



Augmenting perception: How artificial intelligence transforms sensory substitution



Louis Longin ^{a,*}, Ophelia Deroy ^{a,b,c}

^a Faculty of Philosophy, Philosophy of Science and the Study of Religion, LMU-Munich, Geschwister-Scholl-Platz 1, 80359 Munich, Germany

^b Munich Center for Neurosciences-Brain & Mind, Großhaderner Str. 2, 82152 Planegg-Martinsried, Germany

^c Institute of Philosophy, School of Advanced Study, University of London, London WC1E 7HU, United Kingdom

ARTICLE INFO

Keywords:

Sensory substitution
Sensory extension
Intelligent sensory augmentation
Information quality
Senses
Artificial intelligence

ABSTRACT

What happens when artificial sensors are coupled with the human senses? Using technology to extend the senses is an old human dream, on which sensory substitution and other augmentation technologies have already delivered. Laser tactile canes, corneal implants and magnetic belts can correct or extend what individuals could otherwise perceive. Here we show why accommodating intelligent sensory augmentation devices not just improves but also changes the way of thinking and classifying former sensory augmentation devices. We review the benefits in terms of signal processing and show why non-linear transformation is more than a mere improvement compared to classical linear transformation.

1. Introduction

Artificial sensors now largely outperform human sensory capacities: artificial noses can identify thousands of odours (Hu, Wan et al., 2019) and distinguish between an infected and a non-infected wound (Haalboom et al., 2019); driverless cars detect secluded objects with laser radar and infra-red cameras (Pulikkaseril & Lam, 2019); robots can use photosensors to recognise materials by their sounds (Eppe et al., 2018). But what happens if humans get equipped with such artificial sensors? Can these sensors be genuinely coupled with humans in such a way that they extend our perception beyond the use of external tools? And if so, do they really represent a substantial novelty compared to previous sensory substitution and extension devices?

Sensory augmentation builds on the idea that the senses can be modulated and even enhanced through sensory technology. This coupling is of a profound, interactive nature and arguably extends our cognition (Kiverstein & Farina, 2012; Wheeler, 2015). Sensory substitution, in particular, captures the process of transferring sensory signals from one sensory modality to another. One common and successful application has been the transfer of visual information to sound to compensate for a vision impairment like ‘the vOICe’ (Auvray et al., 2007; Meijer, 1992; Proulx et al., 2008). Other applications include vision-to-tactile sensory substitution devices (SSDs) like TVSS (Arnold & Auvray, 2018; Bach-Y-Rita et al., 1969) and vestibular-to-tactile SSDs (Tyler et al., 2003). Findings on neural plasticity, for instance, have demonstrated that SSDs can at least partially restore a lost sense through neural re-organisation and practice (Amedi et al., 2007; Bach-y-Rita & W. Kercel, 2003; Cohen et al., 1997; Collignon et al., 2008).

Much of the applications of artificial intelligence to sensory augmentation are in their infancy. Pairings of sensors and AI are developing in various directions and are driven by marketing and technological opportunities, rather than a systematic taxonomy:

* Corresponding author.

E-mail address: louis.longin@campus.lmu.de (L. Longin).

'Smart wearables' (see [Saganowski et al. \(2020\)](#) for review) encompass AI-infused smartwatches and wristbands, whereas internet-connected textiles seem to fall under the different label of 'smart clothing' ([Chen et al., 2016](#); [Fernández-Caramés & Fraga-Lamas, 2018](#)). Additionally, intelligent prosthetics and advanced sensory substitution utilise AI to improve users' acceptance and performance under the traditional prosthetics and sensory substitution frameworks ([Hu, Wang et al., 2019](#); [Pinheiro Lima Neto et al., 2019](#)).

To understand the impact of AI in sensory augmentation on the field of sensory augmentation, what is needed first is a map of the space of conceptual possibilities. A systematic taxonomy of what already exists, but also of what is possible, requires two things: the first one is a conceptual delimitation of the domain of sensory augmentation; the second is an analysis of what distinguishes kinds of augmentation devices ([Section 2](#)). Here, we suggest a new, less controversial principle that distinguishes sensory augmentation devices by their input and output rather than their perceptual function. Based on this new way of distinguishing sensory augmentation devices, we explore how AI can be integrated into sensory augmentation and how intelligent sensory augmentation (ISA) changes the traditional taxonomy ([Section 3](#)). Then the important thing to ask is whether the proposed difference by ISADs in mapping, from linear to non-linear, genuinely changes sensory augmentation technologies as they currently stand or simply provide ways to improve them ([Section 4](#))? The 'mere improvement' view, we argue, misses fundamental ways in which AI changes sensory augmentation: ISAD genuinely provide a new kind of augmentation based on improving the quality rather than the quantity of sensory information delivered to the user. By clarifying this conceptual issue in this paper, we establish clear grounds for the future development of sensory augmentation.

2. Non-intelligent sensory augmentation

2.1. Defining sensory augmentation

The idea of improving human perception through wearable devices is not new: arguably, it goes back far into popular culture (RoboCop, Ghost in the Shell, Inspector Gadget) and now is realised by prosthetics, sensory substitution and extension devices. Here we count technologies as sensory augmentation, as long as they deliver additional sensory cues to convey pertinent information for a perceptual task. Sensory augmentation, in other words, requires:

- (i) that the **input** of a sensory augmentation device is a **sensible** property, set of properties or object,
- (ii) that the **output** of a sensory augmentation device is causally related to the input and delivered as additional information to the user in a **sensory** format,
- (iii) with the **goal** to provide or improve perceptual functions.

The first requirement on the input places restrictions on the kind of inputs that lead to sensory augmentation but also deserves to be qualified. The main issue is whether one should count virtual reality (VR) glasses as 'augmenting perception' as they also display additional information in a sensory format. In many cases, the objects perceived in VR are not generated from sensible properties or objects: VR glasses used in gaming, for instance, do not gather sensible properties from the user's internal or external environment but generate their displayed objects. Hence, here VR glasses do not perform sensory augmentation.

Following [Dilworth \(2010\)](#), it is true that the term "VR" covers only a loosely related set of technologies: there are some cases where VR takes real objects as inputs, albeit usually distant ones; for instance, when a surgeon perceives in VR the organ of a distant patient that he is operating on. In such cases, the input is a sensible - rather than a purely algorithmically generated - object, and a causal relation obtains between the input and the sensory-delivered output. Such cases, arguably, can count as sensory augmentation but also bring attention to a possible disjunction in the first condition: the input needs to be a sensible property or object, and this sensible object is typically in one's immediate environment. However, it could also eventually be in one's remote environment. In this second case, the spatial relations between the perceiver's body and the object seen in VR are not causally related to the actual spatial relations between the perceiver's body and the object used as input, which opens questions. We leave the second category open to debate and focus on inputs that are clearly taken in one's immediate environment.

The very concept of 'immediate' vs 'remote', we reckon, is also eventually graded: the patient in Oslo operated upon by a doctor in Cape Town is obviously spatially remote, as there is no other way it could be in a perceptual relation to the surgeon. However, the cup captured by the head-mounted camera of a sensory substitution device like the vOICE ([Meijer, 1992](#)), which is 50 cm away from the perceiver, is clearly in one's immediate environment as the perceiver can also touch it. An object that would be 500 m away from a driver and is delivered to the senses via a technological device would be an intermediate case.

If the first requirement addresses VR cases, the second one separates sensory augmentation from both cognitive augmentation and sensory tools. Cognitive augmentation devices provide additional but symbolic or linguistic information. The human perceiver then uses this information to form cognitive judgments. Examples include extended-memory devices ([Lee et al., 2016](#); [Smart, 2017](#)), personal assistants ([Canbek & Mutlu, 2016](#); [Hoy, 2018](#)); and many digital smart wearables ([Fernández-Caramés & Fraga-Lamas, 2018](#); [Sun et al., 2017](#)): smartwatches, for instance, take a sensible property like the heart rate as input and deliver it as a numerical value; digital personal assistants or car-based navigation systems produce verbal outputs that facilitate cognitive tasks like navigation. Similarly, text-to-speech devices classify as cognitive augmentation devices. They take in sensory signals but produce a linguistic output. Because these devices produce information in a linguistic and not strictly sensory format, they cannot be considered sensory augmentation.

At the other end of the spectrum, sensory tools like ordinary glasses or a cane transfer information in a sensory format but fail to add information ([Wright & Ward, 2018](#)). The long cane represents a popular example as it enables the blind to detect obstacles through

extension of their tactile field. The underlying transfer of tactile information relies on the sensory capacities of the user. Because sensory tools only mediate sensory information, they, however, cannot be considered as sensory augmentation.

