# Auditory and Haptic Solutions for Access and Feedback in Internet of Digital Reality Applications

György Wersényi
*Department of Telecommunications*
*Széchenyi István University*
Győr, Hungary
wersenyi@sze.hu

Ádám Csapó
*Department of Informatics*
*Széchenyi István University*
Győr, Hungary
csapo.adam@sze.hu

***Abstract*—The concept of Internet of Digital Reality (IoD) was introduced as the next level organization of cognitive entities following the concept of the Internet of Things (IoT) and Internet of Everything (IoE). As virtual-immersive environments are a fundamental component of IoD which allow human and non-human entities to interact in real time, the ability a wide range of communication modalities is crucial. This paper briefly presents the concept of IoD together with an overview of various I/O solutions for human users, with a focus on research directions and (re)emerging technologies in the near future.**

***Keywords—Internet of Digital Reality, Auditory Display, VR, Haptics, Cognitive Infocommunications***

## I. INTRODUCTION

With the dawn of the Internet, a new communication network was created which enabled human users to connect with each other via computers. First, transmitted information was mainly textual (static html files, e-mails), followed by graphical user interfaces and audio / video (as seen on the Web 2.0). As "smart" devices became increasingly available and able to access the public and private networks, the term Internet of Things was coined and developed. In the case of IoT, a large variety of devices (ranging from miniature sensors to large-sized appliances) can be connected to, controlled, and communicated with [1-4]. The concept of Internet of Everything was derived from IoT, by adding smart connections into the concept and bringing together humans, processes, and data in addition to the devices and appliances characterizing IoT [5, 6]. Today, as a result of the mobile computing and AI revolution, we are witnessing a deepening pervasiveness of 2D (and increasingly, 3D) digital environments, leading not only to users being able to obtain information and influence the world faster and at a larger scale, but also more generally changing the way users perceive what is desirable and achievable within both the digital and physical realm, through digital intermediaries.

To better characterize these developments, several recent works have emphasized the emergence of a new kind of infocommunications, called cognitive infocommunications (CogInfoCom) and cognitive entities – a new, co-evolved form of capabilities that are neither purely artificial nor purely biological [7-10]. With the merging of 2D digital environments, virtual/augmented realities, and artificial intelligence, together with cognitive entities supported by cognitive infocommunications, the concepts of Digital Reality and Internet of Digital Reality have also recently been proposed [11,12].

### A. Internet of Digital Reality

The concept of Digital Reality has in fact emerged in various forms in the literature, first in various publications by Deloitte Consulting LLP and Consumer Technology Association (see e.g., [13]) and was extended based on a scientific perspective in [11,12].

IoD can be conceived of as the next level organization of connected "things" in a way that emphasizes emergent, hybrid cognitive capabilities. "Things" in this case are labelled as cognitive entities that are present in the network and which, rather than being mere entities that undergo interactions, instead reflect merged capabilities that are neither purely natural, nor purely artificial. They can be accessed, they can communicate with each other and they can be logically grouped dynamically according to their role and function – a feature that is crucial in creating new digital realities. Importantly, there is no restriction whether a cognitive digital entity is a human user or a non-human "thing", an algorithm or any kind of an artificial intelligence. As the borders are blurred between these components, human users may face new challenges during communication (e.g., depending on whether the counterpart is human or not). Furthermore, IoD assumes that much of the interactions with humans will be carried out in a partially or fully virtual space (VR, AR or Mixed Reality [14, 15]), even entertaining fully immersive scenarios using auditory, visual and haptic/tactile modalities for input and output (feedback) [16].

The current interface to the Internet (Web 2.0) involves web pages using simple 2D GUIs, mostly based on textual information and 2D graphics, sometimes using sound and video. IoD will replace/extend web pages using 3D virtual representations using 3D graphics (e.g., WebGL), 3D audio, video and haptics. Further, it can embed such web pages into a broader "metaverse" of 3D spaces, which can be navigated between using between-space hyperlinks (Fig. 1). However, IoD is more than a metaverse in that it emphasizes the creation of new realities, that is, environments that form a holistic unity for a given purpose.

