



Synthetic Voices Shed Light on the Deep-Rooted Gender Biases Embedded in our Tech

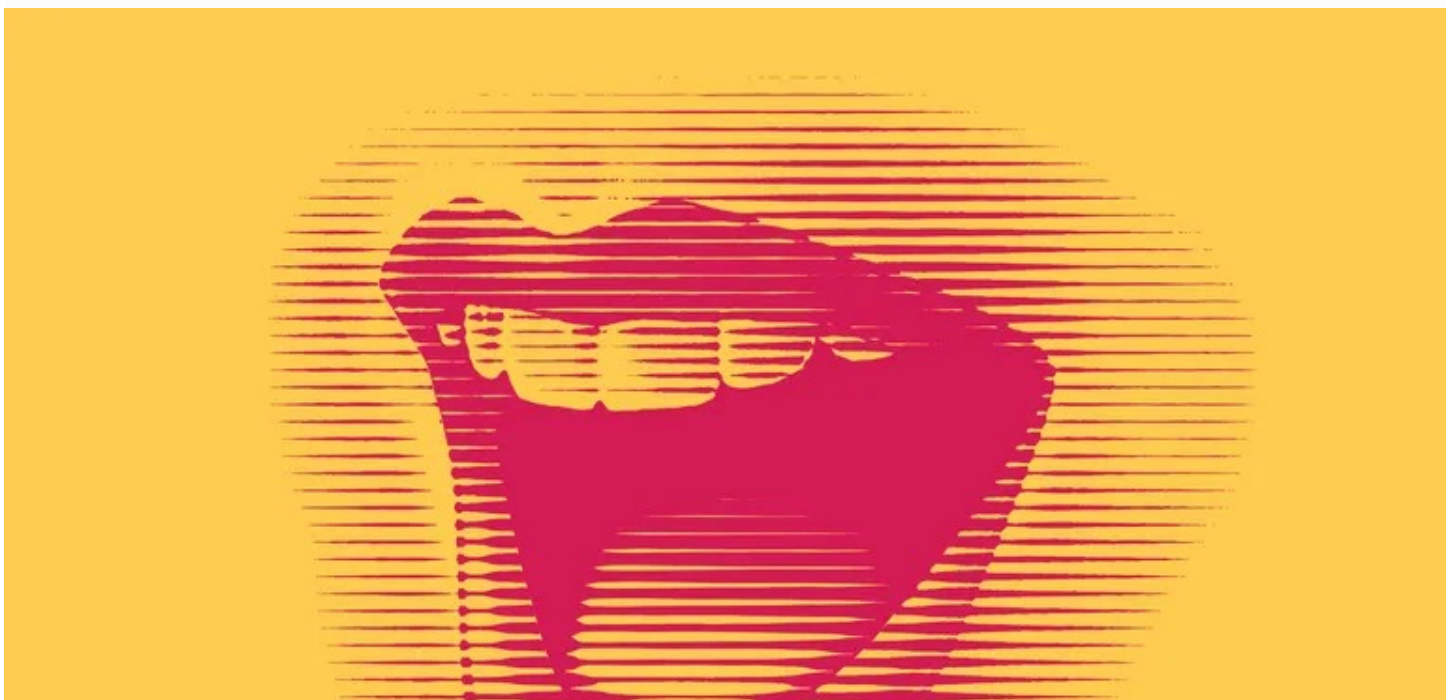
An expert on the impacts of information technologies on society considers how talking machines got their male- and female-sounding voices

Sarah A. Bell, MIT Press Reader

April 9, 2025



Get our newsletter!



Discover more

⚡ voice synthesis

⚡ Communication

⚡ text-to-speech

⚡ Science

⚡ Computer


⚡ computer

⚡ Sciences

⚡ communications

⚡ Speech synthesis





At this point, there is no shortage of stories that have posed some form of the question, “Why are voice assistants always female?” [GeorgePeters/Getty Images](#)

Imagine yourself in the cockpit of a fighter jet, practicing maneuvers over the desert of the American Southwest. Suddenly your altimeter reading is falling, and you must act quickly. The complex panel of instruments in front of you should be second nature to use, but in the moment of crisis, the panels blur together, and your body memory must take over. You begin to make adjustments to solve the problem while simultaneously considering the worst-case scenario. A voice interrupts you, firm but calm, in a soothing alto that reminds you of your mother: “Pull up ... Pull up ... Pull up,” it repeats, and you do what the voice commands, avoiding disaster.