The third requirement holds that sensory augmentation devices have to provide or improve a perceptual function. A perceptual function can be detecting, locating, discriminating or identifying properties and objects in the environment. This excludes devices such as smart textiles where the added sensory output is only the source of additional sensations. Smart textiles such as vibration motors in jackets can provide new tactile sensations on the skin. However, those sensations do not tend to be constructed as the perception of objects or serve a specific perceptual function (see [Tajadura-Jiménez et al. \(2020\)](#) for review and discussion). The output from sensory augmentation devices has to link to some environmental property (including the internal environment, i.e. the body) and help to detect changes in that property. Positive examples include sensory substitution devices such as vOICe, where visual information is recorded and then transferred into auditory frequencies. The produced auditory information improves the perceptual function of spatial awareness by linking changes in the visual field to changes in the produced auditory frequencies.

The conceptual borders between these three domains (cognitive augmentation, sensory augmentation, and mere tools) are clear (see [Table 1](#)). However, fitting actual devices within each category remains a matter of controversy: some see the same technologies for the blind as sensory augmentation ([Kärcher et al., 2012](#)), or as more general as ‘mind-enhancing tools’ ([Auvray & Myin, 2009, p. 1036](#)) or as providing a ‘new set of automatic recognition abilities’ that are not purely perceptual ([Deroy & Auvray, 2014, p. 343](#)). Others would consider that an embodied approach to objects like white canes even blurs the divide between an extension of one’s body and a tool ([Murray, 2008](#)).

2.2. The distinction by function

The function that devices support or perform has been the dominant, but also not uncontroversial, criterion used to classify sensory augmentation devices into different categories.

On the face of it, distinguishing sensory augmentation devices according to their functions makes sense. It allows us to distinguish devices in virtue of their role for the user: adding a new sense, deferring sensory information across the senses to compensate for a sensory loss, or restoring a deficient sense. Sensory substitution devices illustrate the usefulness of such a distinction even to the specific device level. Sensory substitution builds on the idea of replacing sensory information from one sensory modality with sensory information from another. SSDs have achieved a substantial amount of success over the last decades: with the original Tactile Visual Substitution System (TVSS), participants were able to identify visual objects through tactile stimulation within ten seconds after only five rounds of training ([Bach-Y-Rita et al., 1969](#)). Modern versions include vision-to-tactile devices such as TVSS ([Bach-y-Rita & W. Kercel, 2003](#)), vision-to-audition devices such as vOICe ([Meijer, 1992](#)) or vestibular-to-tactile devices ([Tyler et al., 2003](#)). Distinguishing these devices by their function for the user, such as providing ‘visual’ or ‘vestibular’ awareness, provides seemingly clear perceptual and empirical boundaries.

Controversies arise, however, as the individuation by function depends both on views about perception and on how the output is coupled with the existing sensory capacities of the user. The arguments in these debates depend for a large extent on the theory of perception that one embraces, be it direct, enactive or representationalist. Empirical evidence on the phenomenology of the SSD-induced perceptual state enforces this conceptual divide. In principle, the SSD-induced perceptual state can either be associated with an experience in the substituting or the substituted sensory modality or lead to a new kind of phenomenal experience. The advocates of the sensorimotor view have argued that the substituting modality defers its perceptual capacity to the substituted modality ([Hurley & Noë, 2003](#)). In the case of a blind person using a SSD-based visual aid, this means that the blind person sees through SSD’s auditory cues. Contrary, proponents of the representationalist camp have argued that the substituting modality remains dominant ([Block, 2003](#)). In other words, the blind person does not see but instead possesses an enhanced auditory sense. Others like [Kiverstein et al. \(2015\)](#) have argued that one should classify the SSD-induced perceptual experience as part of a new sensory modality. And even others debate whether sensory substitution is only akin to sensory perception through the rise of a new sensory experience or whether it also involves a perceptual judgment or practical skill ([Deroy & Auvray, 2012](#)).

Ultimately, these controversies show that classifying sensory augmentation devices by their performed perceptual function is not as easy as initially thought. Philosophical debates show that functional individuation depends on the endorsed theories of perception, which are as controversial as wide-ranging. Empirical research amplifies this dissonance by struggling to explicate which phenomenal state is induced ([Auvray & Farina, 2017; Deroy & Spence, 2013; Farina, 2013; Nanay, 2017; Proulx, 2010](#)) and how cross-modal plasticity is interpreted ([Amedi et al., 2007; Collignon et al., 2011; Ptito et al., 2018](#)).

Table 1
Summary of sensory augmentation criteria with representative examples.

Criteria for sens. aug. device	Long cane (sens. tools)	Corneal Implants (Prosthetics)	Magnetic belt (sens. exten.)	vOICe (SSD)	VR (fiction)	Smart textiles	Navigation (cogn. aug.)
(i) sensible properties as input	Yes	Yes	Yes	Yes	No	Yes	Yes
(ii) additional, sensory output	No	Yes	Yes	Yes	Yes	Yes	No
(iii) perceptual function	Yes	Yes	Yes	Yes	No [?]	No	Yes [?]

2.3. The distinction by input and output

One way to avoid these controversies is to use a different way of distinguishing categories of sensory augmentation. Here we suggest that looking at the signal processing between input and output is a more robust way to distinguish kinds of sensory augmentation devices, i.e. focusing on criteria (i) and (ii) instead of (iii) of sensory augmentation (see [Section 2.1](#)).

The information processing in a sensory augmentation device occurs through three main components: an artificial sensor, a coupling system and a stimulator ([Elli et al., 2014](#); [Wright & Ward, 2018](#)). The artificial sensor receives the incoming sensory information. The stimulator outputs a sensory signal, and the coupling system connects the artificial sensor with the stimulator. In the sensory substitution terminology, the artificial sensor records the sensory information in the substituted sensory modality, and the stimulator outputs sensory information in the substituting modality. From a technological perspective, the artificial sensor and the stimulator are utilised in hardware while the coupling system connects both pieces of hardware with software.

This process of sensory substitution in SSDs and sensory augmentation more broadly involves implementing a conversion algorithm as the coupling system. The algorithm records the sensory input in one sensory modality, transforms it into a different sensory signal and outputs it into the desired sensory modality. The implementation of the conversion algorithm allows establishing a cross-modal, non-physical connection between artificial sensors and artificial stimulators. The human user must learn how to make sense of the output signal, which is the case for SSDs after extensive training.

By focusing on how the input (i) and the output (ii) are handled, we can clearly distinguish current prosthetics, sensory substitution, and extension devices. Sensory prosthetics and sensory substitution devices are both kinds of sensory augmentation that focus on capturing input information that has been degraded or gone missing (i): such as visual information for people with partial or full blindness. The technologies add missing sensory information that is usually present, though the extent to which they match the one provided by healthy organs varies (ii). Hence, we can distinguish prosthetics and SSDs in virtue of their output (ii) and not only through their performed perceptual function (see [Table 2](#)).

In particular, prosthetics and SSDs vary in two dimensions: whether sensory information is translated into a different sensory modality and how the sensory information is provided to the user. Borrowing [Wright & Ward's \(2018\)](#) terminology, sensory substitution is mostly based on a *between-sense* referral, while prosthetics utilise a *within-sense* referral. These different kinds of referral describe how sensory information is modulated by the sensory augmentation device. *Between-sense* referral devices record and transfer some incoming sensory information to a different sensory modality. In contrast, *within-sense* referral devices retain the sensory modality of the input. Prosthetics or other implants operate *within* a sensory modality: they gather sensory information, amplify it and then forward it within the same sensory modality. SSDs mostly operate *between* sensory modalities: they also add otherwise missing information, but they make it available to another sense such as vision-to-tactile devices TVSS ([Bach-y-Rita & W. Kercel, 2003](#)) and TDU ([Tyler et al., 2003](#)), or vision-to-audition devices vOICe ([Meijer, 1992](#)) or Vibe ([Auvray et al., 2005](#); [Hanneton et al., 2010](#)).

Another difference between prosthetics and SSDs is how the sensory information is provided to the user: through *invasive* or *non-invasive* methods. While prosthetics rely on *invasive* methods such as cochlear, vestibular or corneal implants ([Golub et al., 2014](#); [Zeng et al., 2008](#)), sensory substitution and extension devices use *non-invasive* methods such as external wearable devices, which can be taken on and off.

An additional conceptual distinction that can be drawn with the proposed input/output individuation is the distinction between sensory extensions, sensory substitution devices, and prosthetics. Sensory extension differs from SSDs and prosthetics in the nature of the gathered input (i). Instead of gathering missing but normally available sensory signals, sensory extension devices capture a sensory property not normally available to the human senses. As a result, they add novel information rather than compensate for a missing sense. For instance, [Nagel et al. \(2005\)](#) developed a magnetic belt that grants human perceivers access to a magnetic sense by translating magnetic information into felt tactile vibrations (see also [Hameed et al. \(2010\)](#); [Kärcher et al. \(2012\)](#)).

