Always On: Privacy Implications of Microphone-Enabled Devices

FUTURE OF PRIVACY FORUM

In collaboration with

EY Building a better working world

BY STACEY GRAY

APRIL 2016

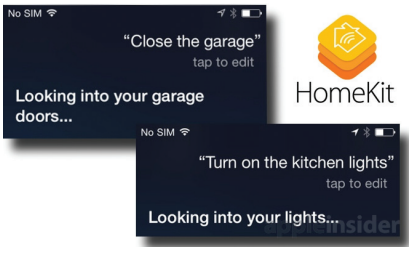
TABLE OF CONTENTS

ALWAYS ON: PRIVACY IMPLICATIONS OF MICROPHONE-ENABLED DEVICES

BY STACEY GRAY† | FUTURE OF PRIVACY FORUM

Is your Smart TV listening to your conversations? Are your children's toys spying on your family? These types of questions are increasingly raised as the next generation of internet-connected devices enter the market. Such devices, often dubbed "always on," include mobile phones, televisions, cars, toys, and home personal assistants—many of which are powered and enhanced by speech recognition technology. There is no doubt that the increasing prevalence of voice integration into everyday appliances enables companies to collect, store, analyze, and share increasing amounts of personal data. But what kinds of data are these devices actually collecting, when are they collecting it, and what are they doing with it?

This paper explores how speech recognition technology fits into a broader scheme of "always listening" technologies, discusses promising current and future applications, and identifies emerging practices by which manufacturers and developers can alleviate privacy concerns and build consumer trust in the ways that data is collected, stored, and analyzed.



We conclude that the colloquial term "always on" is often not an effective way to describe the range of technologies that use audio and video recording hardware. At one end of the spectrum, some devices (such as home security cameras) are designed to be always on. Many others utilize microphones, but are not necessarily always listening, recording, or even retaining information.

Instead, we propose three general categories of microphone-enabled devices:

(1) manually activated (requiring a press of a button, a flip of a switch, or other intentional physical action);

(2) speech activated (requiring a spoken "wake phrase"); and

(3) always on devices (devices, such as home security cameras, that are designed to constantly transmit data, including devices that "buffer" to allow the user to capture only the most recent period of time).

Each category presents different privacy implications, influenced in part by whether data is stored locally (an increasingly rare practice) or whether it is transmitted from the device to a third party or external cloud storage. Another key issue is whether the device is used for voice recognition, the biometric identification of an individual by the characteristics of her voice, or for speech recognition, the mere translation of voice into text. These are among the many factors, discussed in Part V, that must be assessed in order to evaluate potential privacy issues and determine appropriate notice, consent, and default frameworks.

† Stacey Gray is a Legal & Policy Fellow at the Future of Privacy Forum, a Washington, DC based center for privacy thought leadership and the advancement of responsible data practices. The author extends a sincere thank you to Jules Polonetsky, CEO, and staff at the Future of Privacy Forum, and to Ernst & Young for working with us on this important topic.

I. ADVANCES IN SPEECH RECOGNITION

Speech recognition—the ability to speak naturally and contextually with a computer system in order to execute commands or dictate language—used to be considered a dream of science fiction. But over the last forty years, speech recognition technology has improved dramatically. Although the technology is far from perfect—the accuracy is diminished by background noise and recording quality, and certain accents are often more easily understood than others[1] —consumers in 2016 can now interact reasonably well via speech with a range of devices. This includes waking up and asking, "what's on my calendar?" to calibrating a connected thermostat, to dictating a text message or starting a browser search with the likes of "OK, Google," "Hey, Siri," "Hi Alexa," or "Hey, Cortana…"

The benefits of speech recognition technology are undeniable: hands-free control of technology improves the lives of people with physical disabilities, makes healthcare and other professional services more efficient through accurate voice dictation, enhances automobile safety, and makes everyday tasks more convenient.

IMAGE BY: PIOTRUS

A key feature is that by sending data to the cloud, where powerful computing can be applied, speech recognition services can improve over time. Making use of the huge advancements in data processing in recent years, voice-to-text technologies can now adapt to your speech patterns over time and are getting better at understanding speech in context. This aspect led early voice recognition pioneer Raj Reddy to predict that voice recognition technologies would pass the Turing Test in our lifetimes.[3]

"The Star Trek computer is not just a metaphor that we use to explain to others what we're building . . . [it] is the ideal that we're aiming to build—the ideal version done realistically."

– Amit Singhal,
Google Senior VP and Software Engineer[2]

II. EMERGENCE OF PRIVACY CONCERNS AROUND MICROPHONE-ENABLED DEVICES

The same feature of speech recognition technology that makes it useful—its ability to bring voice control into our everyday lives—is the feature that is now understandably raising privacy concerns, as microphone-enabled devices become integrated into our homes and daily environments.