In the 1970s, as the defense manufacturer McDonnell Douglas was developing the F-15 Eagle fighter jet, testing revealed to engineers that pilots’ reactions to warning lights were too slow, especially as the cockpit display increased in complexity. In addition, the development of “heads up” display technology meant that pilots increasingly received information about their aircraft within their field of vision instead of having to look down at a panel of meters and lights. Engineers were concerned that a cacophony of warning bells and buzzers would just add confusion to the mix.

Testing by the U.S. Air Force had shown that a verbal warning system would be more effective—that a human voice breaking into the cockpit would convey a sense of urgency, as well as offer clear and unambiguous directions at the point of need. Systems using recorded warnings had already been installed in some aircraft in the 1960s, but voice synthesis promised to make voice warning systems lighter and more reliable.

Engineers purportedly chose a female voice for the warnings because they believed it would stand out to male fighter pilots. A young actress was recruited to record a series of words that were integrated into the warning system of the F-15. That actress, Kim Crow, recalls that after one of the test flights, the pilot was asked how everything worked; he said, “It was wonderful, except for that Bitching Betty.” The name stuck.

According to Green’s Dictionary of Slang, a “Betty,” meaning an attractive woman, came into use with reference to long-suffering Stone Age housewife Betty Rubble from the cartoon “The Flintstones.” In the days of recorded warning systems, the B-58 Hustler flight crews referred to that aircraft’s warning system as “Sexy Sally.” There were also systems that used male voices, the nickname for which was “Barking Bob.” Although “Bitching Betty” seems derogatory, some pilots have said that they use it as a term of endearment; the voice warnings can save their lives, after all.

Until the 1980s, consumer-grade synthesized voices were in a pitch range that most listeners associated with a male gender. These voices didn’t come close to approximating the prosody or timbre of human voices, but they could produce recognizable language and were often identified with the personal pronoun “he.” Early attempts at synthesizing female-sounding voices consisted of scaling the formants—the peak frequencies that define vowel sounds—of the “male” voice, but this did not succeed in turning the male voice “into a convincing female speaker,” as MIT research scientist Dennis Klatt noted. In one paper, Klatt, who studied both speech perception and speech generation, expressed dissatisfaction with the state of research, stating that “it is not inconceivable that the sound spectrograph has had an overall detrimental influence over the last 40 years by emphasizing aspects of speech spectra that are probably not direct perceptual cues.”

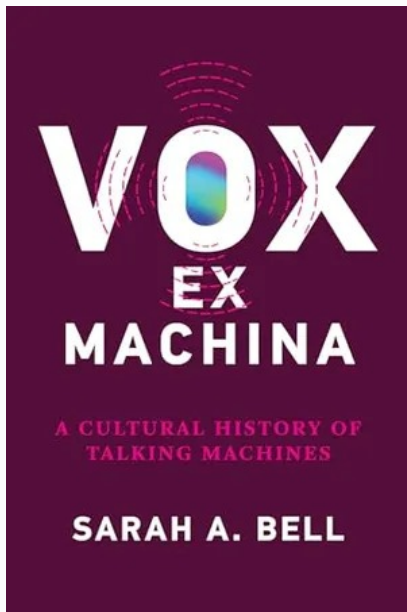
The difficulties were both technical and social. As one technical paper explained, “For high vowels and voiced consonants uttered by women, the first formant is often very close to the first fundamental of the voice source spectrum. This makes it harder to accurately measure the first formant.” This paper didn’t acknowledge limitations with the measurement instrument of spectrographs themselves, but it sited the problem in the female voice, and, by extension, women. It went on to speculate, “The range of possible voices for females are hemmed in on one side by male voices, on the other side by child voices. This indicates that listeners will be more critical toward a female synthetic voice than towards a male or even a child synthetic voice.”