Traditionally, sensory substitution and extension devices have operated *across senses* as they translate a physical property that can be sensed by typical humans or non-human animals (e.g. magnetic field) into sensory information accessible to an existing modality. *Within-sense* sensory devices fall in a grey area: they fulfil the criteria for sensory augmentation. However, they may in some cases transmit sensory properties that are in principle perceivable by the human user. Take, for example, night vision goggles, which transmit visual information also to the visual modality. If they introduce changes in some sensible properties, such as orientation, size or intensity, or extrapolate previously invisible contrasts and make otherwise non-perceptible objects visible, they should count as sensory extension devices. But if the transmitted sensory properties are perceptible in other contexts (e.g. in better illumination conditions), *within-sense* sensory devices only seem to provide a local, contextual augmentation.

The advantage of approaching sensory augmentation via the input/output characteristics is two-fold. On the one hand, the approach is less controversial. The individuation by input/output does not presuppose a certain theory of perception to understand the role the device performs for the user. On the other hand, the individuation by input/output remains open to classifying new kinds of sensory augmentation devices such as those based on artificial intelligence. Intelligent sensory augmentation devices (ISAD), as we will

Table 2
Classification of traditional categories of sensory augmentation through the distinction by input/output.

Sensory Augmentation Distinctions	Prosthetics	Sensory Substitution	Sensory Extension
(i): compensation / novelty	compensation	compensation	novelty
(ii): within / between sense referral	within	between	both
(ii): invasive / non-invasive	invasive	non-invasive	non-invasive

show in the next section, introduce a new way of transforming input to output sensory signals. This is not captured under the traditional framework of functional individuation.

3. How sensory augmentation becomes intelligent

3.1. The pressure to become intelligent

SSDs have achieved a substantial amount of success over the last decades: with the original Tactile Visual Substitution System (TVSS), participants were able to identify visual objects through tactile stimulation within ten seconds ([Bach-Y-Rita et al., 1969, p. 19](#)); in recent experiments, some blind users of SSDs ([Striem-Amit et al., 2012](#)) managed to perform so well that they exceeded the threshold for the World Health Organization definition of blindness, meaning that they were at least “legally” no longer functionally fully blind but rather on par with the severely visually impaired (see [Maidenbaum et al. \(2014\)](#) for review).

However, despite the success of these between-sense SSDs in the laboratory environment, SSDs and SADs have not been extensively utilised outside the lab ([Auvray & Harris, 2014](#); [Lloyd-Esenkaya et al., 2020](#)). This goes back to two main reasons: cognitive overload and perceived low usability ([Elli et al., 2014](#)). Cognitive overload describes the phenomenon that human subjects can feel overwhelmed and stressed by the amount of additional sensory information. Only through extensive training the human user can distinguish sensory noise from meaningful sensory signals ([Reynolds & Glenney, 2012](#); [Striem-Amit et al., 2012](#); [Ward & Meijer, 2010](#)). Training is also required with retinal implants ([Dagnelie, 2012](#)) or sensory extension devices ([Auvray & Myin, 2009](#); [Neugebauer et al., 2020](#)). Hence, for SSDs like vOICe to become effective, the task of making sense of the additional information is fully unloaded onto and has to be solved by the human user through extensive training. However, even after the human user has learned to use the SAD sufficiently, the overall usability is perceived as low. The human user constantly needs to exert cognitive effort to distinguish and interpret SAD-induced sensory information. The added value through using SSDs compared to not using SSDs has been perceived as low because the enhanced sensory awareness does not sufficiently outweigh the cost of the exerted effort.

These challenges have pushed the field of sensory substitution and sensory augmentation at large to turn towards new solutions for improving the usability of sensory devices. Traditional approaches have focused on improving training schemes by adapting the training scheme to the individual user ([Chebat et al., 2015](#); [Stronks et al., 2015](#)) or providing different settings to different users ([Brown et al., 2011](#)). However, with the recent success in artificial intelligence, a new approach has emerged: combining machine learning methods with existing sensory substitution and augmentation technologies.

3.2. Two ways of integrating AI into sensory augmentation

The term ‘artificial intelligence’ (AI) has no single agreed definition. A well-known modern distinction is outlined by [Russell & Norvig \(2016\)](#), who conceptualise AI under four categories: thinking humanly, thinking rationally, acting humanly and acting rationally. These categories vary in their ambition of what AI seeks to accomplish and their notion of intelligence. In the narrowest sense, AI represents a human-like thinking system ([Haugeland, 1985](#)). In the broadest sense, AI performs complex, intelligent behaviour remaining neutral or even agnostic on any computational similarities between humans and AI. We understand intelligence more broadly as the ability to achieve goals in different environments ([Legg & Hutter, 2007](#)) or as the ability to acquire skills efficiently ([Chollet, 2019](#)). This broader understanding of AI as a computer program that learns and adapts is sufficient to capture the ongoing development of AI-based sensory augmentation.

A wide range of AI methods has been used to manipulate sensory signals. The field of robotics utilises three different kinds of sensors (environmental, spatial and proprioceptive) to learn how to interact with the world reliably ([Russell & Norvig, 2016](#)). The field of computer vision focuses on extracting primitive and complex features like edges, texture and objects from low-level visual signals. One key driver for the developments in both fields has been the progress in machine learning. Through the use of data and training time, machine learning algorithms can learn to improve their performance automatically. Some learning goals include pattern recognition of unstructured data (unsupervised learning), input-output matching of structured data (supervised learning), and strategy development for reward maximisation (reinforcement learning).

Performing machine learning algorithms requires a model that specifies how data is processed and where the training occurs. The most predominant model has been the artificial neural network (ANN). Inspired by the biological brain, artificial neurons are interconnected and learn to adapt their connection strength to optimise the ANN’s performance. Application areas include signal processing ([Albawi et al., 2017](#); [Egmont-Petersen et al., 2002](#)), as well as natural language processing ([Goldberg, 2017, 2016](#)) and decision-making models ([Hramov et al., 2018](#); [Zhang et al., 2019](#)).

There are two ways AI can be implemented in sensory augmentation devices: *before* and *after* the output is presented to the user. If AI is introduced *before* the output is presented, then AI changes how sensory information is translated within or across sensory modalities. If AI is introduced *after* the output is presented to the user, then AI facilitates the encoding process by the user.

Intelligent sensory augmentation devices utilising AI *after* the output have implicitly already been developed. Here, AI methods are used to evaluate the quality of the final sensory signal and hence provide information on how signal quality can be improved. For example, [Kim et al. \(2021\)](#) use a cross-modal generative adversarial network-based evaluation method to find an optimal auditory sensitivity to reduce transmission latency in visual-auditory sensory substitution. [Hu, Wang et al. \(2019\)](#) use machine learning to evaluate different encoding schemes for a visual-to-auditory SSD based on the user’s needs. Given that the late-blind and congenitally-blind differ in their previous exposure to visual stimulation, different encoding schemes are needed to facilitate the recognition of ‘visual’ objects through sound. Late-blind users, in contrast to congenitally-blind users, can utilise pre-existing visual experiences,

which provide a useful reference for any cross-modal perception. While these modern technologies have shown how AI can improve classical sensory augmentation schemes *after* the output is presented to the user, we believe that the substantial improvement has yet to come: by implementing AI *before* the output.

The basic idea for utilising AI before the output and as part of processing incoming sensory signals is illustrated by an environmental navigation study by [Kerdegari et al. \(2016\)](#) who developed an ultrasonic helmet that translates ultrasonic radar reflections into tactile feedback. The conversion algorithm is instantiated through a multilayer perceptron neural network. In a set of experiments, the participants were asked to avoid obstacles and move in a specific direction. The helmet's sensors gather environmental information, then computed and forwarded as simplified and task-specific signals to the human user. [Kerdegari et al. \(2016\)](#) found that participants perceive less cognitive load and reach the goal more reliably when the AI-driven helmet forwards its computation as tactile instead of linguistic signals.

This approach is fundamentally qualitative as additional processing steps are implemented in the sensory helmet that performs task-specific directional cueing. Instead of forwarding a high degree of quantitative information, the sensory helmet performs specific perceptual pre-processing tasks such as camera-based object detection and navigation. This helmet offers a first glimpse at how intelligent pre-processing could lead to sensory augmentation, at least if the forwarded output also provides additional sensory cues about the environment, which can serve a perceptual function such as shape recognition or depth perception.

Besides neural networks, [Wright & Ward \(2013\)](#) have used genetic algorithms - a different machine learning method - to overcome the challenges of traditional SSDs of cognitive overload and low usability. [Wright & Ward \(2013\)](#) have used genetic algorithms to 'evolve' efficient signal encoding schemes. Genetic algorithms are a stochastic search method that uses evolutionary principles to find the 'fittest,' i.e. optimal solution to a search problem (for more, see [Haupt & Haupt \(2003\)](#)). Their interactive genetic algorithms expand the fitness, i.e. optimisation function to include human input. The inclusion of user input allows these interactive genetic algorithms to incorporate an aspect of the user experience, such as ease of use, into the evolutionary process. In this early example of an ISAD called 'Polyglot', [Wright & Ward \(2013\)](#) reimplemented the conversion algorithm of the 'vOICe' by mapping visual signals to sounds. While retaining some conversion principles such as utilising frequency to represent a vertical position, the 'Polyglot' varies and evolves other parameters such as frequency allocation, frequency range and contrast enhancement freely. Subsequent experiments validate the idea of adapting the SSD to the human user. [Wright & Ward \(2013\)](#) report a relatively quick convergence to an optimal balance between performance and ease of use. However, they also report that the optimal settings depend highly on the given task and limits of the sensory capacity.