A variety of microphone-enabled devices and services have generated privacy concerns in recent years, in what MIT Technology Review has called "the Era of Ubiquitous Listening."[4] In 2014, Google's Chrome web browser came under fire for its pre-installed (but not automatically enabled) ability to passively listen for the words "OK, Google" to launch its voice-activated search function, leading the company to remove the feature from its open-source Chromium browser, and later, from Google Chrome all together.[5]

In 2015, citing similar concerns, privacy advocates complained to the FTC that Samsung's microphone-enabled SmartTV was "always on" in violation of federal wiretapping laws.[6] The complaint arose after users noticed that Samsung's Privacy Policy warned that sensitive conversations might be swept up and transmitted to third parties as part of the TV's voice controlled search function.[7] Despite Samsung's clarification that the TVs only recorded and transmitted information when the user pushed a button on the remote control to activate voice searching, [8] many advocates remained skeptical.

1 Speech recognition expert Marsal Gavaldà calls this diminished accuracy for children, seniors, and people with accents "the speech divide." CBC RADIO, *Here's why your phone can't understand your accent* (Sept. 13, 2015), http://www.cbc.ca/radio/spark/292-what-you-say-will-be-searched-why-recognition-systems-don-t-recognize-accents-and-more-1.3211777/here-s-why-your-phone-can-t-understand-your-accent-1.3222569; see also Daniela Hernandez, FUSION, *How voice recognition systems discriminate against people with accents* (Aug. 21, 2015), http://fusion.net/story/181498/speech-recognition-ai-equality/.

2 *See* Farhad Majoo, *iPhone 6S's Hands-Free Siri Is an Omen of the Future,* NEW YORK TIMES (Sept. 22, 2015) (quoting Singhal's comments made during an 2013 SXSW Interactive session entitled "The Future of Google Search in a Mobile World," available in video format at http://www.sxsw.com/interactive/news/2013/video-sxsw-2013-watch-amit-singhai-and-guy-kawasaki-talk-next-generation).

3 Xuedong Huang, James Baker & Raj Reddy, *A Historical Perspective of Speech Recognition*, COMMUNICATIONS OF THE ACM, Vol. 57 No. 1, Pages 94-103, available at http://cacm.acm.org/magazines/2014/1/170863-a-historical-perspective-of-speech-recognition/abstract.

4 David Talbot, *The Era of Ubiquitous Listening Dawns,* MIT TECHNOLOGY REVIEW (Aug. 8, 2013), http://www.technologyreview.com/news/517801/the-era-of-ubiquitous-listening-dawns/.

5 See Tony Bradley, *'OK Google' Feature Removed from Chrome Browser,* FORBES (Oct. 17, 2015), http://www.forbes.com/sites/tonybradley/2015/10/17/ok-google-feature-removed-from-chrome-browser/#16d299a44e27.

6 Electronic Privacy Information Center (EPIC), In the Matter of Samsung Electronics Co., Ltd., *Complaint, Request for Investigation, Injunction, and Other Relief* (submitted to the Federal Trade Commission, Feb. 24, 2015), available at https://epic.org/privacy/internet/ftc/Samsung/EPIC-FTC-Samsung.pdf.

7 Letter from Electronic Privacy Information Center (EPIC) to Attorney General Loretta Lynch and FTC Chairwoman Edith Ramirez (July 10, 2015), available at https://epic.org/privacy/internet/ftc/EPIC-Letter-FTC-AG-Always-On.pdf.

8 Samsung's Privacy Policy was modified to state: *"Samsung will collect your interactive voice commands only when you make a specific search request to the Smart TV by clicking the activation button either on the remote control or on your screen and speaking into the microphone on the remote control."* Samsung Newsroom, Samsung Smart TVs Do Not Monitor Living Room Conversations (Feb. 10, 2015), https://news.samsung.com/global/samsung-smart-tvs-do-not-monitor-living-room-conversations; see also Alex Hern, Samsung Rejects Concern over 'Orwellian' Privacy Policy, THE GUARDIAN (Feb. 9, 2015), http://www.theguardian.com/technology/2015/feb/09/samsung-rejects-concern-over-orwellian-privacy-policy.

Similarly, later in 2015, advocates cited Mattel's "Hello Barbie" as an example of a microphone-enabled device that supposedly brought the specter of surveillance.[9] The Wifi-connected doll, which follows a pre-set script ("What's your favorite color?") and uses speech recognition to respond to simple answers ("Orange is outstanding!"), undoubtedly has unique implications for children. Yet due to its technical and processing limitations, as explained below, it is unlikely to be as effective at pervasive data collection as some have predicted.

In most of these contexts, what critics have called bugs were viewed by others as valuable features, often core selling points of the devices. Speech recognition expands the world of possibilities for meaningful interaction and engagement through our devices. But in order to advance beyond the most rudimentary commands—and to become more accurate over time—speech recognition relies on third party translation and cloud storage. Furthermore, in order for a speech-activated device to be truly useful to a person who (for any reason) cannot use her hands to turn it on in the first place, it can become invaluable to eliminate the intermediate step of turning the device on manually, by introducing a "wake phrase."