In voice models, female voices were defined largely by pitch, falling between children and men. This was less a technical fact than it was a reflection of structural sexism that had primarily male researchers considering male bodies as vocal models, and, bizarrely, blaming women’s voices for being ill suited to the technology rather than the other way around. One count of the phonetic literature found that 40.5 percent of studies had assembled only male speakers, with the vast majority of the rest using more male than female or an even number of male and female participants. Less than 5 percent of studies focused on female speakers.

In the same way that it was claimed that women’s voices were not technically suited to recording and amplification technologies in the early 20th century, structural sexism limited the scope of some research and technical development that accepted this unsuitability as “natural.”

Meanwhile, recordings of female voices providing information and instructions in urban environments—public transportation and security announcements, vending and automatic checkout and teller machines—were increasingly common and chosen to forge what one scholar called a “soft coercion” in the

pitch of the neoliberal city. These are voices that tell you where to go, what to do and how to behave in order to move in an orderly way through the urban environment, and they are meant to maintain calm efficiency, not unlike Bitching Betty.



Vox ex Machina: A Cultural History of Talking Machines

How today's digital devices got their voices, and how we learned to listen to them

[Buy Now](#)

In November 1983, the *New York Times* published an editorial by sociologist Steven Leveen under the title “Technosexism.” Leveen noticed that there were “millions of mechanical objects” now speaking “through the new technology of speech synthesis,” including computers, clocks, elevators, automobiles, vending machines and even bathroom scales. He was concerned that they were perpetuating cultural stereotypes by “associating females with low-level service jobs, while associating males with tasks that are broader in range and higher in status.”

Leveen had done a little bit of research before writing his editorial. He was aware that synthesizing a higher-pitched voice was actually more “expensive,” that it required more data be stored “on a microchip,” and he was aware that

product developers were willing to absorb that cost because of market research. Most of the market research in Leveen's examples amounted to assumptions about gender roles gathered through interviewing mostly professional men. A video game developer: "Have you ever been to a baseball game with a female announcer?" An executive from National Semiconductor: "the [supermarket scanner] systems use exclusively female voices because the male voice ... sounded 'just a little bit strange.'" Coca-Cola vending distributors (mostly male): "felt the male voice was not as pleasing." And Chrysler incorporated a "male" voice into its 1983 cars because testers had stated that when a female voice told them their car's oil pressure was low, it "hit [them] the wrong way."

Leveen concluded that "it is not a coincidence ... that males are usually the ones purchasing the systems, and that they find female voices more desirable," although this preference was domain specific. His concern was that the gendered voice distribution between low-status and higher-status applications would "subtly influence our children's beliefs about which activities and careers are open to them."

Leveen's concerns about "technosexism" in the 1980s are often echoed in today's critiques of female-sounding voice assistant applications like Siri, Alexa and Cortana, all originally defaulted to female in the United States. While his argument didn't gain much traction at the time, it foreshadowed ongoing debates about gender and technology.

At this point, there is no shortage of articles that have posed some form of the question, "Why are voice assistants always female?" Nearly all of them cite the Stanford University communications scholar Clifford Nass, who by the 1990s and 2000s became the go-to expert on human-computer interaction, particularly in the realm of voice interfaces. In the 1990s, Nass' research agenda

focused on investigating computers as social actors, which led him to an interest in “interfaces that talked and listened.” In addition to directing a lab running experiments on how people reacted to computer applications with voices, Nass was consulting on voice interfaces for Microsoft, IBM, BMW, General Magic, Verizon and other corporations.