Despite their limitations, both early and implicit cases of ISAD highlight the wide range of AI methods that can be employed to improve and transform the field of sensory augmentation technologies.

3.3. What makes sensory augmentation intelligent: Linear vs non-linear signal processing

The key difference between ISADs and SADs is the shift in computational processing of the sensory signals. Instead of translating and forwarding the sensory signals through a linear relation between input and output, ISADs can match input and output signals non-linearly. In the context of signal processing, the concept of linearity describes the relation of the incoming to the outgoing signals. Signals bearing a linear relation link a change of the input signal directly to a change in the output signal. Signals bearing a non-linear relation, on the other hand, do not necessarily match a change of the input signal with a change in the output signal. This computational shift allows ISADs to recognise complex, non-linear patterns and extend the role of a mere sensory converter to a sensory pre-processor.

Traditionally, as an example of non-intelligent sensory augmentation, SSDs have utilised a linear mapping of recorded to output sensory information. One of the original SSDs, the vOICe, utilises a linear mapping as a coupling system to convert visual to auditory information. The original mapping algorithm developed by [Meijer \(1992\)](#) instantiates a linear mapping that is "as direct and simple as possible" (p. 113) to "reduce the risk of accidentally filtering out important clues" (*ibid*). The assumption was that the "human brain is far superior to most if not all existing computer systems in rapidly extracting relevant information from blurred and noisy, redundant images" (*ibid.*). The vOICe records incoming visual signals as greyscale values and translates them into auditory frequencies. The vOICe matches the location of the incoming signal with the frequency of the output (the higher the signal, the higher the frequency) as well as the brightness of the incoming signal with the loudness of the input (the brighter the signal, the louder the output). Both relations are linear and match a change of the incoming signal with a change in the outgoing signal. The user then learns to reconstruct the decoded image through the presented sound pattern.

Contrast this with an AI-driven sensory augmentation device: by extending the sensory processing to non-linear models, an ISAD is capable of much more than the traditional SSD. Using non-linear filtering algorithms from computer vision, an ISAD is, for example, able to 1. reduce the sensory complexity to its essential features like edges for images, 2. perform sensory classification like navigation for collision avoidance, 3. integrate a wide range of sensory signals simultaneously, and 4. generate complex, novel sensory patterns from incoming signals.

Kerdegari et al.'s tactile helmet, for instance, uses ultrasound sensors to sense the environment and issues haptic navigation commands to avoid any collision. Functionally, the tactile helmet uses a multilayer perceptron neural network to classify the incoming sensory data into navigation commands. This transformation represents a non-linear relation between the sensory data and the navigation command. Not every change in the incoming signal leads to a change in the outgoing signal. Instead, the neural network solves a non-linearly separable pattern classification problem by matching changing ultrasound patterns with relatively stable tactile signals. This non-linear reduction of initial sensory complexity greatly improves usability and reduces the user's cognitive overload.

When comparing a traditional non-intelligent SAD like the vOICe with an intelligent SAD, the differences become clear (see

Table 3). Due to a change in the underlying computational model from linear to non-linear, ISADs take on a more significant role in the computational processing of sensory signals. While classic SADs are designed to transfer the sensory signal as accurately as possible to the human user, ISADs can significantly modify the sensory signal. Instead of having to learn how to ‘make sense’ of the classic SAD’s sensory signals with an ISAD, the human user receives a much more refined sensory signal.

The main difference with ISA comes then not just from looking at the input or output but rather at how the change in mapping (from linear to non-linear) impacts how we can and should conceive of sensory augmentation with such intelligent forms in mind. By pre-computing and only forwarding minimally noisy sensory signals to the human perceiver, which convey rich environmental cues, ISADs provide substantial improvements to the process of sensory substitution as well as sensory augmentation.

3.4. Refining the input/output distinction

As demonstrated above, using AI at the input level to determine which information to send, or using AI at the output level to make the translated information easier to learn or decipher, are two ways in which intelligent sensory augmentation can occur. The first type, relying on a non-linear connection of incoming with outgoing signals, is the most transformative: such ISADs extend the range of human sensory substitution and augmentation by modulating and possibly enhancing the transferred sensory information. This approach is novel because the perceptual learning of making sense (interpreting) the modulated sensory information is no longer taken on in full by the human perceiver but is achieved now in parts by the self-learning coupling algorithm inside of ISADs.

The main difference between SSD and ISAD is then *how the gathered information is translated* into the output format. It is only through adding this new input/output criterion that one can capture the novelty of ISA: ISAD’s non-linear mapping gives up on the idea of preserving the structure of the original data and instead transforms the data non-proportionally. Existing definitions of sensory augmentation (Section 2) did not need to differentiate kinds of mapping because all mappings were de facto of the same kind. However, with the emerging integration of AI within sensory augmentation, the relation between the input and the output signal becomes a criterion with multiple values, which helps to distinguish two sub-types of augmentation, intelligent and non-intelligent. Again, by calling a SAD intelligent, we are not attributing ISADs human intelligence, but rather we highlight the implementation of AI models as part of the signal conversion algorithm. The fourth criterion for sensory augmentation, which we suggest introducing for this purpose, clarifies:

- (iv) that the sensory augmentation device either forwards low-level sensory information through a linear relation between incoming and outgoing signals (for classical sensory augmentation) or extracts higher-level features through a non-linear relation between incoming and outgoing signals (for intelligent sensory augmentation).

Low-level sensory information is a basic sensory pattern from the environment like light reflections or sound frequencies. High-level sensory information represents a more refined sensory pattern that captures particular sensory features. Take, for example, the representation of a human face. An image-to-sound SAD can either forward all the available sensory signals to the human user, including the lighting in the background, or extract the human face and only forward signals related to its essential features. In other words, instead of forwarding only low-level sensory patterns, an ISAD can filter out task-irrelevant sensory signals and extract higher-level features. The computation of higher-level features is achieved through a non-linear model of the sensory input.

The jump in the computational capacity between linear and non-linear models can be illustrated with the development of neural networks. The first instance of a neural network was a singular perceptron (McCulloch & Pitts, 1943). By either firing or not firing depending on the incoming signals, the perceptron can learn to separate the incoming signals into two classes, i.e. form a linear decision boundary. However, not all decision boundaries are linear or even can be approximated accurately as such. A basic XOR operation, where only one input signal can be true for the perceptron to fire, cannot be solved by a single perceptron. Instead, solving the XOR operation requires a non-linear model (Minsky & Papert, 1972). Combining single perceptrons into a multilayered computational model has enabled ANNs to surpass the limits of early perceptrons. Multilayered perceptrons are the basis for modern deep learning models. Multilayered perceptrons can model not only non-linear XOR operations but also other complex tasks like image classification (Tolstikhin et al., 2021), cancer diagnostics (Lorencin et al., 2020) or disease prediction (Car et al., 2020).

For the field of sensory augmentation, a similar developmental trajectory is possible. Traditional conversion algorithms, as used for many SSDs, have relied on matching incoming with outgoing sensory signal through a linear relation. The linear relation of incoming to outgoing signals entails that changes in the incoming signals are directly related to changes in the outgoing signals. Traditional SADs have achieved great success at the cost of cognitive overload and low usability by utilising crossmodal correspondences like matching

Table 3
Measuring up systems with the new criteria for sensory augmentation.

	vOICe (traditional SSD)	Tactile helmet (ISAD)
(i) sensible properties as input	Yes: camera	Yes: ultrasound sensors
(ii) additional, sensory output	Yes: auditory frequencies	Yes: tactile stimulations
(iii) perceptual function	Yes: sense of ‘vision’ through sounds	Yes: sense of ‘depth’ through vibration
(iv) sensory processing	Low-level: output signal is sensitive to all changes in input signal (linear mapping)	High-level: output signal is invariant to change in input signal (non-linear mapping)

auditory pitch with spatial orientation. Incorporating modern AI methods provides the opportunity of addressing these challenges while retaining a great success rate. By moving to non-linear computational methods, AI-driven SADs can extend the sensory processing to extract sensory features and ultimately provide a higher quality sensory signal to the human user.