All these developments can create new challenges for users when it comes to dealing with increased cognitive load. IoD can in this sense adopt a human centered approach based on related fields. The pillars of IoD, as outlined in [11, 12] include:

- Cognitive entities (any combination of machines, sensors, digital twins, bots, AI, human users etc. forming new, emergent cognitive capabilities)
- Data and control information
- Communication networks (wired or wireless from physical layers to logical segmentation and intelligent management)
- Artificial intelligence as a global entity
- Access devices and interfaces (I/O devices)

978-1-6654-9653-7/22/$31.00 ©2022 IEEE

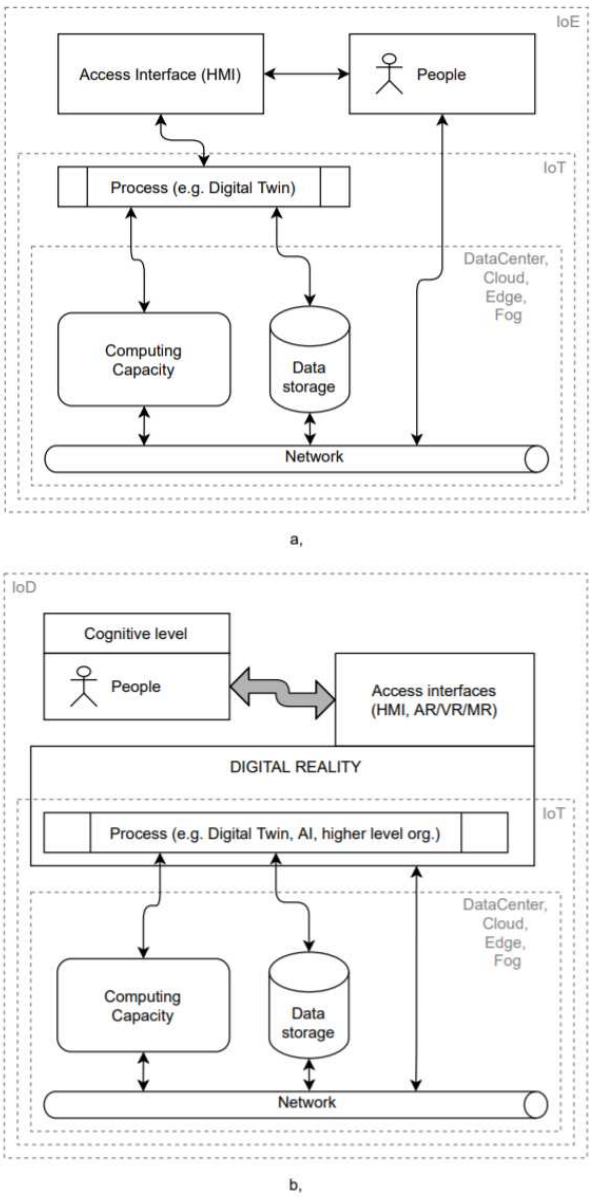


Fig. 1. The concepts of IoT, IoE and IoD and their relationships.

- Cognitive infocommunications technologies (sensation and perception, human factors and HCI, long-term co-evolved capability creation)
- Safety and security
- Digital business and legal issues
- Digital society (education, technology acceptance, digital work, e-government, digital trust via e.g., blockchain, digital arts and gaming)

This paper focuses primarily on the pillar of accessing I/O devices.

## II. ACCESSING IoD

Virtual Reality and 360-degree immersion is a key component of IoD. This means that 2D screens and scenarios are and will be increasingly replaced by 3D spaces on the 2D screens (e.g., desktop VR) as well as 3D displays, helmets, smart glasses etc. These devices can function both as standalone tools and/or as an integrated hardware I/O device providing both input modalities and output (feedback) in a networked environment.

### *A. Input devices*

Today's typical input devices include keyboards and mouse input devices well suited for entering long text inputs and navigation inputs for 2D screens. Gaming consoles, in parallel, use somewhat modified setups (joysticks, controller wheels) for a better gaming experience in 3D, even as keyboard devices and touchscreens (especially on mobile devices) are still an option.