The experiments from Nass’ lab were based on a belief that human speech is an evolutionary adaptation. He reasoned that people “behave toward and make attributions about voice systems using the same rules and heuristics they would normally apply to other humans.” Although Nass attempted to account for the role of culture in his reasoning about gender and social identity, the experimental designs used in his lab’s research could not account for the influence of culture on his test subjects. There is, after all, no way to remove a person’s lifetime of cultural experience to test the degree to which preferences about gendered voices are hardwired or learned, or reflect some degree of both. Furthermore, in pursuing quantitative measures of people’s feelings about voices—how much they like, trust and find credible a certain synthesized voice—test subjects were typically classified simply as “male” and “female,” in a rigid two (gender of voice) by two (gender of participant) experimental design.

While Nass acknowledged sexist biases about voices—that people take male voices “more seriously” than female voices, for example—he ascribed these biases to “learned social behaviors and assumptions,” stating as fact that “each culture defines canonical behaviors for females and males.” He demonstrated his bias when he stated that even machine-like voices trigger “the brain’s obsessive focus on gender.” Stereotyping, Nass argued, was a cognitive shortcut to reduce information overload. While finding gender stereotypes “regrettable” and even describing them as “pernicious,” he concluded that they were evolutionarily ingrained and nevertheless justified their use in voice interfaces, aligning with business interests that capitalized on existing biases.

Some scientists, like phonetician and speech technology entrepreneur Caroline Henton, believed the solution lay in creating better female-sounding voices and argued that synthesized voices could challenge listener prejudices. In a 1999 article in the *Journal of the International Phonetic Association*, Henton echoed Klatt, attributing the lack of quality female-sounding synthesized voices to “insufficient data on female speech production” and “inadequacies in analytic hardware.” However, like Nass, Henton was primarily driven by commercial concerns. Since “the lack of appropriate (female or age-related) voices are the most commonly cited objections to not using [text-to-speech],” she reasoned that the future of speech technology was female. Henton would go on to be a phonetician at Apple, working on Siri.

At Bell Labs, Ann K. Syrdal was also working to create more natural-sounding female voice synthesis. Syrdal’s experience reflected some of the most active voice synthesis research in the U.S., completing her graduate work in psychology at the Center for Cognitive Sciences at the University of Minnesota, and then becoming an affiliate at Haskins Laboratories and spending a five-year National Institutes of Health grant period at the Research Laboratory of Electronics at MIT before joining the speech technology department at Bell Labs. Her Bell Labs team created what some specialists consider the first high-quality female-sounding synthesized voice, called Julia, which won an international competition in 1998.

Henton, Syrdal and others viewed the lack of knowledge about female-sounding voices and the lack of natural-sounding synthesized female voices as itself a form of technosexism. Scholars including Leveen, and more recently Yolande Strengers and Jenny Kennedy, have argued that female-sounding voice assistants are sexist, perpetuating white, middle-class, heteronormative fantasies about women’s compliance with men’s needs, as well as gendered labor

hierarchies. In broad strokes, these criticisms are warranted. The corporations that use voice synthesis for commercial purposes are invested in using voices that seem to correlate with the purchasing and other behaviors that they want to encourage. They have often followed Nass' advice in choosing differently gendered voices to match sexist expectations. Female-sounding voices are supposed to be calming and male-sounding voices more authoritative.

But over the last several years, there has been a shift toward often younger, male-sounding synthesized voices for many applications, including domestic and customer service assistants: In 2015, the U.K. grocery chain Tesco changed the voice of all its self-checkout machines from female to male; IBM's Watson modeled the vocal quality of the typical "Jeopardy!" winner—an educated white man in his mid-20s to 40s, and then became a "Jeopardy!" champion itself; Jibo, a social robot for the home, was supposed to be another member of the family, and developers chose a friendly and enthusiastic young adult male voice for it modeled on Michael J. Fox's performance of Marty McFly in the *Back to the Future* films; and Apple offers several voices for Siri, including male- and female-sounding voices with subtle characteristics of African American Vernacular English, and no longer defaults to the original female unless the user chooses it. According to the *Guinness Book of World Records*, the most downloaded navigation system voice before Google Maps became widely used for personal navigation was the animated oaf Homer Simpson, as voiced by Dan Castellaneta.