With the transition from linear to non-linear signal transformation, the role of sensory pre-processing increases. Now the processing burden of ‘making sense’ of the incoming sensory signals is shared by the human user and the ISAD - and no longer by the human user in full. This shift signifies a possible extension of sensory processing of the human user and even grounds the conceptual notion of an AI extender (Hernández-Orallo & Vold, 2019) or forms of hybrid intelligence (Akata et al., 2020; Pescetelli, 2021). In fact, ISAD can construct high-level perceptual features based on the gathered sensory patterns through machine learning techniques without involving the human user. Hence, ISADs enable the implementation of an extended artificial sensor that outputs constructed high-level features in a sensory format. In the end, the human perceiver can obtain a direct sense of the constructed representation without constructing it in the ordinary sense by herself. The only construction task the perceiver has to do is to decode the forwarded signals from an available sensory modality, where the ISAD-information is received, into the ISAD-constructed representation.

For example, an image-to-sound ISAD can incorporate a wide range of sensory data, for instance, a depth-sensing LIDAR scanner or a thermal camera. After reducing the overall signal complexity with a neural network such as a variational autoencoder (Kingma & Welling, 2019) or a convolutional neural network (Albawi et al., 2017), the device can either forward compressed, low-level sensory signals or process them even further. Further processing can include the detection of human faces close to the human user or a mapping of recorded two-dimensional image data into a three-dimensional soundscape (Rumsey, 2012; Thuillier et al., 2018). A three-dimensional auditory soundscape introduces a feeling of spatiality into the sound environment without additional sensory signals. This technique can be applied for enhancing spatial awareness and, in combination with other sensory classification techniques, boost awareness of fast-moving peripheral objects such as cars on roads. When implementing AI methods into the classic vOICe architecture, the final auditory signal can now convey much richer and more accessible information like a three-dimensional sense of depth.

The main difference with ISA comes then if the change in mapping (from linear to non-linear) impacts how we can and should conceive of sensory augmentation. By pre-computing and only forwarding minimally noisy sensory signals to the human perceiver, which convey rich environmental cues, ISADs provide substantial improvements to the process of sensory substitution as well as sensory augmentation.

4. What does intelligent sensory augmentation really change?

4.1. The mere improvement claim

Two main challenges have accompanied classic SADs, particularly SSDs, alongside their wide success: cognitive overload and the dependency on extensive training. Both challenges have led to perceived low usability by users.

Cognitive overload arises when human users feel overwhelmed and stressed by the SSD-induced sensory input. This is partly because the subject receives the SSD-induced sensory signals alongside the ordinary, non-SSD-induced sensory signals. In other words, the human perceiver needs to exert cognitive effort to distinguish and interpret SSD-induced sensory information. The required cognitive effort is plausibly higher the more complex the SSD-induced sensory stimulation is. A complex sensory stimulation then facilitates cognitive overload (Elli et al., 2014).

Extensive training has shown to improve the human user’s ability to make sense of the additional sensory signals they receive through their device. The human user can learn to focus on only the relevant sensory signals and block out the accompanying irrelevant signals. A signal is considered relevant if it is closely connected to the task at hand. Hence, even if the users can successfully utilise the SSD-produced sensory information, they need to invest a sufficient amount of training before doing so.

ISA provides possible solutions to both challenges of classical SADs by reducing the overall amount of sensory information sent across the sensory modalities to the human user. ISA introduces machine learning algorithms that can filter sensory signals. Focusing on task-relevant instead of task-irrelevant sensory signals consequently allows for retaining all essential sensory information used for a particular perceptual function. Consider, for instance, the conversion from image-to-sound: instead of converting all possible visual information to auditory outputs, reducing the converted signals to a more manageable amount like objects within a five-meter radius yields a much more manageable final sensory output. In other words, eliminating unwanted sensory signals reduces the overall amount of sensory information that needs to be transferred across sensory modalities and the final lower signal complexity.

This approach of reducing signal complexity has been validated in a wide range of empirical studies on white background noise. In a clinical study on cochlear implants, Dawson et al. (2011) found that noise reduction algorithms successfully improve sentence perception in speech-weighted and dynamic background noise. Noise reduction and hence an improvement of the signal-to-noise ratio can significantly impact the overall perceptual performance of the human perceiver. Furthermore, van de Rijt et al. (2019) found that for speech-recognition tasks, audiovisual performance is significantly worse with low signal-to-noise ratios.

Conceptually, ISA reduces unnecessary sensory signals by converting sensory signals non-linearly from input to output. Machine learning methods such as convolutional neural networks or variational autoencoders have shown to be highly potent for signal compression and dimensionality reduction - key methods for signal filtering (Albawi et al., 2017; Fang et al., 2021; Kingma & Welling, 2019). These types of compression are highly efficient because they are non-structure preserving with a non-linear mapping of input and output data (for more see Kingma & Welling (2014); Kingma & Welling (2019); van Hoof et al. (2016); Wetzel (2017)). Implementing these methods after the sensory input and before the sensory output allows an ISAD to forward only the essential, i.e. task-relevant pieces of information, to the human user.

To distinguish between task-irrelevant and task-relevant signals, the ISAD depends on training. Determining what is task-relevant

other words, goal-independent. More recently, some attempts have started to utilise machine learning to segment and categorise 3D visual scenes (Caraiman et al., 2017; Morar et al., 2017): not only can the user choose the maximum number of objects to be encoded, but she can also decide the importance of the object in the final output. Each object is encoded as a weighted sum of its size, average depth, and deviation from where the viewer is looking direction, but the weights are established by the user. Thus, she can select whether more weight should be given to the biggest objects, or the closest ones, or to the objects closest to the direction where the user is looking. Ultimately, these weights could be learned through repeated use and become a flexible source of task-and situation-specific sensory information.

Representational information quality means that the quantity and accuracy of information are adjusted to serve their interpretability and ease of understanding (Wang & Poor, 1998). Classic SSDs are not indifferent to representational quality. The initial design adjusts the codes to already established sensory correspondences. For instance, in the vOICe, it is easier to interpret a high pitch as bright and a low pitch as dark than the reverse. However, besides the design stage, the representational quality is no longer taken into consideration. ISADs, in contrast, guarantee that only the essential information is retained in the reconstructed data.

An ISAD example where the shift in focus from quality over quantity becomes clear is the integration of a generative deep learning model into the signal conversion. Consider a speech-to-vision ISAD, which generates visual images of lip movements from corresponding audio-speech (L. Chen et al., 2018; Tian et al., 2019). The ISAD records incoming auditory frequencies, isolates the speech frequencies, matches the speech frequencies with a corresponding lip movement and then projects the lip movement onto a visual display. Here, the ISAD augments the mere auditory reality with additional, high-quality visual sensations, which would not have been possible to implement with classical sensory augmentation. This ISAD would hence be incredibly useful in scenarios where facial movements are obscured or additional sensory cues to understand speech are needed.

Another, this time more forward-looking form of ISADs are as diagnostic devices. AI-assistance for medical diagnoses can improve the accuracy and sensitivity of medical diagnoses, such as detecting tumours in mammograms (Rodríguez-Ruiz et al., 2018) or classifying liver tumours. Importantly, patients still widely prefer an analysis where both radiologists and AI are involved, compared to either of them alone (Dewey & Wilkens, 2019). Existing solutions, however, all choose to provide the AI-generated diagnosis uniquely in a symbolic format, for instance, as a probability statement about the type and location of a tumour. ISAD could, by contrast, provide its computational results in a sensory format. Hence, the human doctor gains access to an extensive diagnostic system through her senses and forms a medical judgment with all the available information.

5. Conclusion

Sensory substitution and extension devices have paved a way to restore sensory deficiencies through other, healthy senses. This has been achieved by recording sensory information with a sensory substitution device, transforming the signal into a different sensory modality and then forwarding the translated signal to the human perceiver. The standard limitation for these devices has been the high complexity of the outgoing sensory data leading to a high cognitive load and long training hours. This limitation can now be alleviated by artificial intelligence. In fact, implementing AI into sensory augmentation devices challenges the so far dominant assumption that the human can encode the final sensory pattern with similar precision to an encoding algorithm, where more information leads to higher encoding accuracy. However, as the slow adaptation outside of labs shows, without extensive training, no such results can be achieved (Elli et al., 2014; Lloyd-Esenkaya et al., 2020). The human has to learn to make sense of the, from the human perspective, complex and noisy sensory information.

In this paper, we have shown that introducing AI-based, non-linear mapping methods to the field of sensory augmentation leads to a new kind of sensory augmentation. This category of intelligent sensory augmentation challenges the assumption which dominates the field of sensory augmentation: that the perceptual function is better served by providing as much information as possible to the user.

Instead of this quantitative and linear approach, intelligent sensory augmentation introduces perceptual pre-processing of sensory information with the goal to provide a sensory output of the highest sensory quality - both taking into account the context and task - yielding a higher representational accuracy and ease of use. The difference in mapping (from linear to non-linear) points to a difference in principle (from quantity to quality) with suffices to make intelligent sensory augmentation more than 'mere improvements' of existing technologies.

