Voice Activated IoT Devices for Healthcare: Design Challenges and Emerging Applications

Petros Spachos, *Senior Member, IEEE*, Stefano Gregori, *Senior Member, IEEE*, and M. Jamal Deen, *Life Fellow, IEEE*

***Abstract*—The recent pandemic forced substantial changes in our lives, including the way we interact with physical objects. For example, voice-activated systems that enable users to communicate with them through speech commands are becoming more pervasive. At the same time, recent technology developments delivered voice capability to Internet of Things (IoT) devices with low-power audio transducers. Voice-activated IoT devices have the potential to engage patients and caregivers in new and cost-efficient ways, from telehealth and digital health, to portable diagnostics and remotely delivered care. In this brief, we review voice activated IoT devices, discuss their trends, and identify unique challenges when these devices are used in the healthcare sector. Furthermore, we discuss some future application scenarios and their characteristics.**

***Index Terms*—Internet of Things (IoT), IoT healthcare, mobile health, smart healthcare device, speech recognition, voice-activated device, voice-activated IoT healthcare device (VIHD).**

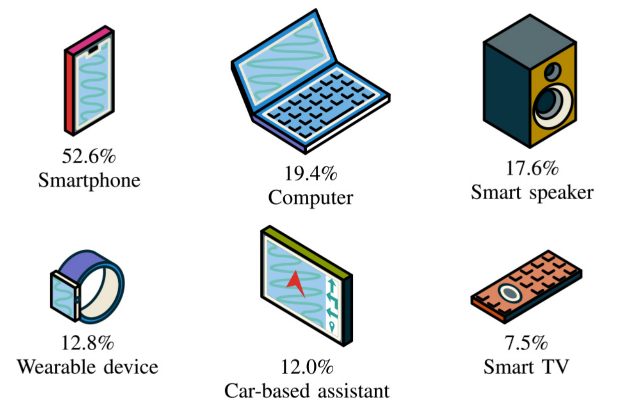

Fig. 1. Devices that consumers use for voice searches related to healthcare [10].

I. Introduction

VOICE technology has improved from limited vocabulary input systems to conversational speech recognition engines that have the same accuracy as human transcriptions [1], and it has jumped from closely controlled settings to autonomous systems navigating unconstrained acoustic environments [2]. The pervasive use of voice technology in tasks that require a human–computer interface stems from advancements in integrated-circuit and microelectromechanical-system (MEMS) technology, improvements in wireless communication systems, and developments in a plurality of areas, including voice-activity detection (VAD) [3], keyword spotting (KWS) [4], automatic speech recognition (ASR) [5], natural language processing (NLP) [6], and text-to-speech synthesis (TTS) [7].

Today, voice is emerging as a powerful interface for a range of applications controlled by spoken language. Voice-activated systems are operated through a user interface implementing a voice technology that enables tasks to be performed hands-free. These devices "wake up" when called, with examples being "Hey Google," "Hey Alexa" (Amazon's devices), "Hey Siri" (Apple's devices), and "Hey Spotify." Consumer awareness of voice technology is growing according to recent surveys [8]–[10]. As an example, over 50% of smartphone users have resorted to a voice assistant application to search for health-related topics (e.g., asking about illness symptoms, finding medication information, locating a healthcare facility). Other commercial products are used for healthcare-related voice searches as well, as shown in Fig. 1.

The user interface of a voice-activated system is based on smart acoustic sensors and relies on intelligent control, processing, and networking capabilities. Such systems are part of the Internet of Things (IoT), the smart objects that connect the physical world to the Internet or other communication networks. IoT systems find applications not only in everyday life, from smart cities and automated buildings to augmented reality and entertainment, but, increasingly, in healthcare settings from e-health and assistive technologies, to portable diagnostics and wearables.

Various diagnostic tools and medical devices can be viewed as a part of the IoT. Voice-activated IoT healthcare devices (hereafter abbreviated VIHD) provide support also for remote health monitoring, fitness programs, rehabilitation, elderly care, and chronic diseases [11], [12]. As limitations, inequities, and gaps in universal healthcare access have been revealed

Manuscript received February 8, 2022; revised April 5, 2022; accepted May 18, 2022. Date of publication June 1, 2022; date of current version June 29, 2022. This brief was recommended by Associate Editor S. Hoyos. *(Corresponding author: Stefano Gregori.)*

Petros Spachos and Stefano Gregori are with the School of Engineering, University of Guelph, Guelph, ON N1G 2W1, Canada (e-mail: petros@uoguelph.ca; sgregori@ieee.org).

M. Jamal Deen is with the Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON L8S 4K1, Canada (e-mail: jamal@mcmaster.ca).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCSII.2022.3179680.

Digital Object Identifier 10.1109/TCSII.2022.3179680

TABLE I
EXAMPLE APPLICATIONS OF VOICE-ACTIVATED IoT IN HEALTHCARE

Settings \ Function	Assistive to healthcare personnel	Assistive to patient	Monitoring, screening and diagnostic
Hospital, health facility	Integrated operation system, resource/supply management, clinical documentation	Robot-assisted therapy, preparation for procedures, smart room control and communication	Clinical patient monitoring, documentation of diagnostic results
Doctor's office, assisted living, home	Hands-free/eye-free system, electronic health records aid, item/supply tracking aid	Smart-home preventive device, intelligent pharmaceutical package, smart bandage	Smart-home monitoring device, smart personal-care device (scale, toothbrush, etc.)
Anywhere	Voice assistant, journaling aid, patient engagement and telemedicine service interface	Medication compliance aid, hearing/speech difficulties aid, telemedicine service interface	Vital-sign monitoring and activity tracking (hearable, wrist-band, ingestible sensor, etc.)