Despite this shift, giving a system a voice—whether a stereotypical "smart wife" or the dulcet tones of Morgan Freeman—reinforces the illusion that corporate informational interactions are personal and personal interactions are purely informational. Put another way, changing the sound of Siri's voice (something that is easy to do) doesn't change the fact that Siri is the "voice" of a U.S.-based technology corporation that manifests a great deal of power by controlling the

information collected and provided through Siri. Tech corporations prioritize using our biases for their benefit, while dismissing the reinforcement of stereotypes as a cultural problem rather than a technological one.

Of course, the cultural problem can *be* a technological problem. We learn to value the humanity of people that we perceive as different from ourselves through experience. As synthesized voices become common, replacing with networked technologies what might have previously been interactions with other people, we lose exposure to the vocal diversity and expressiveness of other human beings and risk losing some of our capacity to truly understand one another.

The temptation to simulate human expressiveness through technology only deepens this disconnect, opening the door to manipulation and deceit rather than fostering meaningful connection.

Sarah A. Bell is a writer and professor who studies the impacts of information technologies on society. She is the author of [Vox ex Machina: A Cultural History of Talking Machines](#), from which this article is adapted. An open access edition of the book is available for download [here](#).



A Note to our Readers

Smithsonian magazine participates in affiliate link advertising programs. If you purchase an item through these links, we receive a commission.

You Might Also Like

How Transgender Women Are Training Their Voices to Sound More Feminine

December 15, 2015

Bionic 'Pilots' Compete for the Gold at the Cybathlon

October 23, 2024

Enhanced Brain Implant Translates Stroke Survivor's Thoughts Into Nearly Instant Speech Using Artificial Intelligence

April 2, 2025

How Will Artificial Intelligence Help the Aging?

March 27, 2017

How Machines Are Getting Better at Making Conversation

April 27, 2016

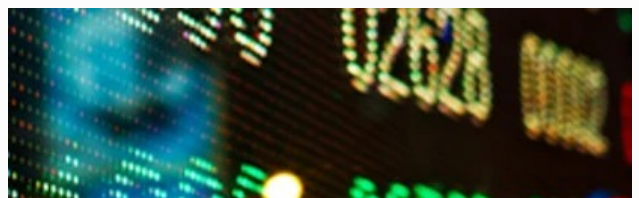
Get the latest stories in your inbox every weekday.

Sign Up

Email Powered by Salesforce Marketing Cloud ([Privacy Notice](#) / [Terms & Conditions](#))

Sponsored Content

dianomi





Wood Mackenzie

Power shifts: the new geopolitics of energy in Asia-Pacific



MarketViews

What is the latest on ETFs?



MarketViews

Stay updated with the newest gold market developments.



MarketViews

Get the latest commodity trends from industry experts.



Wood Mackenzie

Power shifts: the new geopolitics of energy in Asia-Pacific



MarketViews

What is the latest on ETFs?



Wood Mackenzie

Power shifts: the new geopolitics of energy in Asia-Pacific



MarketViews

What is the latest on ETFs?





Wood Mackenzie

Power shifts: the new geopolitics of energy in Asia-Pacific



MarketViews

The latest news on commodity markets from leading experts



MarketViews

What is the latest on ETFs?

More about:

Airplanes

Communication

Computer Science

Gender

Technology

Follow Us



Explore

Smart News
History
Science
Innovation

Subscription

Subscribe
Give a gift
Renew
Manage My Account

About

About Us
Advertising
Contact Us
Content Licensing

Arts & Culture
Travel
At the Smithsonian
Photo Contest
Podcast
Video

Newsletters

Sign Up

FAQ
Feedback
Internships & Employment
Member Services
Smithsonian Institution
Staff

Our Partners

Smithsonian.com
Smithsonian Store
Smithsonian Journeys
Smithsonian Channel
Smithsonian Books
Smithsonian Membership

© 2025 Smithsonian Magazine

Privacy Statement [5/20/25]

Cookie Policy [5/20/25]

Terms of

Use

Advertising Notice

Your Privacy Rights

Cookie Settings