III. DISTINGUISHING BETWEEN ACTIVE AND PASSIVE LISTENING

For the technological reasons to be discussed, it is inaccurate to classify all devices with speech recognition or microphone-enabled features as being "always on." Instead, such devices may be more aptly place into three broad categories, with some being capable of more than one function:

(1) Manually activated speech recognition devices are straightforward: the user presses a button or flips a switch, and the microphone turns on and begins recording and transmitting audio to a voice-to-text translation service, often (but not always) resulting in text appearing simultaneously on the device.

In contrast, **(2) speech activated** devices use the power of energy efficient processors to remain in an inert state of passive processing, or "listening," for a pre-set "wake phrase." The device buffers and re-records locally, without transmitting or storing any information, until it detects the word or phrase that triggers the device to begin actively recording.

The ability of devices to use the microphone to listen for a "wake phrase" is made possible by the same hardware advancements that allow many other mobile sensors to be "always on," such as the gyroscope and accelerometer (to enable fitness tracking), or the compass and GPS (to enable geo-location).

Traditionally, devices had to be awake—that is, accessing the main CPU—in order to process sensory input. That is why many apps that depend on constant input, like the Sleep Cycle App (which uses a mobile device's microphone to assess sleep quality) typically need to remain plugged in to avoid draining the phone's battery. Beginning around 2008, manufacturers began to introduce more energy efficient co-processors into their devices. This enabled the devices to be constantly analyzing sensory information locally (without transmitting data from the device) without draining battery life.[10]

As a result, when a modern smartphone is in a passive state (i.e. asleep, or with the microphone-enabled app in the background), the microphone can still internally (locally) process short stretches of audio, buffering and re-recording every few seconds to detect the device's wake phrase. In other words, it does not record or retain any audio data, or begin to transmit any data until it is "woken up."[11] In this sense, then, it is not really "listening" to its environment, but instead utilizing the microphone as just another environmental sensor.

Finally, **(3) always on** devices are those designed to record and transmit data all of the time.

Most prominently, this includes home security cameras and baby monitors, but also includes a range of new devices, such as the Kapture[12] (a wristband that records audio constantly, buffering every 60 seconds such that the user can capture and save conversational snapshots from daily life) or the OrCam[13] (a wearable video camera, designed for the visually impaired, that translates text to audio in real time). Cities can now detect gunfire via microphone networks, and there are microphones that can detect termite infestations by listening to audio outside of the range of the human ear. [15] [16]

9 *See supra,* note 7.

10 *See generally,* e.g., Tom Kevan, *Always-On Sensing Changes Everything,* SENSORS MAG (Oct. 11, 2013), http://www.sensorsmag.com/sensors-mag/always-sensing-changes-everything-11949.

11 Concerns about remote users who may bypass device controls are reasonable. Although threats of hacking and surveillance are outside the scope of this paper, security will always be a legitimate concern for users of microphone and video-enabled devices. See, e.g., Shodan, https://www.shodan.io/ (an online search engine for unsecured video cameras).

12 *Kapture Audio-Recording Wristband Device,* http://kaptureaudio.com/ (last accessed Mar. 1, 2015).

13 *OrCam – See for Yourself,* http://www.orcam.com/ (last accessed Mar. 1, 2016).

14 *See* Richard Chang, Sacramento police deploy microphones to listen for gunshots, THE SACRAMENTO BEE (July 30, 2015), http://www.sacbee.com/news/local/crime/article29604628.html

15 *HomeSafe Home Services,* http://www.homesafeinspection.com/index.php/licensing/for-pest-inspectors (last accessed Mar. 1, 2016) (describing "high-powered, state-of-the-art infrared and acoustic technologies . . . which enable pest control operators to, in effect, 'see' and 'hear' through a house's walls, floors and ceilings.").

16 Notably, in the public sphere, this category includes body-worn cameras increasingly used by police departments. Although law enforcement is outside the scope of this discussion, this particular example has unique privacy implications as well as potential civil rights benefits. For a discussion of these issues, see JAY STANLEY, AMERICAN CIVIL LIBERTIES UNION, POLICE BODY-MOUNT CAMERAS: WITH RIGHT POLICIES IN PLACE, A WIN FOR ALL (most recent version published March 2015), available at https://www.aclu.org/police-body-mounted-cameras-right-policies-place-win-all.