by the current pandemic, the need for effectively integrating healthcare services with IoT systems has become more pressing. In particular, the accessibility and effectiveness of the healthcare infrastructure can be improved by lowering physical barriers through IoT networks that operate within healthcare centers or connect patients and healthcare providers over longer distances. IoT-based health services are expected to reduce costs, enable remote provision and increase quality of life. Furthermore, the IoT can facilitate scheduling limited resources and tracking the times to replenish supplies for seamless operation. Hence, it is apparent that the use of smart IoT devices will play a critical role in the future healthcare infrastructure, and their applicability will increase when these devices have voice-activation capabilities [13], [14]. In comparison with the aforementioned research, this brief focuses on VIHD, providing a discussion on the design challenges, along with emerging applications.

This brief is organized as follows. Section II reviews the current applications of voice activated IoT devices in healthcare. Section III presents the technological advancements that enable the integration of such devices in small form factors and the available solutions. Section IV discusses the design requirements, challenges, and tradeoffs, and Section V touches on new potential applications. Finally, Section VI will derive the conclusion of the proposed tutorial.

II. VOICE-ACTIVATED IoT IN HEALTHCARE

In healthcare applications, voice and IoT devices have an important role to play in improving access to smart technology for patients and healthcare personnel alike [15]–[17]. Table I captures some applications spanning from assistive functions to diagnostics that are possible in healthcare settings, at home, and elsewhere.

In a hospital or healthcare facility, voice activated devices assist surgeons with performing complex tasks in the operating room while minimizing touches on surfaces, thereby reducing the risk of pathogen transmissions [18]. Voice interfaces are used in computer-aided surgery, robot-assisted activities, and laboratory automation systems [19]. Voice assistants help doctors and nurses with monitoring supplies, providing efficient scheduling of available resources, maintaining patients' electronic medical records, and retrieving relevant information [20]. Automatic speech transcription aids radiologists with the analysis of imaging results and automated recording of findings, thereby facilitating accurate documentation [21]. Voice interfaces assist patients as well, especially when they have restricted mobility, through support for communication, room control, and preparation for medical procedures [22].

Voice-activated devices provide a simple interface also to patients recovering at home and to the elderly for controlling various functions in homecare or assisted-living settings [23], including remote health monitoring [24], preventive and diagnostic smart home [25], home-care robotics [26], intelligent pharmaceutical packages that improve medication compliance [27], and smart bandages and stitches that provide information about the healing progress [28]. Voice-activated devices facilitate access to telemedicine services [16] and provide support to people with hearing or speech difficulties [29]. Finally, real-time speech recognition improves the communication of people with neurodegenerative diseases and speech disorders [30], [31].

Voice technology finds applications also in screening, diagnosing, and monitoring certain health conditions [32]. Breathing and coughing sounds contribute to diagnosing respiratory conditions [33], including coronavirus infections [34]. The detection of vocal biomarkers has been proposed for rapid screening, diagnosing, or monitoring the course of diseases impairing voice production mechanisms [35]–[37]. Recent studies include the identification of depression and post-traumatic stress disorder [38], [39], neurodegenerative diseases [40], [41], and metabolic and cardiovascular diseases [42], [43].

III. STATE OF THE ART

A. Voice-Activated IoT Devices

Practical VIHD designs can incorporate voice capability due to technological advancements in the following areas:

- Audio transducers with integrated MEMS microphones [44], [45], and CMOS circuit interfaces [46]–[48].
- Low-power integrated circuits for signal processing implementing voice techniques (e.g., VAD, KWS, ASR, NLP) [49]–[51].
- Integrated circuits for low-power wireless communications [52]–[54].

Since healthcare voice activated applications must respond at all times, a key challenge is to enable voice activation with a dedicated low-power microcontroller rather than depending on a wireless connection to operate off a cloud-based system. Although this contributes to reducing the power consumption for data transmission, minimizing the power consumption for data acquisition and processing is a critical challenge.

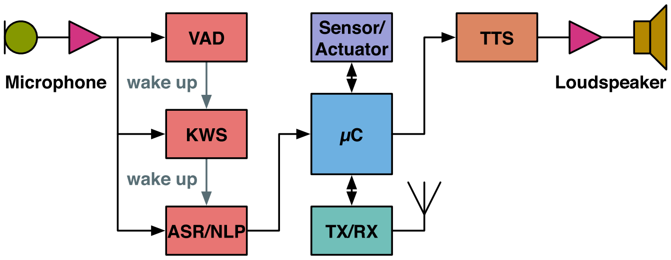

Fig. 2. Typical VIHD block diagram.

As shown in Fig. 2, compared to a conventional low-power embedded system, a typical VIHD includes a multistage wakeup system fed by the microphone (VAD and KWS), a speech recognition and digital classification system (ASR and NLP), and often a text-to-speech synthesis module driving a loudspeaker (TTS) for a two-way user interface. The biomedical sensors or actuators and low-power wireless communication with secure data protocols complete the IoT module.