Our argument is primarily based on what we call the input/output distinction of sensory augmentation: a way of looking at what makes a sub-category of sensory augmentation a separate category based on inputs, outputs, and ways of mapping them. As discussed, many debates in the literature take another perspective on individuating types of sensory augmentation through their assisted perceptual function: is it extension (by acquiring a new sense), substitution (by deferring a function to an existing sense) or restoration (by reinstating a lost sense)? Our approach here does not mean that the functional approach cannot be further applied to these novel intelligent sensory augmentation devices - but that it will depend on which devices are engineered and how they couple with human users - something which cannot be anticipated yet. In addition, we do think that the range of possibilities introduced by intelligent sensory augmentation devices will change the practical and theoretical landscape around sensory augmentation.

Practically, through their qualitative focus, they will change the way the debates about functions have gone so far: that is, centred around how to fit the 'added quantity' of information into one of the sensory modalities. Because of the lower load for the users, they can also eventually provide multiple kinds of information to them, for instance, both tactile and auditory, opening up questions regarding these new augmented multisensory experiences, which would not reduce to the 'which modality' questions.

Theoretically, the role of action in learning sensory substitution devices and sensory extensions has so far been used to promote enactive accounts of perception, whereas action may no longer play such a constitutive role in intelligent sensory augmentation.

In these respects, AI is clearly introducing a technological and conceptual shift in sensory augmentation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Louis Longin was funded by the Bavarian State Ministry of Science and the Arts coordinated by the Bavarian Research Institute for Digital Transformation (bidt).

Ophelia Deroy was supported by a grant from the Excellence Initiative in the LMU, a grant from the NOMIS foundation (acronym DISE), and a grant from the Volkswagen Stiftung (Co-Sense).

References

- Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., ... Welling, M. (2020). A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, 53(8), 18–28. <https://doi.org/10.1109/MC.2020.2996587>
- Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. *International Conference on Engineering and Technology (ICET)*, 2017, 1–6. <https://doi.org/10.1109/ICEngTechnol.2017.8308186>
- Amedi, A., Stern, W. M., Campodon, J. A., Bermphohl, F., Merabet, L., Rotman, S., Hemond, C., Meijer, P., & Pascual-Leone, A. (2007). Shape conveyed by visual-to-auditory sensory substitution activates the lateral occipital complex. *Nature Neuroscience*, 10(6), 687–689. <https://doi.org/10.1038/nrn1912>
- Arnold, G., & Auvray, M. (2018). Tactile recognition of visual stimuli: Specificity versus generalization of perceptual learning. *Vision Research*, 152(November), 40–50. <https://doi.org/10.1016/j.visres.2017.11.007>
- Auvray, M., & Farina, M. (2017). Patrolling the boundaries of synesthesia: A critical appraisal of transient and artificially induced forms of synesthetic experiences. In *Sensory blending: On synesthesia and related phenomena* (pp. 248–274). Oxford University Press.
- Auvray, M., Hanneton, S., Lenay, C., & O'Regan, K. (2005). There is something out there: Distal attribution in sensory substitution, twenty years later. *Journal of Integrative Neuroscience*, 4(4), 505–521. <https://doi.org/10.1142/S0219635205001002>
- Auvray, M., Hanneton, S., & O'Regan, J. K. (2007). Learning to Perceive with a Visuo Auditory Substitution System: Localisation and Object Recognition with "The Voice". *Perception*, 36(3), 416–430. <https://doi.org/10.1088/p5631>
- Auvray, M., & Harris, L. R. (2014). The state of the art of sensory substitution. *Multisensory Research*, 27(5–6), 265–269. <https://doi.org/10.1163/22134808-00002464>
- Auvray, M., & Myin, E. (2009). Perception with compensatory devices: From sensory substitution to sensorimotor extension. *Cognitive Science*, 33(6), 1036–1058. <https://doi.org/10.1111/j.1551-6709.2009.01040.x>
- Bach-Y-Rita, P., Collins, C. C., Saunders, F. A., White, B., & Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, 221(5184), 963–964. <https://doi.org/10.1038/221963a0>
- Bach-y-Rita, P., & Kercel, W. S. (2003). Sensory substitution and the human-machine interface. *Trends in Cognitive Sciences*, 7(12), 541–546. <https://doi.org/10.1016/j.tics.2003.10.013>
- Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. *Perception & Psychophysics*, 57(8), 1151–1162. <https://doi.org/10.3758/BF03208371>
- Block, N. (2003). Tactile sensation via spatial perception. *Trends in Cognitive Sciences*, 7(7), 285–286. [https://doi.org/10.1016/S1364-6613\(03\)00132-3](https://doi.org/10.1016/S1364-6613(03)00132-3)
- Brown, D., Macpherson, T., & Ward, J. (2011). Seeing with sound? Exploring different characteristics of a visual-to-auditory sensory substitution device. *Perception*, 40(9), 1120–1135. <https://doi.org/10.1088/p6952>
- Canbek, N. G., & Mutlu, M. E. (2016). On the track of Artificial Intelligence: Learning with Intelligent Personal Assistants. *Journal of Human Sciences*, 13(1), 592–601.
- Car, Z., Baressi Šegota, S., Anelić, N., Lorenčin, I., & Mrzljak, V. (2020). Modeling the Spread of COVID-19 Infection Using a Multilayer Perceptron. *Computational and Mathematical Methods in Medicine*, 2020, 1–10. <https://doi.org/10.1155/2020/5714714>
- Caraiman, S., Morar, A., Owczarek, M., Burlacu, A., Rzeszotarski, D., Botezatu, N., Hergheliegiu, P., Moldoveanu, F., Strumillo, P., & Moldoveanu, A. (2017). Computer Vision for the Visually Impaired: The Sound of Vision System. *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 1480–1489.
- Chebat, D.-R., Maidenbaum, S., & Amedi, A. (2015). Navigation Using Sensory Substitution in Real and Virtual Mazes. *PLOS ONE*, 10(6), Article e0126307. <https://doi.org/10.1371/journal.pone.0126307>
- Chen, L., Li, Z., Maddox, R. K., Duan, Z., & Xu, C. (2018). Lip Movements Generation at a Glance. In V. Ferrari, M. Hebert, C. Sminchisescu, & Y. Weiss (Eds.), *Computer Vision ECCV 2018* (pp. 538–553). Springer International Publishing.
- Chen, M., Ma, Y., Song, J., Lai, C.-F., & Hu, B. (2016). Smart Clothing: Connecting Human with Clouds and Big Data for Sustainable Health Monitoring. *Mobile Networks and Applications*, 21(5), 825–845. <https://doi.org/10.1007/s11036-016-0745-1>
- Chollet, F. (2019). On the Measure of Intelligence. arXiv:1911.01547 [Cs]. <https://arxiv.org/abs/1911.01547>.
- Cohen, L. G., Celnik, P., Pascual-Leone, A., Corwell, B., Faiz, L., Dambrorsia, J., Honda, M., Sadato, N., Gerloff, C., Catala, M. D., & Hallett, M. (1997). Functional relevance of cross-modal plasticity in blind humans. *Nature*, 389(6647), 180–183. <https://doi.org/10.1038/38278>
- Collignon, O., Champoux, F., Voss, P., & Lepore, F. (2011). Sensory rehabilitation in the plastic brain. In *Progress in brain research* (Vol. 191, , 211–231. <https://doi.org/10.1016/B978-0-444-53752-2.00003-5>
- Collignon, O., Voss, P., Lassonde, M., & Lepore, F. (2008). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. *Experimental Brain Research*, 192(3), 343. <https://doi.org/10.1007/s00221-008-1553-z>
- Dagnelie, G. (2012). Retinal implants: Emergence of a multidisciplinary field. *Current Opinion in Neurology*, 25(1), 67–75. <https://doi.org/10.1097/WCO.0b013e32834f02c3>
- Dawson, P. W., Mauger, S. J., & Hersbach, A. A. (2011). Clinical Evaluation of Signal-to-Noise Ratio in Nucleus Cochlear Implant Recipients. *Ear and Hearing*, 32(3), 382–390. <https://doi.org/10.1097/AUD.0b013e318201c200>
- Deroy, O., & Auvray, M. (2012). Reading the world through the skin and ears: A new perspective on sensory substitution. *Frontiers in Psychology*, 3(NOV). <https://doi.org/10.3389/fpsyg.2012.00457>
- Deroy, O., & Auvray, M. (2014). A crossmodal perspective on sensory substitution. In *Perception and its modalities* (pp. 327–349). Doi: 10.1093/acprof:oso/9780199832798.003.0014.
- Deroy, O., & Spence, C. (2013). Training, hypnosis, and drugs: Artificial synesthesia, or artificial paradises? *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00660>
- Dewey, M., & Wilkens, U. (2019). The Bionic Radiologist : Avoiding blurry pictures and providing greater insights. *Npj Digital Medicine*, 2(1), 1–7. <https://doi.org/10.1038/s41746-019-0142-9>
- Dilworth, J. (2010). Realistic virtual reality and perception. *Philosophical Psychology*, 23(1), 23–42. <https://doi.org/10.1080/09515080903533942>
- Egmont-Petersen, M., de Ridder, D., & Handels, H. (2002). Image processing with neural networks a review. *Pattern Recognition*, 35(10), 2279–2301. [https://doi.org/10.1016/S0031-3203\(01\)00178-9](https://doi.org/10.1016/S0031-3203(01)00178-9)