Categories of Microphone-Enabled Devices		
	Description	**Selected Examples**
Manually Activated	Devices **begin** recording and transmitting audio only when manually switched on (by remote or button) and **stop** recording automatically or when the button or remote is released.	**Samsung TV** **LG Smart TV** **Sony Android TV** **Apple TV** **Fire TV** **Hello, Barbie**
Speech Activated	Devices **begin** recording and transmitting audio only after the microphone detects a "key word" and **stop** recording automatically after a short amount of time. Until then, they remain in an inert state of buffering and re-recording, allowing the microphone to passively "listen" for a key word without recording or transmitting information.	**Amazon Echo ("Alexa" or "Amazon")** **iPhone 6S ("Hey, Siri")** **Google Chrome ("OK, Google")** **Microsoft Cortana ("Hey, Cortana")** **Motorola X Phone (customizable)**
Always On	Devices **begin** recording and transmitting audio when turned on, and are designed to **continue** recording and transmitting data 100% of the time or until manually turned off.	**Nest Cam** **Baby monitors** **Kapture** **OrCam**

These devices, because they are designed to be always on, evoke different privacy concerns from those that are manually or speech activated, and call for notice and consent frameworks in sync with the more extensive data collection that they enable.

As discussed below, microphone-enabled devices (whether manual or speech activated) are more limited in the scope of their privacy implications than devices that are designed to be always on. In fact, the ability of devices such as televisions and home personal assistants to be activated using a spoken command, rather than a push of a button, is often a helpful step towards integration of speech functionality into everyday life.

In the next few years, we will likely see an increase in the flexibility of "wake phrases"—for instance, while Apple iOS 9 retains the classic "Hey, Siri," Motorola permits users to generate their own 3 to 5 syllable "launch phrase."[17] This hands-free functionality is a game-changer for anyone with a physical disability, as well as for professionals that need to access software hands-free (e.g. surgeons), and consumers who seek the functionality of hands-free engagement with their devices.

Another benefit will be contextual awareness—the ability of the device to adjust itself in accordance with the environment. For instance, a phone's microphone can detect when you're in a crowded, noisy situation and adjust its ring volume accordingly, without the need to record, transmit, or save audio. In this sense, use of the microphone is again similar to use of other environmental sensors, such as the gyroscope or accelerometer, to allow devices to adjust to their surroundings in useful ways.

IV. PRIVACY IMPLICATIONS WILL VARY BY SOCIAL AND LEGAL CONTEXT

Despite the fact that many devices dubbed "always on" are in fact only using the microphone to detect a wake phrase, the fact remains that microphones and specifically voice data retain unique social and legal significance. In some instances, laws that protect biometric information may apply. In general, sector-specific laws and regulations will also apply on the basis of the content of the voice communications.

Biometric Identification

The collection of certain voice characteristics for the purpose of recognizing an individual currently implicates a range of laws. At the federal level, a "voice print" is considered either a biometric or personal record in the context of the Privacy Act,[18] FERPA,[19] and

17 As users of the Moto X (2d gen) discovered, Moto Voice can also be launched using a whistled tune. *See* Kellex, *Moto X Tip: Use a Whistle Instead of a Cheesy Phrase to Launch Moto Voice,* DROIDLIFE (Dec. 4, 2014), http://www.droid-life.com/2014/12/04/moto-x-tip-use-a-whistle-instead-of-a-cheesy-phrase-to-launch-moto-voice/.

18 22 C.F.R. § 308.3 ("Record means any document, collection, or grouping of information about an individual maintained by the agency, including but not limited to . . . any other personal information which contains . . . a finger or voiceprint.").

19 34 C.F.R. § 99.3 ("Biometric record, as used in the definition of personally identifiable information, means a record of one or more measurable biological or behavioral characteristics that can be used for automated recognition of an individual. Examples include . . . voiceprints").

HIPAA,[20] and thus subjected to greater regulatory restrictions. Similarly, several states have expanded their legal definitions of personally identifiable information in certain identity theft or breach notification laws to include some form of biometrics.[21] Two states, Illinois and Texas, have broad-reaching statutes that cover biometric data in commercial contexts, and strictly curtail its use.[22] While some ambiguity currently exists in distinguishing between the record itself—say, a photograph of a face, or an audio file of a voice—and the use of that record for biometric purposes,[23] industries collecting voice data would be well served to be aware of the growing body of laws and regulations around biometric identification.

However, there is an important difference between speech recognition and voice recognition—the latter indicating biometric identification. The majority of speech enabled devices on the market today are not designed for the purpose of uniquely identifying a person through the biometric characteristics of her voice. Instead, they aim to create products for which speech is a useful interface for engagement. In the future, however, it can be foreseen that unique voice recognition might become a useful consumer tool—for example, to permit only a specific person to access a device, or to enable parental controls by distinguishing between user accounts. Companies considering adding such features should be aware of the growing body of federal and state laws regarding biometric identification.