The requirement of continuous listening poses a severe power-consumption constraint for the modules that are always on. The microphone biasing network and preamplifier are critical parts of the design. Capacitive MEMS microphones are biased at a voltage over 5 V to achieve sufficient sensitivity, which can be delivered by an on-chip charge pump increasing a standard supply voltage. The biasing network and preamplifier typically consume a power of a few hundred microwatts, which can be decreased by reducing biasing voltage and sensitivity in standby mode. Piezoelectric MEMS microphones can reduce the stand-by power to a few ten microwatts because they do not require a biasing voltage [44], [46]. Event-driven approaches are the most successful in reducing the power consumption by acquiring and processing data only when relevant voice data is detected [55]. Wake-up stages for preliminary event detection based on spectral decomposition by analog band-pass filtering can achieve a power consumption below hundred microwatts [56], [57]. Event-driven digital signal processing can be power efficient as well [58], but it requires an always-on analog-to-digital converter, which consumes a few hundred milliwatts [47]. The power consumption of the remaining modules has a lesser impact because they operate with a low duty cycle.

The microphone interface circuits, including preamplifiers and analog to digital converters (ADC) are critical for the audio performance in terms of bandwidth, dynamic range (DR), and total harmonic distortion. A minimum bandwidth of 3.5 kHz and DR of 70 dB are sufficient for basic voice interfaces and can be achieved with extremely low power consumption, which allows always running functionality on battery power. However, some healthcare applications require high fidelity to enable biometric authentication, accurate speech recognition or diagnostic functions. Therefore, larger bandwidth (e.g., 20 kHz) and DR (e.g., 90 dB) are required at the cost of more power (e.g., in the order of 1 mW). For always running applications, the microphone interface circuits should be designed to operate in low-power, low-DR mode for extending battery life, and to switch to high-power, high-DR mode when an audio input signal is detected.

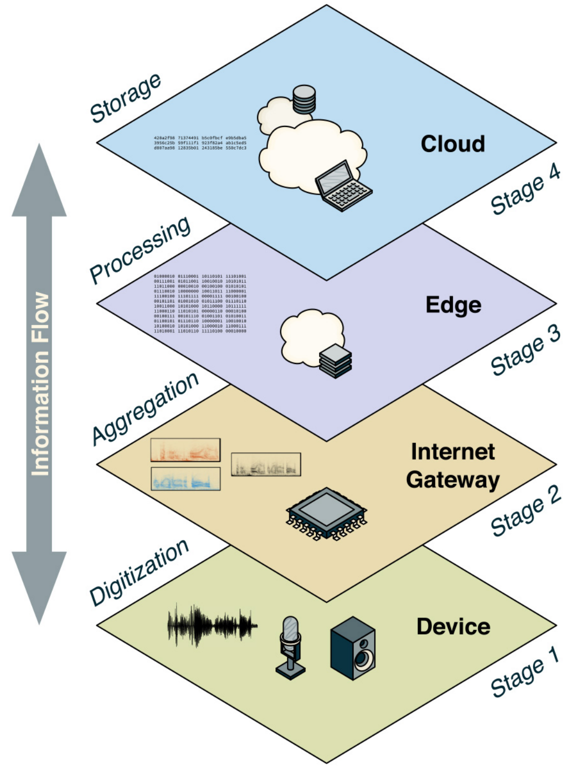

Fig. 3. Stages of a VIHD architecture along with information flow.

Low-power wake-up circuit for acoustic detectors that recognize human speech are implemented with discrete components [57], programmable devices [59], or ASIC designs [56]. As shown in Fig. 2, typical implementations use VAD, which allows the system to sleep while listening, thereby lowering power dissipation as well as processing demand. Once an audio signal of a predefined type is detected, the system resumes full power and performs the operations required [60]–[65]. Effective solutions include multiple wake-up stages, where each element increases in power consumption and computational complexity, starting from analog feature extraction (VAD), proceeding to keyword spotting (KWS), and ending with speech recognition and digital classification (ASR and NLP). Design strategies for processing the resulting audio signal include time-domain and frequency-domain approaches and the application of machine learning (ML) techniques [66]–[70].

B. Voice-Activated IoT Architectures

IoT system architecture is the four-stage process illustrated in Fig. 3 in which data are generated by the device, flow through the network and intermediate stages, and eventually are stored in a data center.

1) Device: This stage is the home of the “thing” in an IoT system. Devices include sensors that generate data, and sometimes actuators that act on their environment. Information from the physical world is converted into digital data before being forwarded to the next stage. In our framework, a microphone

TABLE II
EXAMPLE OFF-THE-SHELF IMPLEMENTATIONS

Device	Features	Standby power
VM1010 [71]	Analog piezoelectric MEMS microphone with low-power voice detection mode (WoS)	18 μW, WoS 153 μW, nml
VM3011 [72]	Digital piezoelectric MEMS microphone with low-power listening mode (ZPL)	18 μW, ZPL 176 μW, nml
CX20921 [73]	SoC with 2 microphone preamplifiers and ADCs with 106 dB dynamic range, dual-core 32-bit DSP with voice wakeup	70 mW
QCS403 [74]	SoC with 4 microphone channels, dual-core CPU, DSP processor with voice wakeup, Wi-Fi, ZigBee and Bluetooth transceiver	100–500 mW

can convert voice data from a surgeon during an operation to a digital form.

2) Internet Gateway: In this stage, the raw data are collected by a data acquisition system. The system aggregates and formats the data before forwarding them, via wires or wireless, through the gateway to the next stage. The gateway can be attached to the device, or it can be stand-alone. A simple microcontroller can take the raw digital data, format them and forward them to a router inside the operation room.

3) Edge Computing: In this stage, the data are processed, close to the devices. After the data have been digitized and aggregated, they are processed for quick analysis and to reduce their volume, before going to the cloud. The router inside the room can decide which data need immediate action during the operation, which need further processing after the operation, and which data should be discarded.