- Elli, G. V., Benetti, S., & Collignon, O. (2014). Is there a future for sensory substitution outside academic laboratories? *Multisensory Research*, 27(5–6), 271–291. <https://doi.org/10.1163/22134808-00002460>
- Eppe, M., Kerzel, M., Strahl, E., & Wermter, S. (2018). Deep Neural Object Analysis by Interactive Auditory Exploration with a Humanoid Robot. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 284–289. Doi: 10.1109/IROS.2018.8593838.
- Fang, H., Carbalal, G., Wermter, S., & Gerkmann, T. (2021). Variational Autoencoder for Speech Enhancement with a Noise-Aware Encoder. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 676–680. Doi: 10.1109/ICASSP39728.2021.9414060.
- Farina, M. (2013). Neither touch nor vision: Sensory substitution as artificial synesthesia? *Biology & Philosophy*, 28(4), 639–655. <https://doi.org/10.1007/s10539-013-9377-z>
- Fernández-Caramés, T., & Fraga-Lamas, P. (2018). Towards the internet-of-smart-clothing: A review on IoT wearables and garments for creating intelligent connected E-textiles. *Electronics*, 7(12), 405. <https://doi.org/10.3390/electronics7120405>
- Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. *Synthesis Lectures on Human Language Technologies*, 10(1), 1–309. <https://doi.org/10.2200/S00762ED1V01Y201703HLT037>
- Goldberg, Y. (2016). A Primer on Neural Network Models for Natural Language Processing. *Journal of Artificial Intelligence Research*, 57, 345–420. <https://doi.org/10.1613/jair.4992>
- Golub, J. S., Ling, L., Nie, K., Nowack, A., Shepherd, S. J., Bierer, S. M., Jameyson, E., Kaneko, C. R. S., Phillips, J. O., & Rubinstein, J. T. (2014). Prosthetic implantation of the human vestibular system. *Otology & Neurotology: Official Publication of the American Otological Society, American Neurotology Society [and] European Academy of Otology and Neurotology*, 35(1), 136–147. Doi: 10.1097/MAO.00000000000000000003.
- Haalboom, M., Gerritsen, J. W., & van der Palen, J. (2019). Differentiation between infected and non-infected wounds using an electronic nose. *Clinical Microbiology and Infection*, 25(10), 1288.e1–1288.e6. <https://doi.org/10.1016/j.cmi.2019.03.018>
- Hameed, J., Harrison, I., Gasson, M. N., & Warwick, K. (2010). A novel human-machine interface using subdermal magnetic implants. In *2010 IEEE 9th International Conference on Cybernetic Intelligent Systems, CIS 2010*. Doi: 10.1109/UKRICIS.2010.5898141.
- Hanneton, S., Auvray, M., & Durette, B. (2010). The Vibe: A versatile vision-to-audition sensory substitution device. *Applied Bionics and Biomechanics*, 7(4), 269–276. <https://doi.org/10.1080/11762322.2010.512734>
- Haugeland, J. (1985). *Artificial intelligence: The very idea*. MIT Press.
- Haupt, R. L., & Haupt, S. E. (2003). *Practical Genetic Algorithms*. John Wiley & Sons, Inc. Doi: 10.1002/0471671746.
- Hernández-Orallo, J., & Vold, K. (2019). AI Extenders: The Ethical and Societal Implications of Humans Cognitively Extended by AI. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 507–513. Doi: 10.1145/3306618.3314238.
- Hoy, M. B. (2018). Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants. *Medical Reference Services Quarterly*, 37(1), 81–88. <https://doi.org/10.1080/02763869.2018.1404391>
- Hramov, A. E., Frolov, N. S., Maksimenko, V. A., Makarov, V. V., Koronovskii, A. A., Garcia-Prieto, J., Antón-Toro, L. F., Maestú, F., & Pisarchik, A. N. (2018). Artificial neural network detects human uncertainty. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(3), Article 033607. <https://doi.org/10.1063/1.5002892>
- Hu, D., Wang, D., Li, X., Nie, F., & Wang, Q. (2019). Listen to the Image. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 7972–7981.
- Hu, W., Wan, L., Jian, Y., Ren, C., Jin, K., Su, X., Bai, X., Haick, H., Yao, M., & Wu, W. (2019). Electronic Noses: From Advanced Materials to Sensors Aided with Data Processing. *Advanced Materials Technologies*, 4(2), 1800488. <https://doi.org/10.1002/admt.201800488>
- Hurley, S., & Noë, A. (2003). Neural Plasticity and Consciousness. *Biology and Philosophy*, 18(1), 131–168. <https://doi.org/10.1023/A:1023308401356>
- Illari, P., & Floridi, L. (2014). Information Quality, Data and Philosophy. In L. Floridi, & P. Illari (Eds.), *The Philosophy of Information Quality* (pp. 5–23). Springer International Publishing. https://doi.org/10.1007/978-3-319-07121-3_2
- Kärcher, S. M., Fenzlaff, S., Hartmann, D., Nagel, S. K., & König, P. (2012). Sensory augmentation for the blind. *Frontiers in Human Neuroscience, MARCH 2012*, 1–15. Doi: 10.3389/fnhum.2012.00037.
- Kerdegari, H., Kim, Y., & Prescott, T. J. (2016). Head-Mounted Sensory Augmentation Device: Comparing Haptic and Audio Modality. In N. F. Lepora, A. Mura, M. Mangan, P. F. M. J. Verschure, M. Desmulliez, & T. J. Prescott (Eds.), *Biomimetic and Biohybrid Systems* (pp. 107–118). Springer International Publishing.
- Kim, M., Park, Y., Moon, K., & Jeong, C. Y. (2021). Analysis and Validation of Cross-Modal Generative Adversarial Network for Sensory Substitution. *International Journal of Environmental Research and Public Health*, 18(12), 6216. <https://doi.org/10.3390/ijerph18126216>
- Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. *arXiv:1312.6114 [Cs, Stat]*. <https://arxiv.org/abs/1312.6114>.
- Kingma, D. P., & Welling, M. (2019). An Introduction to Variational Autoencoders. *Foundations and Trends in Machine Learning*, 12(4), 307–392. <https://doi.org/10.1561/2200000056>
- Kiverstein, J., & Farina, M. (2012). Do sensory substitution devices extend the conscious mind? In F. Paglieri (Ed.), *Advances in Consciousness Research* (pp. 19–40). John Benjamins Publishing Company. <https://doi.org/10.1075/aicr.86.02kiv>
- Kiverstein, J., Farina, M., & Clark, A. (2015). Substituting the senses. In *The Oxford handbook of philosophy of perception* (pp. 659–675). Oxford University Press.
- Lee, H., Upright, C., Eliuk, S., & Kobsa, A. (2016). Personalized object recognition for augmenting human memory. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*, 1054–1061. Doi: 10.1145/2968219.2968568.
- Legg, S., & Hutter, M. (2007). A Collection of Definitions of Intelligence. *arXiv:0706.3639 [Cs]*. <https://arxiv.org/abs/0706.3639>
- Lloyd-Esenkaya, T., Lloyd-Esenkaya, V., O'Neill, E., & Proulx, M. J. (2020). Multisensory inclusive design with sensory substitution. *Cognitive Research: Principles and Implications*, 5(1), 37. <https://doi.org/10.1186/s41235-020-00240-7>
- Lorençin, I., Anelić, N., Španjol, J., & Car, Z. (2020). Using multi-layer perceptron with Laplacian edge detector for bladder cancer diagnosis. *Artificial Intelligence in Medicine*, 102, Article 101746. <https://doi.org/10.1016/j.artmed.2019.101746>
- Maidenbaum, S., Abboud, S., & Amedi, A. (2014). Sensory substitution: Closing the gap between basic research and widespread practical visual rehabilitation. *Neuroscience & Biobehavioral Reviews*, 41, 3–15. <https://doi.org/10.1016/j.neubiorev.2013.11.007>
- Marks, L. E., Hammele, R. J., Bornstein, M. H., & Smith, L. B. (1987). Perceiving Similarity and Comprehending Metaphor. *Monographs of the Society for Research in Child Development*, 52(1), i. <https://doi.org/10.2307/1166084>
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115–133. <https://doi.org/10.1007/BF02478259>
- Meijer, P. B. L. (1992). An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, 39(2), 112–121. <https://doi.org/10.1109/10.121642>
- Minsky, M., & Papert, S. A. (1972). *Perceptrons: An introduction to computational geometry* (2. print. with corr). The MIT Press.
- Morar, A., Moldoveanu, F., Petrescu, L., & Moldoveanu, A. (2017). Real Time Indoor 3D Pipeline for an Advanced Sensory Substitution Device. In S. Battiatto, G. Gallo, R. Schettini, & F. Stanco (Eds.), *Image Analysis and Processing - ICIAP 2017* (pp. 685–695). Springer International Publishing. Doi: 10.1007/978-3-319-68548-9_62.
- Murray, C. D. (2008). Embodiment and Prosthetics. In P. Gallagher, D. Desmond, & M. MacLachlan (Eds.), *Psychoprosthetics* (pp. 119–129). Springer. Doi: 10.1007/978-1-84628-980-4_9.
- Nagel, S. K., Carl, C., Krings, T., Märtin, R., & König, P. (2005). Beyond sensory substitution learning the sixth sense. *Journal of Neural Engineering*, 2(4), R13–R26. <https://doi.org/10.1088/1741-2560/2/4/R02>
- Nanay, B. (2017). Sensory substitution and multimodal mental imagery. *Perception*, 46(9), 1014–1026. <https://doi.org/10.1177/0301006617699225>
- Neugebauer, A., Rifai, K., Getzlaff, M., & Wahl, S. (2020). Navigation aid for blind persons by visual-to-auditory sensory substitution: A pilot study. *PLOS ONE*, 15(8), Article e0237344. <https://doi.org/10.1371/journal.pone.0237344>
- Pescetelli, N. (2021). A Brief Taxonomy of Hybrid Intelligence. *Forecasting*, 3(3), 633–643. <https://doi.org/10.3390/forecast3030039>
- Pinheiro Lima Neto, E., Martins da Costa, R., Silva Alves Fernandes, D., & Alphonsus Alves de Melo Nunes Soares, F. (2019). Sensory Substitution of Vision: A Systematic Mapping and a Deep Learning Object Detection Proposition. *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI)*, 1815–1819. Doi: 10.1109/ICTAI.2019.00274.
- Proulx, M. J. (2010). Synthetic synesthesia and sensory substitution. *Consciousness and Cognition*, 19(1), 501–503. <https://doi.org/10.1016/j.concog.2009.12.005>