One and Two-Party Consent

When considering microphone-enabled devices, such as security cameras or audio recording devices, both users and manufacturers should be aware of potentially applicable anti-surveillance statutes. Federally, the Wiretap Act prohibits the intentional interception of the contents of any wire, oral, or electronic communication without the prior consent of at least one of the parties.[24] Most states have similar statutes, such that conversations in private settings may be lawfully recorded so long as one party (usually the party doing the recording) has consented.[25]

In twelve states, however—California, Connecticut, Florida, Illinois, Maryland, Massachusetts, Michigan, Montana, Nevada, New Hampshire, Pennsylvania, and Washington—it is only permissible to record a private conversation if all parties to the conversation have given their prior consent (so-called "two party consent" laws).[26] In these states, the question of whether a user risks violating an anti-surveillance statute usually turns on whether the communication being recorded is confidential, such that one of the parties has a reasonable expectation that no one is listening in (excluding, for example, recordings in public spaces).[27] If the conversation is confidential, then all parties must give consent, although consent can often be implied from the surrounding circumstances.[28]

Not all of these two-party consent laws are identical. Massachusetts, for instance, makes it a crime to "secretly" record a conversation.[29] Although two of these state statutes—Connecticut and Nevada—apply only to "wire" or telephonic conversations, most state laws additionally apply to oral or in-person conversations. Most laws also include the requirement that a recording be "intentional" (or "purposeful," or "willful"), but some do not. And like their federal counterpart, most also contain a variety of exceptions, including for communication service providers or "common carriers."[30]

As a result of the variety in applicable state laws, manufacturers will be wise to be aware of the legal landscape and design devices to assist users in avoiding legal complications. For example, many video recording devices have the option to disable the microphone. Similarly, devices may be designed with prominent visual cues to alert passersby to their recording functionality. See infra, Part V(4) (discussing prominent visual cues).

20 45 C.F.R. § 164.514 (listing "[b]iometric identifiers, including finger and voice prints" as examples of personal information that must be removed from a data set before that data set can be considered properly de-identified and thus no longer subject to HIPAA regulations).

21 *See, e.g.,* Conn. Gen. Stat. § 38a-999b; Iowa Code § 715C.1; Neb. Rev. Stat. § 87-802; N.C. Gen. Stat. § 75-66; Or. Rev. Stat. § 165.800; Or. Rev. Stat. § 336.184 (regulating student educational records); Wis. Stat. § 943.201; Wyo. Stat. § 6-3-901.

22 Illinois Biometric Information Privacy Act, 740 Ill. Comp. Stat. 14/; Tex. Bus. & Com. Code § 503.001. *See also* Fla. Stat. § 1002.222 (prohibiting the collection of biometric information by any state educational institution or agency).

23 Facebook, Google, and Shutterfly have all been targeted by litigation under the Illinois Biometric Information Privacy Act (BIPA) over the issue of their facial recognition technologies. *See* Alex Perala, *Google the Latest to Run Up Against Illinois Biometrics Law,* FindBiometrics (Mar. 7, 2016), http://findbiometrics.com/google-the-latest-to-run-up-against-illinois-biometrics-law-303074/.

24. 18 U.S. Code § 2511.

25. *See generally,* Digital Media Law Project, *Recording Phone Calls and Conversations,* http://www.dmlp.org/legal-guide/recording-phone-calls-and-conversations (last accessed Mar. 1, 2016).

26. *Id.* Cal. Penal Code § 632; Conn Ge. Stat. § 52-570d (2016) (applying only to telephonic conversations); Fla. Stat.§ 934.03 (2016); 720 Ill. Comp. Stat. 5/14-1 et seq (2016); Md. Code, CTS & Jud. Proc. § 10-402 (West 2016); Mass. Gen Law ch. 272, § 99 (2016); Mich. Comp Laws § 750.539c (2016); Mont. Code Ann. § 45-8-213 (2016); N.H. Rev. Stat. § 200.620 (2015) (applying only to "wire" or telephonic conversations); N.H. Rev. Stat. Ann. § 570-A:2 (2016); 18 PA. Cons Stat. § 5703 (2016); Wash. Rev. Code § 9.73.030 (2015).

27. *See, e.g.,* Stevenson v. State, App. 1 Dist., 667 So.2d 410 (1996) (finding Florida anti-surveillance statute inapplicable because defendant had no reasonable expectation of privacy in conversation which took place outside van stopped in public roadway).

28. Implied consent is a highly fact-specific question that requires consideration of all of the surrounding circumstances to a recording. *See, e.g.,* Berry v. Funk, 146 F.3d 1003, 1011 (D.C. Cir. 1998) ("[W]ithout actual notice, consent can only be implied when the surrounding circumstances convincingly show that the party knew about and consented to the interception.")

29. Mass. Gen. Laws ch. 272, § 99.

30. *See supra,* note 26.