4) Cloud Computing: In this stage, the data are stored while advanced processing can take place. ML methods can be applied to extract useful information from large amounts of data after the operation.

C. Example Implementations

Table II lists examples of off-the-shelf components for VIHD implementations. Piezoelectric MEMS microphones can detect voice activity while consuming a very low power (e.g., 18 μW) [71], [72]. When the microphone detects a sound in the voice band above a configurable threshold, the system switches to normal mode in a time short enough to capture and process the sound that exceeds the threshold (e.g., 200 μs). While the output of an analog microphone is processed by a preamplifier and an ADC, a digital microphone provides a pulse density modulation output directly that is processed by a codec. Such components are typically part of a system on a chip (SoC) with a digital signal processor (DSP). Typical implementations dedicated to smart audio include preamplifiers and ADCs for multiple microphones, a built-in DSP-based voice wakeup system and communication circuits [73], [74]. However, their power consumption can exceed 100 mW depending on the implementation and the active modules. To enable battery-operated always-listening applications, the average power can be reduced to a few milliwatts by maintaining the SoC is sleep mode and relying on a microphone with VAD to resume normal operation when voice is detected.

IV. TRENDS AND CHALLENGES IN VIHD

A. Trends

1) Speech-Recognition Artificial Intelligence: There are increasing efforts to develop more robust speech recognition systems using advances in intelligent software that include artificial intelligence (AI), ML and NLP. These techniques are becoming better at handling accents and distinguishing speech from background noise. There is also much progress in automated speech recognition using intelligent software driven by language models combined with linguistics, experimental psychology and advanced data analytics. In addition, significant advancements are being made in voice interface design and voice application development. It is expected that continued progress in speech recognition for VIHD will enhance user experience further [75]–[77].

2) Low-Power IoT Devices: A key component of VIHD is MEMS microphones with associated electronics and software for capturing and processing high-quality voice communications while suppressing environmental noise and extraneous audio signals. For portable or wearable devices, compact size and low power are significant design challenges being addressed to satisfy user needs. In fact, in the last years, there were continual improvements in the development of ultra-low power audio and voice capabilities of VIHD. Techniques such as VAD allow a VIHD system to listen (e.g., using only one of an array of several microphones) even when in low-power mode to reduce total power usage. Only when a speech signal above a predefined threshold is detected, the system wakes up for normal audio signal processing and action.

3) Personalized Health Services: A continuing trend in VIHD is the provision of personalized services with a high-level of data security. This is because it is expected that VIHD will help to improve personalized patient experience such as updating their medical record in real-time or providing relevant information about medications. Stimulated by the ongoing pandemic, VIHD is expected to continue advancing digital health by providing improved, personalized, predictive and preventative healthcare through real-time monitoring or self-care using AI powered diagnostics [78], [79].

B. Challenges

1) Medical Data Security and Privacy: As other medical devices with communication capabilities, VIHD are prone to security risks that may include theft of patient data, service denial, manipulation of therapy, or damage to assets. These security risks become more important as these devices are connected to a network for real-time data transmission. Sometimes, security may be traded at the expense of user convenience, but VIHD may reduce the risks with robust biometric authentication based on integrated speech recognition and voice identity verification techniques [80], [81]. Further, audio files are stored in the VIHD with a high-level of sensitive data security, and secure data protocols are implemented. The ongoing challenges are to ensure adequate audio performance for biometric authentication (i.e., signal-to-noise ratio and dynamic range), and to improve blockchain cryptographic algorithms for robust encryption of users' data to keep them secure and private.

2) Training of Speech Recognition Technologies: In VIHD, a continuing challenge is for large, diverse training sets which are critical for ML or AI software. Further, expectations of users of these VIHD are increasingly demanding higher accuracy since it is less acceptable for AI-systems to make errors. Therefore, a continuing challenge is to make such devices speaker independent (except for user authentication using speech-based biometrics), so that multiple persons such as doctors, nurses and other healthcare professionals can use them without additional speech recognition training for tasks such as record transcription over voice. For this to become a reality, the voice-activated systems should identify different languages, accents, pitches, volumes and speeds of speech, as well as variations of speech related to emotions, mental/psychological states, and other atypical speech patterns, all of which require large heterogeneous training datasets. In speech recognition technologies, it is also challenging to know when to use the global approach, where an entire word is recognized, versus the analytical approach that uses the linguistic structure of a word to detect and identify its basic components of phonemes and syllables. In addition, it is still challenging for the software in speech recognition technologies to handle voice in the presence of background noise [82], [83].

3) Reliability and Robustness: Reliability and robustness of network access may be dependent on geographic location as smaller rural centers may not have the same level of access as bigger urban areas, and there may be network outages which can be frustrating for users of VIHD. Reliability can be enhanced with blockchain technology, which will also help with data security and privacy. Another challenge is to improve the reliability, robustness and battery lifetime of the hardware such as MEMS microphones, data and signal processing integrated circuits and networking infrastructure.