- Proulx, M. J., Stoerig, P., Ludowig, E., & Knoll, I. (2008). Seeing "Where" through the Ears: Effects of Learning-by-Doing and Long-Term Sensory Deprivation on Localization Based on Image-to-Sound Substitution. *PLoS ONE*, 3(3), Article e1840. <https://doi.org/10.1371/journal.pone.0001840>
- Ptito, M., Iversen, K., Auvray, M., Deroy, O., & Kupers, R. (2018). Limits of the Classical Functionalist Perspective on Sensory Substitution. In *Sensory Substitution and Augmentation* (First). Oxford University Press.
- Pulikkasari, C., & Lam, S. (2019). Laser eyes for driverless cars: The road to automotive LIDAR. *2019 Optical Fiber Communications Conference and Exhibition, OFC 2019 - Proceedings*. Doi: 10.1364/ofc.2019.tu3d.2.
- Reynolds, Z., & Glenney, B. (2012). When Sensory Substitution Devices Strike Back: An Interactive Training Paradigm. *Philosophy Study*, 2(6), 432–438. <https://doi.org/10.17265/2159-5313>
- Rodríguez-Ruiz, A., Krupinski, E., Mordang, J.-J., Schilling, K., Heywang-Köbrunner, S. H., Sechopoulos, I., & Mann, R. M. (2018). Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System. *Radiology*, 290(2), 305–314. <https://doi.org/10.1148/radiol.2018181371>
- Rumsey, F. (2012). Spatial Audio (Zeroth). *Routledge*. <https://doi.org/10.4324/9780080498195>
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Malaysia: Pearson Education Limited.
- Saganowski, S., Kazienko, P., Dzieżyc, M., Jakimów, P., Komoszyńska, J., Michalska, W., Dutkowiak, A., Polak, A., Dziadek, A., & Ujma, M. (2020). Review of Consumer Wearables in Emotion, Stress, Meditation, Sleep, and Activity Detection and Analysis. *arXiv:2005.00093 [Cs]*. <https://arxiv.org/abs/2005.00093>.
- Smart, P. (2017). Extended cognition and the internet: A review of current issues and controversies. *Philosophy and Technology*, 30(3), 357–390. <https://doi.org/10.1007/s13347-016-0250-2>
- Striem-Amit, E., Guendelman, M., & Amedi, A. (2012). "Visual" Acuity of the Congenitally Blind Using Visual-to-Auditory Sensory Substitution. *PLOS ONE*, 7(3), Article e33136. <https://doi.org/10.1371/journal.pone.0033136>
- Stronks, H. C., Nau, A. C., Ibbotson, M. R., & Barnes, N. (2015). The role of visual deprivation and experience on the performance of sensory substitution devices. *Brain Research*, 1624, 140–152. <https://doi.org/10.1016/j.brainres.2015.06.033>
- Sun, W., Liu, J., & Zhang, H. (2017). When Smart Wearables Meet Intelligent Vehicles: Challenges and Future Directions. *IEEE Wireless Communications*, 24(3), 58–65. <https://doi.org/10.1109/MWC.2017.1600423>
- Tajadura-Jiménez, A., Välijamäe, A., & Kuusk, K. (2020). Altering One's Body-Perception Through E-Textiles and Haptic Metaphors. *Frontiers in Robotics and AI*, 7. <https://doi.org/10.3389/frobt.2020.00007>
- Thuijlier, E., Camper, H., & Tashev, I. J. (2018). Spatial Audio Feature Discovery with Convolutional Neural Networks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 6797–6801. Doi: 10.1109/ICASSP.2018.8462315.
- Tian, G., Yuan, Y., & Liu, Y. (2019). Audio2Face: Generating Speech/Face Animation from Single Audio with Attention-Based Bidirectional LSTM Networks. *2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 366–371. Doi: 10.1109/ICMEW.2019.00069.
- Tolstikhin, I., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., Lucic, M., & Dosovitskiy, A. (2021). MLP-Mixer: An all-MLP Architecture for Vision. *arXiv:2105.01601 [Cs]*. <https://arxiv.org/abs/2105.01601>.
- Tyler, M., Danilov, Y., & Bach-Y-Rita, P. (2003). Closing an open-loop control system: Vestibular substitution through the tongue. *Journal of Integrative Neuroscience*, 2 (2), 159–164. <https://doi.org/10.1142/S0219635203000263>
- van de Rijt, L. P. H., Roye, A., Mylanus, E. A. M., van Opstal, A. J., & van Wanrooij, M. M. (2019). The principle of inverse effectiveness in audiovisual speech perception. *Frontiers in Human Neuroscience*, 13. <https://doi.org/10.3389/fnhum.2019.00335>
- van Hoof, H., Chen, N., Karl, M., van der Smagt, P., & Peters, J. (2016). Stable reinforcement learning with autoencoders for tactile and visual data. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, 3928–3934. <https://doi.org/10.1109/IROS.2016.7759578>
- Wang, X., & Poor, H. V. (1998). Blind multiuser detection: A subspace approach. *IEEE Transactions on Information Theory*, 44(2), 677–690. <https://doi.org/10.1109/18.661512>
- Ward, J., & Meijer, P. (2010). Visual experiences in the blind induced by an auditory sensory substitution device. *Consciousness and Cognition*, 19(1), 492–500. <https://doi.org/10.1016/j.concog.2009.10.006>
- Wetzel, S. J. (2017). Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. *Physical Review E*, 96(2), Article 022140. <https://doi.org/10.1103/PhysRevE.96.022140>
- Wheeler, M. (2015). Not what it's like but where it's like: Phenomenal consciousness, sensory substitution, and the extended mind. *Journal of Consciousness Studies*, 22 (3–4), 129–147.
- Wright, T., & Ward, J. (2013). The evolution of a visual-to-auditory sensory substitution device using interactive genetic algorithms. *Quarterly Journal of Experimental Psychology*, 66(8), 1620–1638. <https://doi.org/10.1080/17470218.2012.754911>
- Wright, T., & Ward, J. (2018). Sensory Substitution Devices as Advanced Sensory Tools. In *Sensory Substitution and Augmentation* (Vol. 219, pp. 188–204). The British Academy. <https://doi.org/10.5871/bacard/9780197266441.003.0012>
- Zeng, F. G., Rebscher, S., Harrison, W., Sun, X., & Feng, H. (2008). Cochlear implants: System design, integration, and evaluation. *IEEE Reviews in Biomedical Engineering*. <https://doi.org/10.1109/RBME.2008.2008250>
- Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys*, 52(1), 1–38. <https://doi.org/10.1145/3285029>