The evolution of speech recognition also allows simple speech commands to be initiated by users, but the technology is still highly language dependent. In the "larger" languages spoken by hundreds of millions of people (e.g., English, Chinese), existing solutions are robust, whereas dictating long text inputs in smaller languages is farther away from widespread adoption. Even so, extensive use of speech interfaces requires a quiet environment insulated from external sounds, which is a hindrance to extended use beyond short spoken commands.

Haptic devices (gloves, gesture controls) using accelerometers and/or optical devices communicating with a base station are also spreading, especially along with VR headsets for entertainment and Industry 4.0 solutions. However, these are still relatively heavy and often require wired connections which are cumbersome for end users.

With respect to IoD, virtual immersion is almost a prerequisite, therefore the most evident next level solution would be some kind of headset or helmet. Entering text information using a (real or virtual) keyboard will still be a preferred method for a long time. This can be replaced by speech, first using simple speech commands (i.e., "send", "delete", "save", "close"), later if technology reaches the appropriate level, even complex textual information can be "dictated" [17]. Speech, as an input method, when independent of the task, is therefore highly dependent on the underlying AI capability. Amazon's Alexa and Apple's Siri are recent examples that serve as an I/O device and personal assistant for information exchange based on a speech module in different languages that can be regarded as one of the first prototypes of an IoD input device (where the AI is in the cloud, and the device exists in a physical form without a virtual 3D immersion) [18, 19].

Regarding textual information as input, speech recognition techniques will be the most important element and R&D field of IoD. This also includes the human factors, such as emotion detection, appropriate use of language, humor etc. In the case of input commands that we usually execute by "clicking with something on something" the visual content is already there (i.e., icons, menu items, checkboxes, scrollbars). Here, single-word iconic speech commands can be a first step.

Another solution is haptics, whereby the mouse can be replaced by a targeting device with at least two clicking options (left and right button of the mouse) and a means for scrolling up and down. Special devices can be developed with more intuitive functions of navigation in 3D besides the one hand control: using two hands and arm(s), the head or body movements.

Gesture control is special, because a dedicated capture device is needed in the "line of sight" of the hands [20-22].

Furthermore, accuracy and usability are still an issue. This technique is widely used in high-end automobiles as a way to control the embedded infotainment system (volume up/down, switching on/off) [23].

In the medium term, we expect a great explosion of integrated command devices incorporating speech and tactile/haptic methods, especially in the realm of 3D VR scenarios.

### B. *Feedback devices*

In general, speech synthesis is more developed than speech recognition [24]. The first "robotic" voices using synthetized speech were characterized by a restricted quality due to the limited resolution and frequency bandwidth of parametric (articulatory, formant-based or concatenative) synthesis methods. Evolution of speech feedback methods can be well tracked by testing today's smartphone capabilities or by listening to sophisticated GPS routing commands of a navigation device: such solutions are characterized by a relatively higher quality, low distortion output, leading to good speech intelligibility [25, 26]. Today, like in many domains, state-of-the-art solutions for text-to-speech rely on end-to-end training of artificial neural networks [27]. Better intonation and emotion expressions can be expected to widen the possibilities of meaningful communication and make the underlying systems more human-like.

In contrast to audio input, audio output can rely on much more than speech, especially if simplicity is a requirement and/or language-independent solutions are preferred. Sonification is a wide area of research dealing with various kinds of audio signals, rendering methods and cognitive aspects [28]. Earcons, auditory icons, spearcons, auditory emoticons, lyricons, spindices etc. can serve as sound events in a wide range of scenarios, ranging from informal to high-level conceptual feedback [29-31]. In many cases, users can respond better to familiar everyday sound events or to well-known music samples; even cartoonification has been shown as a viable tool in resolving ambiguous situations [32]. Regarding ways to map meaning on to (concrete or abstract) parametric audio signals, both direct forms of mapping as well as association-based mapping schemes can be applicable in a wide range of scenarios [33].

Spatial rendering of sound sources ranging from simple panoramic stereo rendering to fully immersive HRTF solutions and auralization methods can also greatly improve user experience [34-37]. Depth (monaural distance) information is usually associated with the loudness of the sound signal. Another question that often arises in the case of headsets is whether head-tracking is involved or not. 3D VR scenarios are often associated with spatially distributed objects and sound sources; thus, psychoacoustics can be expected to play a significant role.