4) Standardization and Interoperability: Standardization across networks, application programming interfaces and protocols are key challenges to increase VIHD popularity and to make them more secure while addressing users' concerns on security and data privacy. Technological standards are necessary, including network protocols, communication protocols and data aggregation. There is also a need for regulatory standards related to issues such as security and privacy of data and data accountability, as well as collection, modification, deletion, use and storage of data. The collected data from multiple VIHD may be of different formats at different sampling rates, and structured or unstructured, so there is a need to establish standards for their handling, aggregation and transformation. Standards for data brokers, i.e., companies that sell data collected from a variety of sources, are desirable as well. For interoperability among VIHD, blockchain technology is of great benefit. Also, the IoT platform should ensure interoperability and scalability among information-technology systems, voice-activated healthcare devices and service interfaces. This is especially challenging as the IoT platforms will have non uniform user interfaces catering different physical abilities, such as hearing or speech impairments, and which must satisfy specific standards and be easily upgradeable. Challenges also exist in interoperability in heterogeneous sensor systems related to communication protocols; scalability; and the storage, exchange and security of collected data primarily because most sensor systems are proprietary and customized [84].

V. Future Applications

Voice activated IoT devices have a great potential, not only to improve existing healthcare activities, but also to make novel applications possible. Some new applications enabled by VIHD include the following.

1) Voice-Tech Patient Screening: AI-powered virtual assistants and chatbots played a significant role in the fight against Covid-19. Chatbots operating through VIHD can help to screen individuals faster and decrease the number of patients that needed to meet with healthcare providers.

2) Personalized Healthcare and Improved Patient Experience: VIHDs can collect crucial information such as vital signs in real time, update medical records and notify doctors immediately. The patients will use their voices to issue the necessary commands, minimizing the risk of not using the device properly.

3) Telehealth and e-Health: Low and inexpensive IoT devices will make telehealth more affordable while they will reduce the number of patients that need to access the limited healthcare facilities.

4) Smart Home and Environment: VIHDs can be used in a smart home or autonomous vehicle and provide important data that through ML can reveal useful correlations between patient everyday activities and their health.

5) Predictive and Prescriptive Health Data Analysis Through Advanced Voice Recognition: Predictive analysis aims to foretell issues before they occur, while prescriptive analysis recommends solutions for upcoming problems. Both approaches have the potential to make e-health revolutionary.

Finally, as trivial tasks are performed through VIHDs, medical personnel can focus on critical tasks, perform more complex operations and assist patients more effectively. This would not only reduce the workload, but also improve accessibility and facilitate new healthcare services.

VI. Conclusion

This brief has presented a review of the most recent voice activated IoT devices focusing on healthcare, along with current trends and challenges and some future applications. Several VIHD devices are available in the market and there is a technological trend toward the further development of such systems. However, their usage should follow some basic standards while their applicability in healthcare is highly based on the level of performance and provided security. When these challenges are properly addressed, they can be used in a plethora of future healthcare applications.

REFERENCES

[1] W. Xiong *et al.*, "Toward human parity in conversational speech recognition," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 25, no. 12, pp. 2410–2423, Dec. 2017.

[2] A. Schmidt, H. W. Löllmann, and W. Kellermann, "Acoustic self-awareness of autonomous systems in a world of sounds," *Proc. IEEE*, vol. 108, no. 7, pp. 1127–1149, Jul. 2020.

[3] M. Croce, B. Friend, F. Nesta, L. Crespi, P. Malcovati, and A. Baschirotto, "A 760-nW, 180-nm CMOS fully analog voice activity detection system for domestic environment," *IEEE J. Solid-State Circuits*, vol. 56, no. 3, pp. 778–787, Mar. 2021.

[4] I. López-Espejo, Z.-H. Tan, J. H. L. Hansen, and J. Jensen, "Deep spoken keyword spotting: An overview," *IEEE Access*, vol. 10, pp. 4169–4199, 2022.

[5] S. Alharbi *et al.*, "Automatic speech recognition: Systematic literature review," *IEEE Access*, vol. 9, pp. 131858–131876, 2021.

[6] P. M. Nadkarni, L. Ohno-Machado, and W. W. Chapman, "Natural language processing: An introduction," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 5, pp. 544–551, Sep. 2011.

[7] A. Tjandra, S. Sakti, and S. Nakamura, "Machine speech chain," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, no. 1, pp. 976–989, Mar. 2020.

[8] S. Higginbotham, "Do you need a smart microwave? [Opinion]," *IEEE Spectr.*, vol. 56, no. 2, p. 22, Feb. 2019.

[9] "Consumer Intelligence Series: Prepare for the Voice Revolution." PwC. 2018. [Online]. Available: https://www.pwc.com/us/en/advisory-services/publications/consumer-intelligence-series/voice-assistants.pdf

[10] "How Voice Assistants are Changing the Way Patients Search for Healthcare." Chatmeter. 2022. [Online]. Available: https://www.chatmeter.com/blog/how-voice-assistants-are-changing-the-way-patients-search-for-healthcare/

[11] S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The Internet of Things for health care: A comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.

[12] S. Latif, J. Qadir, S. Farooq, and M. A. Imran, "How 5G wireless (and concomitant technologies) will revolutionize healthcare?" *Future Internet*, vol. 9, no. 4, p. 93, Dec. 2017.

[13] M. N. Bhuiyan, M. M. Rahman, M. M. Billah, and D. Saha, "Internet of Things (IoT): A review of its enabling technologies in healthcare applications, standards protocols, security, and market opportunities," *IEEE Internet Things J.*, vol. 8, no. 13, pp. 10474–10498, Jul. 2021.

[14] M. J. Baucas, P. Spachos, and S. Gregori, "Internet-of-Things devices and assistive technologies for health care: Applications, challenges, and opportunities," *IEEE Signal Process. Mag.*, vol. 38, no. 4, pp. 65–77, Jul. 2021.

[15] C. E. Koop *et al.*, "Future delivery of health care: Cybercare," *IEEE Eng. Med. Biol. Mag.*, vol. 27, no. 6, pp. 29–38, Nov./Dec. 2008.

