

# Speak to me!

## PROS AND CONS OF VOICE CONTROLLED DEVICES FOR CHILDREN AND ADULTS

David Kleeman

**The author discusses the advantages and disadvantages of voice as the navigation medium of smart speakers, phones, games, toys, apps and metaverse experiences.**

We're still not sure what "Sally" was, but there it was on our Alexa-generated shopping list. Clearly, one of us had used voice activation to add some item we needed at the grocery store, but "she" misheard it. It was a good reminder that smart speakers are only as smart as their ability to understand what we want, and that can be a challenge with little children's voices. The results can be ... alarming.

### SMART SPEAKER ACCESS RISING FOR KIDS

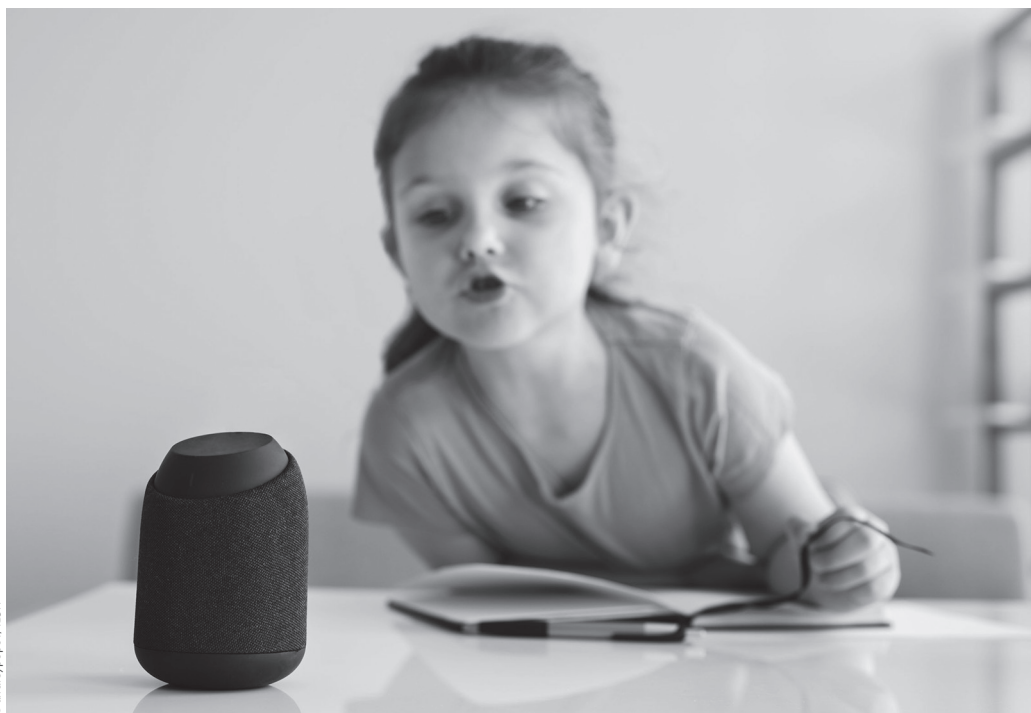
Voice activated devices continue to grow in presence in households worldwide. Across 16 countries surveyed by Dubit Trends<sup>1</sup> as of October 2021, 46% of children 2-15 had access to a smart speaker, with about a third of those owning one for themselves. In the US, access and ownership is at almost 56%, almost double what it was in 2018.

Likely the actual number is higher, given how many devices now have voice control, including smartphones.

More than 50% of US 2- to 15-year-olds now own a smartphone of their own, with around 43% ownership across the other surveyed countries.

59% of children in the US say they often or sometimes use voice search to find what they want to watch, play or listen to; this is up from 50% from October 2020. A stand-alone smart speaker is the most-used device for doing so, at 35%, followed by smartphone voice command (29%) and television remote (27%).

Listening to music is the primary application of a smart speaker, with listening to stories and podcasts and seeking information just behind.



© andreyppov/123rf

Ill. 1: Voice activated devices continue to grow in presence in households worldwide

### KID AND VOICE ACTIVATION: PLUSES AND MINUSES

Clearly, voice control has some advantages for children, particularly those who are pre-literate and so have trouble using text-based search or games, or those who are still mastering the fine motor skills of touchscreens. At the same time,

## RESEARCH

there are limitations to speech as an interface.

Smart speakers tend to give less helpful feedback than screen-based media, when a young user's question doesn't hit the mark. Generally, the device will either respond that it can't find anything matching the request, or it will return a solution based on what it "thought" it heard. With our Alexa at home, e.g., a request for a particular streaming news service frequently results in a playlist by a band with a vaguely-similar name.