Based on the above, in the case of IoD, both speech synthesis and sonification will have a major role to play in improving perceived ease of use [38]. At the same time, tactile and haptic feedback can also be expected to gain increased traction. Often, tactile feedback can be implemented in the same device that is used for input. Tactile feedback can be either electrotactile, or more generally vibrotactile, however, in both cases, the spatial resolution and dynamics of sensitivity (saturation) in different parts of the body (and of different parts of the hand) need to be taken into consideration.

### C. *Hardware developments*

Although the software part (i.e., signal processing, AI, simulations, rendering methods) of IoD is the first component we think of, hardware development especially in case of I/O devices will gain increased. Ergonomic, intuitively designed and user-friendly interfaces cover microphones, 2-channel or multi-channel speaker arrangements, 2D and 3D screen miniaturization and placement, vibrators or new hardware elements integrated into a one-in-all solution. Furthermore, hardware design has to focus on special areas and user needs (users with disabilities, visually impaired users, users with hearing impairments, children, and elderly people).

Controllers for gaming allow navigation not just in the up/down/left/right, but also in the forward/backward directions, thus, depth information is covered in case of VR scenarios. This extra degree of freedom will be essential for IoD, thus, game controllers could be the next solution for navigation.

Furthermore, vehicle simulators include pedals for controlling speed and braking, which can introduce another channel for control. VR scenarios are widely used in action and FPS games, and experience gathered in this field can be transferred to IoD devices (Fig. 2).

Fig. 2. Classic design for a game controller used for PS4. Ergonomic design helps two-handed handling. Some controllers have integrated tablet holders or steering wheel and pedal extensions.

It is of great importance that real-time feedback should be provided following input. If we type on a keyboard looking at the screen, letters will be displayed and errors can be detected easily. Accessing devices for IoD also have some kind of screen for the same purpose, along with sound and haptics.

For all I/O devices the question of cables have to be discussed. Although wired connections are still more reliable, the freedom of movement and mobility will also play a role in future developments. Miniaturization, energy supply, high-speed ultra-reliable wireless connections between I/O devices and network as well as in and between networks (5G) will highlight new development strategies.

### D. *I/O devices supporting IoD applications*

Based on the recent advances in Digital Reality and Internet of Digital Reality, as well as several other related fields, it seems to be the case that modern computational paradigms are evolving towards less platform-dependent use-cases than their more traditional counterparts. Indeed, DR and IoD (as well as CogInfoCom, even earlier) all emphasize cognitive capabilities and functionally holistic perspectives over any

specific medium, format or modality. Hence, above all else, I/O devices that support interoperability with a wide variety of platforms can be expected to be favored in IoD applications. This constraint, in turn suggests that much of the processing behind I/O can be expected to become increasingly outsourced into the cloud, where cutting-edge AI solutions can be deployed more effectively. At the same time, as described in [39], interfaces can be expected to become increasingly dynamic, through an understanding of human behaviors, human limits, human needs and human cognition.

## III. CONCLUSIONS

This paper briefly introduced the concept of IoD, focusing on the pillar of access devices and methods for input and output. IoD will include cognitive digital entities where "things", AIs, algorithms and human users will interact and communicate. Moving toward a fully immersive virtual scenario, new possibilities emerge, but increased cognitive load may be an issue. Input devices will use more speech-based commands and high-quality speech recognition methods. Output devices will have more opportunity using sonification and spatial audio rendering methods. In the case of haptics, glove-like controller devices are expected as I/O devices with the function of clicking and scrolling, and at the same time, vibrators allowing tactile/haptic feedback. Full body movements and body parts can be also used for input. Besides signal processing, hardware development, psycho- and vibroacoustics, an increasing evolution towards cloud-based personalized AI solutions can be expected to play a key role in I/O-related R&D behind IoD applications.

## ACKNOWLEDGMENT

This research was supported by the Digital Development Center in the national framework GINOP-3.1.1-VEKOP-15-2016-00001 "Promotion and support of cooperations between educational institutions and ICT enterprises".