[16] P. Pawar, V. Jones, B.-J. F. van Beijnum, and H. Hermens, "A framework for the comparison of mobile patient monitoring systems," *J. Biomed. Inform.*, vol. 45, no. 3, pp. 544–556, Jun. 2012.

[17] S. Latif, J. Qadir, A. Qayyum, M. Usama, and S. Younis, "Speech technology for healthcare: Opportunities, challenges, and state of the art," *IEEE Rev. Biomed. Eng.*, vol. 14, no. 1, pp. 342–356, Jan. 2021.

[18] A. Perrakis, W. Hohenberger, and T. Horbach, "Integrated operation systems and voice recognition in minimally invasive surgery: Comparison of two systems," *Surg. Endosc.*, vol. 27, no. 2, pp. 575–579, Feb. 2013.

[19] A. Di Lallo, R. Murphy, A. Krieger, J. Zhu, R. H. Taylor, and H. Su, "Medical robots for infectious diseases: Lessons and challenges from the COVID-19 pandemic," *IEEE Robot. Autom. Mag.*, vol. 28, no. 1, pp. 18–27, Mar. 2021.

[20] S. V. Blackley, V. D. Schubert, F. R. Goss, W. Al Assad, P. M. Garabedian, and L. Zhou, "Physician use of speech recognition versus typing in clinical documentation: A controlled observational study," *Int. J. Med. Inform.*, vol. 141, Sep. 2020, Art. no. 104178.

[21] D. Ganeshan *et al.*, "Structured reporting in radiology," *Acad. Radiol.*, vol. 25, no. 1, pp. 66–73, Jan. 2018.

[22] M. Amiribesheli, A. Benmansour, and A. Bouchachia, "A review of smart homes in healthcare," *J. Ambient Intell. Humanized Comput.*, vol. 6, no. 4, pp. 495–517, Aug. 2015.

[23] H. Thapliyal, R. K. Nath, and S. P. Mohanty, "Smart home environment for mild cognitive impairment population: Solutions to improve care and quality of life," *IEEE Consum. Electron. Mag.*, vol. 7, no. 1, pp. 68–76, Jan. 2018.

[24] K. Nisar, A. A. A. Ibrahim, L. Wu, A. Adamov, and M. J. Deen, "Smart home for elderly living using wireless sensor networks and an android application," in *Proc. IEEE Int. Conf. Appl. Inf. Commun. Technol. (AICT)*, Oct. 2016, pp. 1–8.

[25] S. Majumder *et al.*, "Smart homes for elderly healthcare—Recent advances and research challenges," *Sensors*, vol. 17, no. 11, p. 2496, Oct. 2017.

[26] G. Yang *et al.*, "Homecare robotic systems for healthcare 4.0: Visions and enabling technologies," *IEEE J. Biomed. Health Inform.*, vol. 24, no. 9, pp. 2535–2549, Sep. 2020.

[27] Z. Pang, J. Tian, and Q. Chen, "Intelligent packaging and intelligent medicine box for medication management towards the Internet-of-Things," in *Proc. Int. Conf. Adv. Commun. Technol. (ICACT)*, Feb. 2014, pp. 352–360.

[28] X. Jin, C. Liu, T. Xu, L. Su, and X. Zhang, "Artificial intelligence biosensors: Challenges and prospects," *Biosens. Bioelectron.*, vol. 165, Oct. 2020, Art. no. 112412.

[29] J.-C. Edelmann and T. Ussmueller, "Can you hear me now? Challenges and benefits for connectivity of hearing aids and implants," *IEEE Microw. Mag.*, vol. 19, no. 7, pp. 30–42, Nov./Dec. 2018.

[30] J. A. Gonzalez-Lopez, A. Gomez-Alanis, J. M. Martín Doñas, J. L. Pérez-Córdoba, and A. M. Gomez, "Silent speech interfaces for speech restoration: A review," *IEEE Access*, vol. 8, pp. 177995–178021, 2020.

[31] B. G. Schultz *et al.*, "Automatic speech recognition in neurodegenerative disease," *Int. J. Speech Technol.*, vol. 24, no. 3, pp. 771–779, Sep. 2021.

[32] E. Anthes, "Alexa, do I have COVID-19?" *Nature*, vol. 586, pp. 22–25, Oct. 2020.

[33] M. Schatz *et al.*, "Reliability and predictive validity of the asthma control test administered by telephone calls using speech recognition technology," *J. Allergy Clin. Immunol.*, vol. 119, no. 2, pp. 336–343, Feb. 2007.

[34] W. Jiang *et al.*, "A wearable tele-health system towards monitoring COVID-19 and chronic diseases," *IEEE Rev. Biomed. Eng.*, vol. 15, no. 1, pp. 61–84, Jan. 2022.

[35] G. Fagherazzi, A. Fischer, M. Ismael, and V. Despotovic, "Voice for health: The use of vocal biomarkers from research to clinical practice," *Digit. Biomark.*, vol. 5, no. 1, pp. 78–88, Apr. 2021.

[36] G. Muhammad and M. Alhussein, "Convergence of artificial intelligence and Internet of Things in smart healthcare: A case study of voice pathology detection," *IEEE Access*, vol. 9, pp. 89198–89209, 2021.

[37] C. Robotti *et al.*, "Machine learning-based voice assessment for the detection of positive and recovered COVID-19 patients," *J. Voice*, to be published.

