, I should have a clear path to do so. We need to enable competition.

Sarah Andrew Wilson is excited for Matchbox.io's role in assisting to help spread the use of voice that respects individual data privacy and security in entertainment. She says:

> Smart speaker sales may have leveled off, but use of voice is climbing, in your car, as a calling device, and in so many other uses. It is our opportunity and responsibility to bring those users along.

Another senior digital marketing executive at a major movie studio is optimistic about the future of voice at studios and beyond. He says:

> I'd like to get to a point where we are using the voice technology to complete the experience, from initial awareness to transaction. You just walk into the theater, no ticket stub required, just your voice. I'm excited to go there.

Brandon Kaplan predicts the majority of voice interactions will be split between mobile devices and platforms like Alexa and Google Home, with more and more mobile usage and smart screen usage. He is optimistic for the future of voice:

> Once we figure out how to connect voice more smoothly to mobile, web, and smart home, voice becomes a high-quality touch point in a user's journey. I correlate it to the frequency of use of in-home devices in favor of what was done previously on a phone. It is not easy now to jump across modalities and create an omni channel initiative that jumps in and out of voice.

> Talk to people who have a smart device in their home and ask them how often they take their phones out of their pocket to check weather, or set timers or ask simple questions. They are using the voice device because it is a frictionless experience that is really clean and smooth. On the entertainment side there is still a lot of friction. Once we make that as easy as asking about the weather, behavior will change.

Conclusion

Voice technology has gained public attention as a highly accessible way to engage with digital devices by speaking. It offers media, entertainment, advertising, and other industries a broad and deep range of possibilities to generate value by creating accessible ways to engage with audiences, connect communities, and maximize the value of existing digital assets through voice interfaces and platforms.

But, there's work to be done to ensure the voice tech industry maintains a strong foundation. Supporting the Open Voice Network's work on reliable functionality through system interoperability and best practices for information processing that respect individual privacy and security are essential for making voice worthy of public trust.

The best voice-based interactions are transparent about how they protect users and provide new and powerful ways to create value. It turns out you actually *can* change the world by thinking differently, but only if you execute.

About the Open Voice Network

The Open Voice Network is a neutral, non-profit industry association dedicated to the development of the standards and ethical use guidelines that will make voice worthy of user trust. It operates as an open source community within The Linux Foundation, and is independently funded and governed with participation from more than 120 voice practitioners and enterprise leaders from 12 countries worldwide.

Open Voice Network community's work at present is focused in four areas:

1. Interoperability, defined as the ability for conversational agents to share dialogs (and accompanying context, control, and privacy),

2. Destination registration and management, the ability of users to confidently find a destination of choice through specific requests, and for the providers of goods and services to register a verbal "brand" (similar to the DNS of the internet);

3. Privacy, with voice-specific guidance for both the protection of individual user data and that of commercial users; and

4. Security, with a focus on voice-specific threats and harms.

Ready to have your industry's voice heard? Please support the Open Voice Network by visiting https://openvoicenetwork.org

About the Linux Foundation

Founded in 2000, the Linux Foundation is supported by more than 1,000 members and is the world's leading home for collaboration on open source software, open standards, open data, and open hardware. Linux Foundation's projects are critical to the world's infrastructure including Linux, Kubernetes, Node.js, and more. The Linux Foundation's methodology focuses on leveraging best practices and addressing the needs of contributors, users and solution providers to create sustainable models for open collaboration. For more information, please visit us at linuxfoundation.org.

About the authors

Donald Buckley is the former Chief Marketing Officer of Showtime Networks, Inc., and former senior marketing executive at Warner Bros. Pictures, where he founded the company's first digital marketing division. He is an independent consultant for streaming TV, movies & marketing and is the Entertainment and Media Industry Advisor to the Open Voice Network.

Janice K. Mandel is a communications consultant who crafts the stories of profit and nonprofit organizations that foster community connection, facilitate ethical technology integration, and result in positive outcomes for society. She is an Open Voice Network Ambassador and a leader in the organization's Ethical Use Task Force Community and Outreach Committee.