Employment and Workplaces

Generally speaking, workers in the United States have a lower expectation of privacy in the context of an employer-employee relationship, and workplace monitoring is fairly **commonplace.**[31] Nonetheless, direct surveillance of voice communications, without notice and if unrelated to legitimate business purposes, may run afoul of federal and state laws.[32] As microphone-enabled devices make monitoring of employees' conversations easier, expectations around the appropriate use of these devices in the workplace may shift. In contrast, international norms around workplace privacy are often much more protective of employees.[33]

Hospitals and Medical Environments

The use of Smart TVs and other microphone-enabled smart devices is already beginning to be commonplace in hospital settings. Many hospitals, for example, equip patient rooms with Smart TVs to allow patients to benefit both from the entertainment and the ability to receive health-related instructional materials designed to reduce readmission.[34] In a hospital or a longer-term assisted living facility, it's easy to imagine how a speech-activated device enabled with speech recognition features can enable higher quality of life and improved recovery.

Nonetheless, because health information is specifically regulated by the HIPAA Privacy and Security Rules, hospitals and other facilities will be obligated to meet high standards of data security and to protect voice data with the same protections as other covered health records.[35] Similarly, if a person's voice is used for biometric identification, i.e. by generating a "voice print," this identifier must be removed from data sets in order for protected health information to be considered de-identified.[36]

Homes (Historically Protected Spaces)

Under the auspices of the Fourth Amendment, the home has historically been considered a sacred space, embedded with a higher expectation of privacy against government intrusion.[37] However, under the "third party doctrine" arising in the twentieth century, information shared with third parties loses its private status under the assumption that there is no reasonable expectation of privacy in information shared with the outside world.[38] As sensor-embedded devices begin to integrate into today's Smart Home, it becomes increasingly possible that courts will be unable to reconcile the third-party doctrine with the historical notion of the home as a constitutional sanctuary. Until judicial solutions are reached, the distinction between local and external processing (discussed below, Part V) may be of particular importance.

V. EMERGING PRIVACY QUESTIONS AND BEST PRACTICES

In determining the appropriate framework of privacy protections around a device, manufacturers should keep in mind the utility of speech recognition features, and whether the device is one that uses the microphone as an essential feature of the device. For example, a device like a television, which for most users does not require the microphone in order to perform its essential functions, may evoke an entirely different set of expectations than a device like the Amazon Echo, for which speech activation and speech recognition are clearly the core features of the device. In all cases, manufacturers should emphasize user awareness, consent-based features, and control over the device.

The following are key privacy questions to consider, with examples of some emerging privacy-conscious practices where relevant:

(1) Does processing and storage occur locally or externally (i.e. cloud-based)?

Cloud storage and computing bring huge value to microphone-enabled devices, not only in cost savings and accessibility, but in improving speech recognition by permitting a device to adapt to a person's speech patterns over time. Nonetheless, for many consumers, understanding when a device is transmitting and storing data externally is of great importance for reasons involving security, law enforcement access, future use, or retention. For this reason, we may begin to see greater market emphasis of local processing and storage as a selling point for privacy-conscious consumers[39] (see Fig. 1), as well as a growing awareness of the implications of cloud storage. On the other hand, for many others the benefits of cloud storage and computing may prove

31 *See generally,* Privacy Rights Clearinghouse, Fact Sheet 7: Workplace Privacy and Employee Monitoring (rev. Jan. 2016), https://www.privacyrights.org/workplace-privacy-and-employee-monitoring (last accessed Mar. 1, 2016).

32 *Id.*

33 *See generally,* Tim Wybitul, Part 11: *Data Protection in the Workplace,* HOGAN LOVELLS CHRONICLE OF DATE PROTECTION (July 1, 2015), http://www.hldataprotection.com/2015/07/articles/international-eu-privacy/part-11-data-protection-in-the-workplace/.

34 *See, e.g.,* Tom Foley, *Transforming Health—The Smart Patient Room,* CDW HEALTHCARE (Oct. 5, 2015), http://www.cdwcommunit.com/perspectives/expert-perspectives/transforming-health-the-smart-patient-room/; Megan Headley, *Today's Trends in Healthcare Televisions,* TELEHEALTH SERVICES (July 31, 2014), http://www.telehealth.com/sites/default/files/31_TodaysTrendsinHealthcareTelevisions.pdf.

35 Under HIPAA, health information means "any information . . . ***whether oral or recorded in any form or medium,*** that: (1) Is created or received by a health care provider, health plan, public health authority, employer, life insurer, school or university, or health care clearinghouse; and (2) Relates to the past, present, or future physical or mental health or condition of an individual; the provision of health care to an individual; or the past, present, or future payment for the provision of health care to an individual." 45 C.F.R. § 160.103 (emphasis added).

36 45 C.F.R. § 164.514(b)(2)(i)(P).

37 *See, e.g.,* Miller v. United States, 357 U.S. 301, 307 (1958) (quoting the oft-cited statement attributed to William Pitt from a 1763 address to Parliament: *"The poorest man may in his cottage bid defiance to all the force of the crown. It may be frail—its roof may shake—the wind may blow through it—the storm may enter, the rain may enter—but the King of England cannot enter—all his force dares not cross the threshold of the ruined tenement."*).