[38] Z. Huang, J. Epps, and D. Joachim, "Investigation of speech landmark patterns for depression detection," *IEEE Trans. Affective Comput.*, vol. 13, no. 12, pp. 666–679, Apr.–Jun. 2022.

[39] C. R. Marmar *et al.*, "Speech-based markers for posttraumatic stress disorder in U.S. veterans," *Depress. Anxiety*, vol. 36, no. 7, pp. 607–616, Jul. 2019.

[40] K. C. Fraser, J. A. Meltzer, and F. Rudzicz, "Linguistic features identify Alzheimer's disease in narrative speech," *J. Alzheimer's Dis.*, vol. 49, no. 2, pp. 407–422, 2016.

[41] J. Laguarta and B. Subirana, "Longitudinal speech biomarkers for automated Alzheimer's detection," *Front. Comput. Sci.*, vol. 3, Apr. 2021, Art. no. 624694.

[42] D. Chitkara and R. Sharma, "Voice based detection of type 2 diabetes mellitus," in *Proc. Int. Conf. Adv. Electr. Electron. Inf. Commun. Bio-Inform. (AEEICB)*, Feb. 2016, pp. 83–87.

[43] E. Maor, J. D. Sara, D. M. Orbelo, L. O. Lerman, Y. Levanon, and A. Lerman, "Voice signal characteristics are independently associated with coronary artery disease," *Mayo Clin. Proc.*, vol. 93, no. 7, pp. 840–847, Jul. 2018.

[44] S. A. Zawawi, A. A. Hamzah, B. Y. Majlis, and F. Mohd-Yasin, "A review of MEMS capacitive microphones," *Micromachines*, vol. 11, no. 5, p. 484, May 2020.

[45] Z. Lin *et al.*, "A personalized acoustic interface for wearable human–machine interaction," *Adv. Funct. Mater.*, vol. 32, no. 9, Feb. 2022, Art. no. 2109430.

[46] Y. Yoo and B.-D. Choi, "Readout circuits for capacitive sensors," *Micromachines*, vol. 12, no. 8, p. 960, Aug. 2021.

[47] P. Malcovati and A. Baschirotto, "The evolution of integrated interfaces for MEMS microphones," *Micromachines*, vol. 9, no. 7, p. 323, Jul. 2018.

[48] C. De Berti, P. Malcovati, L. Crespi, and A. Baschirotto, "A 106 dB A-weighted DR low-power continuous-time $\Sigma\Delta$ modulator for MEMS microphones," *IEEE J. Solid-State Circuits*, vol. 51, no. 7, pp. 1607–1618, Jul. 2016.

[49] D. Du and K. Odame, "An energy-efficient spike encoding circuit for speech edge detection," *Analog Integr. Circuits Signal Process.*, vol. 75, no. 3, pp. 447–458, Jun. 2013.

[50] E. Shi, X. Tang, and K. P. Pun, "A 270 nW switched-capacitor acoustic feature extractor for always-on voice activity detection," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 68, no. 3, pp. 1045–1054, Mar. 2021.

[51] D. Rossi *et al.*, "Vega: A ten-core SoC for IoT endnodes with DNN acceleration and cognitive wake-up from MRAM-based state-retentive sleep mode," *IEEE J. Solid-State Circuits*, vol. 57, no. 1, pp. 127–139, Jan. 2022.

[52] C. Gomez, J. Oller, and J. Paradells, "Overview and evaluation of Bluetooth low energy: An emerging low-power wireless technology," *Sensors*, vol. 12, no. 9, pp. 11734–11753, Sep. 2012.

[53] E. Garcia-Espinosa, O. Longoria-Gandara, I. Pegueros-Lepe, and A. Veloz-Guerrero, "Power consumption analysis of Bluetooth low energy commercial products and their implications for IoT applications," *Electronics*, vol. 7, no. 12, p. 386, Dec. 2018.

[54] C. Li *et al.*, "Overview of recent development on wireless sensing circuits and systems for healthcare and biomedical applications," *IEEE J. Emerg. Sel. Topics Circuits Syst.*, vol. 8, no. 2, pp. 165–177, Jun. 2018.

[55] Y. Tsividis, "Event-driven data acquisition and digital signal processing—A tutorial," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 57, no. 8, pp. 577–581, Aug. 2010.

[56] T. Delbruck, T. Koch, R. Berner, and H. Hermansky, "Fully integrated 500 μW speech detection wake-up circuit," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2010, pp. 2015–2018.

[57] D. Oletic, V. Bilas, M. Magno, N. Felber, and L. Benini, "Low-power multichannel spectro-temporal feature extraction circuit for audio pattern wake-up," in *Proc. Des. Autom. Test Eur. Conf. Exhibit. (DATE)*, Mar. 2016, pp. 355–360.

[58] S. Mourrane, B. Larras, A. Cathelin, and A. Frappé, "Event-driven continuous-time feature extraction for ultra low-power audio keyword spotting," in *Proc. IEEE Int. Conf. Artif. Intell. Circuits Syst. (AICAS)*, Jun. 2021, pp. 1–4.

[59] S. Shah and J. Hasler, "Low power speech detector on a FPAA," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2017, pp. 1–4.

[60] S. Lauwereins, W. Meert, J. Gemmeke, and M. Verhelst, "Ultra-low-power voice-activity-detector through context- and resource-cost-aware feature selection in decision trees," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process. (MLSP)*, Sep. 2014, pp. 1–6.