When a screen is added (with text or with voice command), additional contextual clues are provided. I just now said to Alexa, "Peppa Pig," and it replied, "It sounds like you're asking about Peppa Pig; would you like to hear something about it?" An affirmative answer brings a wiki-like narration of the theme and voice actors for the series – not especially helpful to a 3-year-old. By contrast, a Google search of, simply, "Peppa Pig" return images (products and screenshots), video links, game links, websites, descriptions and more. Overwhelming, perhaps, to a toddler, but at least it offers multiple cues from which to launch successive approximations.

AI speech interpretation can be challenging even for non-kid voices. Recently, a joint US-India academic study (Ramesh et al., 2022; summarized in Simonite, 2022) found that 40% of more than 7,000 YouTube videos from popular children's channels had "taboo" words in the captions auto-generated from the program soundtracks.

Of course, it's still early days for voice interfaces, and they are improving as developers enhance the artificial intelligence. The more real-world experiences programmers have to analyze – a wide range of voices, accents, keywords, syntax, off-target frustrations, etc. – the better the AI will perform, growing its massive internal libraries to more quickly ascertain what a child wants.

## VOICE IN THE METAVERSE

As we move toward the metaverse (Kleeman, 2021), voice control will be extremely interesting. In immersive worlds, while much can be done via hand-held controllers, having to use a text interface for more complex activities or to move from place to place would break the illusion and the "suspension of disbelief." These worlds will be designed for social play, and so we will want to communicate with others by speaking.

In the metaverse, though, we'll have more to work with than just our words to generate the context that is currently lacking. In virtual reality, for example, the voice-activated AI could supplement what it hears with eye tracking, body positioning, and inference from the surroundings. Meta recently demoed a technology called "BuilderBot" that placed items in a virtual environment based on spoken words by users as well as contextual clues in the virtual space (Ghaffary, 2022).

We want to be certain that the metaverse is an inclusive and accessible space, and voice will play a strong role in delivering a satisfying experience for disabled and able-bodied users. Voice can be the navigation medium for those whose motion is restricted and prevents them from using hand controls or moving their real-world bodies in virtual spaces.

## SPEAK TO ME!

Remember when tablets first came out, and kids thought the entire world was made of interactive touchscreens? YouTube was full of videos with toddlers trying to swipe and tap books and magazines.

Perhaps sometime soon, little children will think they can command everything by talking to it. And, with more and more devices voice controlled –

smart speakers, smart phones, smart home systems, smart remotes and toys (also, I suppose, smart) – will they be wrong? ■

## NOTE

<sup>1</sup> *Dubit Trends is conducted every April and October in the USA (2,000 children/parents 2-18) and the UK (1,300 children/parents 2-18), with 18 other countries surveyed on a rotating basis (1,000 children/families 2-16). It is an online survey, representative of the national population profile and capturing ethnicity, household income, age of parents, and more. The study is syndicated to subscribers globally, and is used as the foundation of Dubit's research and strategy consulting as well as its development and production of games, apps and metaverse experiences.*

## REFERENCES

Ghaffary, Shirin (2022). Mark Zuckerberg wants to build a voice assistant that blows Alexa and Siri away. Vox recode (February 23, 2022). Available at: <https://www.vox.com/recode/22948097/mark-zuckerberg-voice-assistant-metaverse-ai-announcement> [28.4.22]

Kleeman, David (2021). What is the "metaverse" & why does it matter? Available at: <https://staging.kidscreen.com/2021/09/14/what-is-the-metaverse-why-does-it-matter/> [28.4.22]

Ramesh, Krithika, KhudaBukhsh, Ashiqur & Kumar, Sumeet (2022). "Beach" to "bitch": Inadvertent unsafe transcription of kids' content on YouTube. Preprint (February 17, 2022). Available at: [https://github.com/sumeetkr/UnsafeTranscriptionofKidsContent/blob/main/YouTube\\_Transcription\\_Final.pdf](https://github.com/sumeetkr/UnsafeTranscriptionofKidsContent/blob/main/YouTube_Transcription_Final.pdf) [28.4.22]

Simonite, Tom (2022). YouTube's captions insert explicit language in kids' videos. Wired (February 24, 2022). Available at: <https://www.wired.com/story/youtubes-captions-insert-explicit-language-kids-videos/> [28.4.22]

## THE AUTHOR

David Kleeman is Senior Vice President of Global Trends for Dubit, a strategy and research consultancy based in Leeds, England.