## REFERENCES

[1] C. Srinivasan, B. Rajesh, P. Saikalyan, K. Premsagar, and E.S. Yadav, "A Review on the Different Types of Internet of Things (IoT)," Journal of Advanced Research in Dynamical and Control Systems, vol. 11, no. 1, pp. 154–158, 2019.

[2] D. Jones, C. Snider, A. Nassehi, J. Yon, and B. Hicks, "Characterising the Digital Twin: A systematic literature review," CIRP Journal of Manufacturing Science and Technology, vol. 29, pp. 36–52, 2020.

[3] S. Greengard, The Internet of Things, MIT Press, 2015.

[4] G. Sallai, "The cradle of cognitive infocommunications," Acta Polytechnica Hungarica, vol. 9, no. 1, pp. 171-181, 2012.

[5] Cisco, "The Internet of Everything - Global Private Sector Economic Analysis,"https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoE_Economy_ FAQ.pdf, 2013.

[6] D. Williams, "The Internet of Everything - Cisco IoE Value Index Study," https://www.huffpost.com/entry/the-history-of-augmented-_b_9955048, 2017.

[7] P. Baranyi, Á. Csapó, and G. Sallai, Cognitive Infocommunications (Coginfocom), Springer, 2015.

[8] L.I. Komlósi, and P. Waldbuesser, "The cognitive entity generation: Emergent properties in social cognition," in 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 439-442, 2015.

[9] P. Baranyi, and Á. Csapó, "Revisiting the concept of generation CE-Generation of Cognitive Entities," in 6th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), pp. 583-586, 2015.

[10] P. Baranyi, and Á. Csapó, "Definition and synergies of cognitive infocommunications," Acta Polytechnica Hungarica, vol. 9, no. 1, pp. 67–83, 2012.

[11] P. Baranyi, Á. Csapó, T. Budai, and G. Wersényi, "Introducing the Concept of Internet of Digital Reality - Part I," Acta Polytechnica Hungarica, vol. 18, no. 7, pp. 225–240, 2021.

[12] G. Wersényi, Á. Csapó, T. Budai, and P. Baranyi, "Internet of Digital Reality: Infrastructural Background - Part II," Acta Polytechnica Hungarica, vol. 18, no. 8, pp. 91–104, 2021.

[13] P. D. Ramani Moses, Nikita Garia, "Digital Reality – A technical primer," https://www2.deloitte.com/insights/us/en/topics/emerging-technologies/ digital-reality-technical-primer.html, 2021.

[14] L. Xue, C.J. Parker, and H. McCormick, "A virtual reality and retailing literature review: Current focus, underlying themes and future directions," Augmented Reality and Virtual Reality, pp. 27–41, 2019.

[15] M. Handosa, H. Schulze, D. Gracanin, M. Tucker, and M. Manuel, "An Approach to Embodiment and Interactions with Digital Entities in Mixed-Reality Environments," in 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 569–570, 2018.

[16] J. Katona, "A Review of Human–Computer Interaction and Virtual Reality Research Fields in Cognitive InfoCommunications," Applied Sciences, vol. 11, no. 6, p. 2646, 2021.

[17] D. Yu, L. Deng, Automatic Speech Recognition - A Deep Learning Approach. Springer, 2015.

[18] V. Këpuska, and G. Bohouta, "Next-generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home)," in 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), pp. 99-103, 2018.

[19] A. Reis, D. Paulino, H. Paredes, I. Barroso, M.J. Monteiro, V. Rodrigues, and J. Barroso, "Using intelligent personal assistants to assist the elderlies An evaluation of Amazon Alexa, Google Assistant, Microsoft Cortana, and Apple Siri," in 2018 2nd International Conference on Technology and Innovation in Sports, Health and Wellbeing (TISHW), pp. 1-5, 2018.

[20] G. Hackenberg, R. McCall and W. Broll, "Lightweight palm and finger tracking for real-time 3D gesture control," 2011 IEEE Virtual Reality Conference, pp. 19-26, 2011.