38 *See generally,* Marley Degner, Riley and the Third-Party Doctrine, 32 WESTLAW JOURNAL COMPUTER AND INTERNET 1 (2015).

compelling. Transparency around whether and when devices transmit data externally will help build consumer trust.

Fig. 1. Image of the packaging of Fisher-Price's "Smart Toy".

Although some devices may incorporate localized speech recognition exclusively—for example, a television that processes simple voice commands ("volume up," "volume down")—other microphone-enabled devices may use local processing for discrete functions while relying on cloud-based processing for others. For example, when speech-activated devices use the microphone to passively listen for a wake phrase, they are using a limited form of local processing, and not transmitting or storing data. Once "woken up," of course, they begin to transmit audio outside of the device. Companies can build consumer trust by promoting a clear understanding of this boundary through prominent, reader-friendly privacy explanations.

(2) Does the device arrive with speech recognition, or other audio recording functionality, pre-enabled?

The question of whether a device should arrive "out of the box" with audio recording or speech recognition functionality enabled will depend on the consumer's reasonable expectations of the default capabilities of the device. As speech recognition becomes increasingly more integrated into our lives, these user expectations will evolve. Nonetheless, expectations do exist and are often dependent on context and the nature of the device. For example, it would be onerous to require today's users, upon purchasing a new mobile phone, to go into the settings and turn on the microphone—because a phone is obviously designed with the microphone as a core feature of the device. Similarly, there is most likely no cause for concern if a device like the Amazon Echo arrives with speech recognition enabled, because it is being marketed as a "virtual personal assistant," obviously designed to interact through a voice interface.

In many other contexts, enhanced notice and choices about speech recognition will be the better path. This will be especially important as voice control becomes a new way to command devices that have traditionally been manually operated, such as cars and televisions. Many companies have already taken the privacy-conscious step of asking users to opt in: for instance, Samsung ships its Smart TVs with the speech activation feature disabled, requiring users to affirmatively enable it in the settings.[40] For others, enhanced notice might take the form of prominent privacy explanations during initial set-up. As speech recognition becomes integrated into more aspects of life, the appropriateness of this kind of enhanced notice will evolve.

(3) Does the device contain a hard on/off switch that can disable the microphone?

In devices that utilize the microphone to passively listen for a "wake phrase," it may be both useful and privacy-conscious to provide a way to manually disable the microphone. Primarily, such a feature is one of convenience: with the ubiquity of speech in daily lives, a hard "off" eliminates the possibility that the device will activate at inconvenient or unintended times.[41] To address such concerns, the Amazon Echo was not only designed with more than one possible wake phrase ("Alexa" or "Amazon") but with a hard "mute" button, permitting users to easily de-activate the microphone without having to un-plug the device. If this kind of function is directly tied to the hardware of the microphone, it can help to also alleviate concerns, around surveillance or infiltration. Of course, such a feature would not be appropriate in all circumstances, making little sense, for example, on a baby monitor that would typically be turned off entirely.

(4) Does the device provide visual cues that clearly indicate when it is recording and/or transmitting information?

The core principles of trust and informed consent dictate that users should understand when a device is on and recording, and many companies have incorporated prominent visual cues into their devices when they are recording. As an example of this, the Hello Barbie has a series of distinct visual cues that inform the user whether the doll is listening, transmitting, or looking for a Wi-Fi signal, based on the color and pattern of the LED lights in the doll's necklace.

Fig. 2. Mattel's Hello Barbie, released 2015.

Similarly, visual cues can be built directly into the form of the device: for instance, the Kapture, a wearable wristband with an embedded microphone, was purposely designed in bright colors with a microphone-like gridded appearance for the purpose of being obvious to users and observers.[42]

The importance of such cues and the level of their appropriate prominence will vary depending on context, user expectations, and form factor. Visual cues may not be as necessary on

39 *See, e.g.,* Kickstarter Campaign, *Sense: The intelligent camera and hub for your modern home,* https://www.kickstarter.com/projects/gal/sense-personalized-intelligence-for-your-connected (stating that it will use local processing to "make it impossible for anyone not in possession of your smartphone, or Sense, to view your private moments.") (last accessed Mar. 1, 2016).

40 *See* Samsung Newsroom, *Samsung Smart TVs Do Not Monitor Living Room Conversations* (Feb. 10, 2015), https://news.samsung.com/global/samsung-smart-tvs-do-not-monitor-living-room-conversations.

41 In one humorous example of this sort of inconvenience, a recent NPR radio news segment about the Amazon Echo accidentally activated a number of home devices in homes where the Echo was present and word "Alexa" spoken on the radio was indistinguishable from a person speaking the word in the room. *See* NPR, *Listen Up: Your AI Assistant Goes Crazy for NPR, Too* (Mar. 6, 2016), http://www.npr.org/2016/03/06/469383361/listen-up-your-ai-assistant-goes-crazy-for-npr-too.

