

Voice Recognition Still Has Significant Race and Gender Biases

[Joan Palmiter Bajorek](#)

MAY 10, 2019



EVERSOFINE/GETTY IMAGES

Voice AI is becoming increasingly ubiquitous and powerful. Forecasts suggest that voice commerce will be [an \\$80 billion business by 2023](#). Google reports that [20% of their searches](#) are made by voice query today — a number that's predicted to climb to [50% by 2020](#). In 2017, Google announced that their speech recognition had a [95% accuracy rate](#). While that's an impressive number, it begs the question: 95% accurate for whom?

Speech recognition has [significant race and gender biases](#). As with [facial recognition](#), [web searches](#), and even [soap dispensers](#), speech recognition is another form of AI that performs worse for women and non-white people. To be clear, I do not believe that the creators of these systems set out to build racist or sexist products. It's doubtful these biases are intentional, but they are still problematic. The fact is that speech recognition understands white male voices well...but what about the rest of us?

Accuracy rates are more important than playing music. Speech recognition now influences important aspects of people's lives, including [immigration decisions](#), [job hiring](#), and transportation, among many other things. That means that speech recognition accuracy — or lack thereof — could prevent you from immigrating to a new country, getting a job, or traveling safely. Did you see that [episode of Silicon Valley](#) where a car drives someone to an abandoned island? It's funny on TV; not so funny in real life.

[Automakers have admitted for years](#) that their speech recognition doesn't work as well for women. [The recommended remedy](#) has been that women do extensive training ("Women could be taught to speak louder, and direct their voices towards the microphone...") that their male peers don't have to do. Same for minorities and people with non-standard accents. Seriously?

Recognition Accuracy by Gender and Race

[Research](#) by Dr. Tatman published by the North American Chapter of the Association for Computational Linguistics (NAACL) indicates that Google's speech recognition is 13% more accurate for men than it is for women. And Google is regularly the [highest performer](#) — as compared to Bing, AT&T, WIT, and IBM Watson systems.

So here's a thought experiment: Let's consider three Americans who all speak English as a first language. Say my friend Josh and I both use Google speech recognition. He might get 92% accuracy and I would get 79% accuracy. We're both white. If we read the same paragraph, he would need to fix about 8% of the transcription and I'd need to fix 21%. My mixed-race female friend, Jada, is likely to get [10% lower accuracy](#) than me. So, our scorecard would look something like:

Josh (white male) = A-, 92%

Joan (white female) = C+, 79%

Jada (mixed race female) = D+, 69%

Dialects also affect accuracy. For example, Indian English has a [78% accuracy rate](#) and Scottish English has a [53% accuracy rate](#). [Amazon](#) and [Google](#) teams are working to improve that accuracy, but the problem has not yet been solved.

Real World Consequences

These biases have serious consequences in people's life. For example, an Irish woman [failed a spoken English proficiency test](#) while trying to immigrate to Australia, despite being a highly-educated native speaker of English. She got a score of 74 out of 90 for oral fluency. Sounds eerily familiar, right? This score is most likely a failure of the system.

Why does this bias exist? Disparities exist because of the way we've structured our data analysis, databases, and machine learning. Similar to how [cameras are customized to photograph white faces](#), audio analysis [struggles with breathier and higher-pitched voices](#). The underlying reason may be that databases have [lots of white male data](#), and less data on female and minority voices. For example, TED Talks are frequently analyzed by speech scientists, and [70% of TED speakers are male](#).

AI is therefore set up to fail. Machine learning is a technique that finds patterns within data. When you use speech recognition, the system is answering the question "given this audio data, which words best map onto this data, given the patterns and data in the database?" If the database has mostly white male voices, it will not perform as well with data it sees infrequently, such as female and other more diverse voices.

This is absolutely a matter of social injustice. But if that alone doesn't convince companies to fix the problem, they should consider that the accuracy of speech recognition also affects customer purchasing decisions. I have affluent English-Spanish bilingual friends who have chosen not to buy smart fridges because they know that the fridges will not understand them. What other IoT devices would they buy if these devices actually understood them?

Melinda Gates, who frequently discusses [financial blind spots related to diversity](#), has said: "We [as a society] care about diversity, but we *really* care about how much money we make ... Women are [responsible for] 85% of consumer dollars spent. Women control 70% of financial decisions in the house. So, you're missing an opportunity... you're leaving money on the table."

Potential Solutions

As voice AI becomes more ubiquitous and powerful, this technology will affect our daily lives more and more. Let's work on building a world where everyone's voices are heard clearly.

What can companies do? Be more transparent about your voice statistics, and encourage competition in the area. For example, companies can report their accuracy rates for women and diverse speakers in their marketing and sales pitches. Is your target user a working class woman? Then cite how well your system understands that demographic. Remember that women and minorities have huge purchasing power — why wouldn't you want to solve this problem?

Lastly, it's something we all need to keep talking about. Because everyone deserves to have their voice heard.

[Joan Palmiter Bajorek](#) is the Founder of [Women in Voice](#). She's a speech technologist, linguist, and Ph.D. candidate at the University of Arizona.

This article is about [TECHNOLOGY](#)

[FOLLOW THIS TOPIC](#)

Related Topics: [DIVERSITY](#)