[61] S. Lecoq, J. Le Bellego, A. Gonzalez, B. Larras, and A. Frappé, "Low-complexity feature extraction unit for 'wake-on-feature' speech processing," in *Proc. IEEE Int. Conf. Electron. Circuits Syst. (ICECS)*, Dec. 2018, pp. 677–680.

[62] E. Fallis *et al.*, "A testbed for adaptive microphones in ultra-low-power systems," in *Proc. IEEE Sustainability Through ICT Summit (StICT)*, Jun. 2019, pp. 1–6.

[63] M. Yang, C.-H. Yeh, Y. Zhou, J. P. Cerqueira, A. A. Lazar, and M. Seok, "Design of an always-on deep neural network-based 1-μW voice activity detector aided with a customized software model for analog feature extraction," *IEEE J. Solid-State Circuits*, vol. 54, no. 6, pp. 1764–1777, Jun. 2019.

[64] E. Fallis, P. Spachos, and S. Gregori, "A power-efficient audio acquisition system for smart city applications," *Internet Things*, vol. 9, Mar. 2020, Art. no. 100155.

[65] M. Lipski, M. James, P. Spachos, and S. Gregori, "Low power data acquisition system for noise pollution monitoring," in *Proc. IEEE Can. Conf. Elect. Comput. Eng. (CCECE)*, Aug. 2020, pp. 1–4.

[66] D. Oletic, L. Korman, M. Magno, and V. Bilas, "Time-frequency pattern wake-up detector for low-power always-on sensing of acoustic events," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (I2MTC)*, May 2018, pp. 1–6.

[67] M. Price, J. Glass, and A. P. Chandrakasan, "A low-power speech recognizer and voice activity detector using deep neural networks," *IEEE J. Solid-State Circuits*, vol. 53, no. 1, pp. 66–75, Jan. 2018.

[68] A. d. S. P. Soares, W. D. Parreira, E. G. Souza, C. d. D. do Nascimento, and S. J. M. de Almeida, "Voice activity detection using generalized exponential kernels for time and frequency domains," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 66, no. 6, pp. 2116–2123, Jun. 2019.

[69] D. Josh, J.-A. Elenis, H. Muresan, P. Spachos, and S. Gregori, "Low-power low-cost audio front-end for keyword spotting," in *Proc. IEEE Can. Conf. Elect. Comput. Eng. (CCECE)*, Aug. 2020, pp. 1–4.

[70] J. H. Teo, S. Cheng, and M. Alioto, "Low-energy voice activity detection via energy-quality scaling from data conversion to machine learning," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 67, no. 4, pp. 1378–1388, Apr. 2020.

[71] "VM1010." Vesper. 2017. [Online]. Available: https://vespermems.com/products/vm1010/

[72] "VM3011." Vesper. 2017. [Online]. Available: https://vespermems.com/products/vm3011/

[73] "CX20921." Synaptics. 2017. [Online]. Available: https://www.synaptics.com/sites/default/files/audiosmart-dual-mic-far-field-voice-cx20921.pdf

[74] "QCS403." Qualcomm. 2019. [Online]. Available: https://www.qualcomm.com/products/technology/processors/application-processors/qcs400-series/qcs403

[75] M. Ma, M. Nirschl, F. Biadsy, and S. Kumar, "Approaches for neural-network language model adaptation," in *Proc. Conf. Int. Speech Commun. Assoc. (INTERSPEECH)*, Aug. 2017, pp. 259–263.

[76] "Visualize Speech Data With Speech Analysis Framework." Google Cloud Architecture Center. [Online]. Available: https://cloud.google.com/architecture/visualize-speech-data-with-framework (Accessed: Jan. 2022).

[77] J. P. Bajorek (Harvard Business Review, Watertown, MA USA). *Voice Recognition Still Has Significant Race and Gender Biases*. (May 2019). [Online]. Available: https://hbr.org/2019/05/voice-recognition-still-has-significant-race-and-gender-biases

[78] P. Radanliev *et al.*, "COVID-19 what have we learned? The rise of social machines and connected devices in pandemic management following the concepts of predictive, preventive and personalized medicine," *EPMA J.*, vol. 11, no. 3, pp. 311–332, Sep. 2020.

[79] H. Tarkkala, I. Helén, and K. Snell, "From health to wealth: The future of personalized medicine in the making," *Futures*, vol. 109, pp. 142–152, May 2019.

[80] J. P. Campbell, W. Shen, W. M. Campbell, R. Schwartz, J.-F. Bonastre, and D. Matrouf, "Forensic speaker recognition," *IEEE Signal Process. Mag.*, vol. 26, no. 2, pp. 95–103, Mar. 2009.

[81] E. G. Spanakis, M. Psaraki, and V. Sakkalis, "Congestive heart failure risk assessment monitoring through Internet of Things and mobile personal health systems," in *Proc. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 2925–2928.

[82] G. López, L. Quesada, and L. A. Guerrero, "Alexa vs. Siri vs. Cortana vs. Google assistant: A comparison of speech-based natural user interfaces," in *Advances in Human Factors and Systems Interaction (AHFE 2017)*, I. L. Nunes, Ed., vol. 592. Cham, Switzerland: Springer, 2018, pp. 241–250.

[83] C.-C. Chiu *et al.*, "Speech recognition for medical conversations," Jun. 2018, *arXiv:1711.07274*.

[84] K. Panetta. "Interoperability, AI comprise healthcare IoT platform essentials." IoTAgenda. Jan. 2021. [Online]. Available: https://internetofthingsagenda.techtarget.com/tip/Interoperability-AI-comprise-healthcare-IoT-platform-essentials