42 *See* Kapture User Guide (one-page pamphlet that shipped with device in March 2016) (on file with author).

some devices, such as handheld microphones, that are clearly designed with the sole purpose of recording. The form factor of a device may also make visual cues on certain devices impractical because of the size or shape of the device. Many devices, however, like home assistants, may still be relatively new to many passersby, and as a result would be better served by lights or other prominent indicators.

(5) Use limitation is appropriate to alleviate concerns over misuse of audio.

The principle that companies should not collect or disclose recorded voice data (or text translations) beyond what the user reasonably expects applies broadly to many forms of data collection. Many consumers feel more strongly about voice information, in part because of the thoughts of wiretapping and surveillance that such information evokes. Even in the commercial sphere, however, microphone-embedded devices can transmit a wealth of information both about content of communications and potential biometric patterns in a person's voice. As a result of this unique status of voice, it will be incumbent on forward-looking companies to use voice data in ways limited to what is necessary and reasonably expected. For example, if a device purports to engage in speech recognition (translation of voice-to-text), it should be reliably clear to the user that the company is not analyzing audio files of voices to detect user identity.

One way to permit product improvement, as well as general research and analysis, while protecting individual privacy is to de-link and aggregate voice files from their original accounts. As an example, Apple collects and stores information from audio searches made via Siri, its mobile personal assistant. When a mobile user makes a request to Siri, the audio file is sent to Apple's servers using a random identifier (rather than the Apple ID, MAC address, or other persistent identifier).[43] This identifier can be re-set at any time, and after a period of time, the identifier itself is dis-associated with the audio file, leaving no remaining connection between the audio and the original account. Although retention even after this de-linking need not be indefinite, this privacy-conscious step can permit longer-term analysis while making it impossible for voice data to be linked to an individual.

(6) Ability to access and delete stored audio files will build consumer trust.

For some devices, a main selling point is the ability to improve over time by adapting to a user's preferences and habits, a feature that requires keeping audio files correlated with an individual account. For example, the Amazon Echo adapts to a user's speech patterns (e.g. timbre or accent) in order to get better at understanding voice commands. When this is the case, a strong step to build consumer trust is to permit that user to access the audio files and delete them. In particular, it makes sense to allow users to delete all audio files at once, since the nature of many microphone-enabled devices is that voice interaction becomes ubiquitous and quickly too impractical to delete files individually.[44]

When voice data is deleted, it is most likely reasonably understood that companies will often retain text translations in de-linked or de-identified aggregate forms, for purposes such as product improvement. However, unless companies provide other notification, users should be confident that their deletion has real effect: that an audio file of their voice is no longer in existence, and the text translation thereof is no longer correlated with their account.

(7) The difference between far-field or near-field microphone technology will influence appropriate privacy frameworks.

The appropriate privacy controls may differ between far-field listening technology and near-field listening technology. Consumers have different expectations between these two types of technology, understanding for example that a microphone adapted to a near-field range (such as a mobile phone) is unlikely to capture a conversation in the next room. In contrast, a company that has designed a microphone-embedded device for a specific far-field listening purpose—e.g. to detect gunfire or termite infestation—may alleviate privacy concerns by making it impossible for that device to detect near-field audio or ranges at the level of the human voice. This distinction may create one way for companies to use the capacities of the technology itself to design products that protect privacy.

CONCLUSION

The integration of speech recognition into our lives will bring an array of benefits, both for the day-to-day convenience of consumers, as well as in professional settings and in enabling vast improvements in quality of life for people in hospital care or living with physical disabilities. Nonetheless, moving forward it will be important to recognize that voice data is unique in its historical protection, communicative content, and biometric features. As we enter 2016, useful guiding principles are beginning to emerge,[45] and the conversation will continue to evolve on this subject as social norms shift about when and where we should expect to be able to speak to our devices. In considering the benefits of speech-enabled devices in parallel to their legitimate privacy implications, forward-looking companies will be well-served to use the power of technology itself to enable the power of speech recognition while protecting consumer privacy and control.

43 *See* Apple Privacy, http://www.apple.com/privacy/approach-to-privacy/ (last accessed Mar. 1, 2016).

44 By way of example, as of this writing, the ToyTalk dashboard for Hello Barbie does not permit mass deletion, a structural feature which means that a parent who wished to delete their child's voice data would be compelled to go through each short audio file one at a time, a task which quickly becomes impractical to the point of impossibility.

45 *See* Alta Associates' Executive Women's Forum, Voice Privacy Guiding Principles (March 2016), http://c.ymcdn.com/sites/www.ewf-usa.com/resource/collection/CAA076AF-9566-4E1E-9F07-6421154DE0EA/Voice_Privacy_Guiding_Principles_Public_(final).pdf.

In collaboration with

1400 EYE STREET, NW | SUITE 450 | WASHINGTON, DC 20005 • FPF.ORG

ALWAYS ON: PRIVACY IMPLICATIONS OF MICROPHONE-ENABLED DEVICES

BY STACEY GRAY

APRIL 2016