[21] P. Vogiatzidakis, and P. Koutsabasis, "Mid-Air Gesture Control of Multiple Home Devices in Spatial Augmented Reality Prototype," Multimodal Technol. Interact., vol. 4, no. 3, pp. 1-22, 2020.

[22] Y. Liu, Z. Jiang, and H.C. Chan, "Touching Products Virtually: Facilitating Consumer Mental Imagery with Gesture Control and Visual Presentation," Journal of Management Information Systems, vol. 36, pp. 823-854, 2019.

[23] S. Akyol, U. Canzler, K. Bengler, and W. Hahn, "Gesture Control for use in Automobiles," in IAPR Workshop on Machine Vision Applications, The University of Tokyo, Japan, pp. 349-352, 2020.

[24] M. Latinus, and P. Belin, "Human voice perception," Current Biology, vol. 21, no. 4, pp. 143–145, 2011.

[25] C.V. Botinhao, and J. Yamagishi, "Speech intelligibility in cars: the effect of speaking style, noise and listener age," in Proc. of InterSpeech 17, pp. 2944-2948, 2017.

[26] J.S. Pardo, L.C. Nygaard, R.E. Remez, and D.B. Pisoni (Eds), The Handbook of Speech Perception, Second Edition, Wiley, 2021.

[27] J. Donahue, S. Dieleman, M. Bińkowski, E. Elsen, and K. Simonyan, „End-to-end adversarial text-to-speech," arXiv preprint arXiv:2006.03575, 2020.

[28] T. Hermann, A. Hunt, and J.G. Neuhoff (Eds), The Sonification Handbook. Logos Publishing House, Berlin, 2011.

[29] Á. Csapó, and G. Wersényi, "Overview of auditory representations in human-machine interfaces," ACM Computing Surveys (CSUR), vol. 46, no. 2, pp. 1–23, 2013.

[30] K. Tislar, Z. Duford, B. Nelson, M. Peabody, and M. Jeon, "Examining the Learnability of Auditory Displays: Music, Earcons Spearcons, and Lyricons," in Proc. of ICAD, pp. 197-202, 2018.

[31] T.S. Amer, and T.L. Johnson, "Earcons versus auditory icons in communicating computing events: Learning and user preference," International Journal of Technology and Human Interaction, vol. 14, no 4, pp. 95-109, 2018.

[32] P.P. Lennox, J.M. Vaughan, and T. Myatt, “3D Audio as an Information-Environment: Manipulating Perceptual Significance for Differentiation and Pre-Selection,” in Proc. of ICAD, pp. 155-160, 2001.

[33] Á. Csapó., et al. "VR as a medium of communication: from memory palaces to comprehensive memory management." 2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom). IEEE, 2018.

[34] K. Ooi, J-Y. Hong, B. Lam, Z-T. Ong, and W-S. Gan, “Validation of 3D headphones for use in soundscape evaluation,” INTER-NOISE and NOISE-CON Congress and Conference Proceedings, InterNoise17, Hong Kong, pp. 4651-4659, 2017.

[35] S. Serafin, M. Geronazzo, C. Erkut, N.C. Nilsson and R. Nordahl, “Sonic Interactions in Virtual Reality: State of the Art, Current Challenges, and Future Directions,” in IEEE Computer Graphics and Applications, vol. 38, no. 2, pp. 31-43, Mar./Apr. 2018.

[36] V. Hohmann, R. Paluch, M. Krueger, M. Meis, and G. Grimm, ”The Virtual Reality Lab: Realization and Application of Virtual Sound Environments,” Ear Hear. 2020;41 (Suppl 1):31S-38S, 2020.

[37] M. Vorländer, Auralization - Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality, Springer/ASA Press, 2020.

[38] A. Baird, E. Parada-Cabaleiro, S. Hantke, F. Burkhardt, N. Cummins, and B. Schuller, “The Perception and Analysis of the Likeability and Human Likeness of Synthesized Speech,” in Proc. of InterSpeech 18, pp. 2863-2867, 2018.

[39] Á. Török. "From human-computer interaction to cognitive infocommunications: a cognitive science perspective." 2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom). IEEE, 2016.