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:
‘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 perceive’s body and the object seen in VR are not causally related to the actual spatial relations between the perceive’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 perceive, is clearly in one’s immediate environment as the perceive 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 perceive 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 the
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 spatial 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 & Noe, 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); Kocher 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 perceivable 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

| 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 SSD sufficiently, the overall usability is perceived as low. The human user constantly needs to exert cognitive effort to distinguish and interpret SSD-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 ‘Polylot’, 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 ‘Polylot’ 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. Not every change in the incoming signal leads to a change in the outgoing signal. Instead, the neural network learns to reconstruct the incoming signal from the output. 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 cable 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...
depending on the incoming signals, the perceptron can learn to separate the incoming signals into two classes, i.e. form a linear de-
level features. The computation of higher-level features is achieved through a non-linear model of the sensory input.
The representation of a human face. An image-to-sound SAD can either forward all the available sensory signals to the human user,
level sensory information represents a more refined sensory pattern that captures particular sensory features. Take, for example,
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 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,
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 de-
cision 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 compu-
tational 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 utilizing 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, audiospatial 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
can then be achieved in two main ways. For a supervised learning architecture, the ISAD learns to filter out task-relevant signals given a set of examples. Convolutional neural networks can learn to perform robust object classification given a training set. For an unsupervised learning architecture, the ISAD learns to classify signals by itself. This is the case for variational autoencoders. An autoencoder learns which piece of information is essential by first compressing and reconstructing it. If the signal does not aid the reconstruction of the incoming data, it is deemed not essential. In this sense, it is possible to create ISAD with an externally defined filtering purpose like extracting only human faces or with an internally defined filtering purpose by dropping information useless for reconstruction. A sound-to-tactile ISAD, for instance, might only be sensitive to speech-related frequencies and allows for a much cleaner sensory conversion and output; or an image-to-sound ISAD might only extract edges and shapes from visual input, which facilitates the construction of much simpler auditory signals designed for object recognition.

Thus, instead of instructing SSD users to focus on predetermined useful information and simplified structures, ISADs can implement self-learning algorithms to determine independently the essential pieces of sensory information that need to be retained. The advantage of ISAD become clear: by reducing the complexity of the forwarded sensory data, the user has to do less work for figuring out their underlying pattern, i.e. ‘meaning’. Instead of a linear mapping of all incoming sensory information, ISADs can transform and forward only the most relevant sensory information for the user to make sense of the stimulation. This approach makes it possible to reduce the cognitive load for human users and reduce training requirements.

4.2. The qualitative difference claim

As discussed above, by utilising a non-linear transfer of sensory information within the coupling system, ISADs have the potential to improve the performance of sensory substitution and extension devices. However, this ‘mere improvement’ arguably underestimates the difference ISA introduces to sensory augmentation. In fact, ISA challenges the underlying principle on which sensory augmentation has rested so far. While traditional, non-intelligent sensory augmentation approaches seek to improve or extend the quantity of sensory information provided to the user, ISA instead seeks to improve the quality of the provided sensory signal.

Under the traditional framework of non-intelligent sensory augmentation, the main method for improving the usability of SSDs has been to rely on cross-modal features of perception (Auvray et al., 2005). This includes, for example, the connection between pitch and vertical positioning (Ben-Artzi & Marks, 1995) and between loudness and luminance (Marks et al., 1987). However, what has remained is the overall dependency on the human user making sense of the underlying sensory patterns. The cross-modal pattern has shown particularly useful because it represents easily identifiable patterns across the respective sensory modalities. If spatial location and luminance are encoded through an image-to-sound SSD, for instance, the final output traditionally varies in pitch and loudness. Adding additional input features requires adding another output parameter to a classical SAD. After adding multiple output parameters, the final conflated output signal becomes highly complex and difficult to encode.

In contrast, 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.

By connecting input and output through non-linear computational models, an ISAD can extract higher-level features from the incoming signals. These higher-level features are higher-level because they possess a higher information quality than the low-level sensory signals taken in and produced by classical SADs. For an intuitive example, consider a vision-to-tactile SAD. Under the classical framework, this SAD relays only low-level sensory information to the human user. The SAD converts images to tactile stimulations while relying on established cross-modal features of perception to aid the decoding of the information. An ISAD, in contrast, manipulates the incoming visual signal either by filtering out background noise or by extracting higher-level features like object or feature classification. A high-quality tactile output signal might be sensitive only to contextual, relevant information like potential hazards in traffic, targets during sports practice or human faces in crowds.

The notion of information quality is multi-dimensional, as pointed out by philosophers and computer scientists (Illari & Floridi, 2014; Wang & Poor, 1998) and basically covers all dimensions of information that are not simply captured by looking at the quantity and accuracy of information. A good example here is the believability of information: two sets of data can be equal in size and accuracy, but users may find one more believable than the other because of trust and reputation. Another qualitative difference comes from accessibility. However, such qualitative dimensions are not the most relevant for ISA, but they point out the difference between the goals of quantitative and qualitative approaches to information. Quantitative approaches are meant to improve the relation between encoded set and decoded set, or world and data: how can information be best encoded and transmitted to be accurately decoded. Qualitative approaches consider that this relation between the world and data is relative to the users’ purposes and constraints. Not all accuracy is equal, as other considerations feed in, and sometimes less accuracy can be better.

Connecting the notion of information quality to the distinction of low-level and higher-level sensory signals as outlined above, it generally holds that higher-level sensory signals possess a higher information quality than low-level sensory signals. In fact, the higher information quality makes them higher-level compared to the low-level sensory signals. The two qualitative considerations relevant to ISADs fall in the categories of contextual information quality and, more importantly, of representational information quality.

**Contextual information quality** means that the quantity and accuracy of information provided depend on the context of use to be relevant and timely (Wang & Poor, 1998). Classic approaches to sensory substitution and extensions usually provide the same amount of information across contexts. They have not considered contextual quality: the quantity of information provided to the user is, in
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 interpret-ability 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 pathed 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-Enskay 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 (GoSense).

References

Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Elben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oltehoek, 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, competitive, 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., Bermpohl, 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/nn1912

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 synaesthesia: A critical appraisal of transient and artificially induced forms of synaesthesic experiences. In Sensory blending: On synaesthesia and related phenomena (pp. 248–274). Oxford University Press.

Auvray, M., Hanneton, S., Lenay, C., & O’Regan, K. (2011). Clinical Evaluation of Signal-to-Noise Ratio in Nucleus Cochlear Implant Recipients. Digital Transformation (bidt).

Brown, D., Macpherson, T., & Ward, J. (2011). Seeing with sound? Exploring different characteristics of a visual-to-auditory sensory substitution device. Frontiers in Psychology, 3(4), 505–521. https://doi.org/10.3389/fpsyg.2012.00505

Ben-Artzi, E., & Marks, L. E. (1995). Visual-auditory interaction in speeded classification: Role of stimulus difference. Perception, 24(3), 416–430. https://doi.org/10.1068/p9361

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.1551-6709.2009.01046.x

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)00013-2

Brown, D., Macpherson, T., & Ward, J. (2011). Seeing with sound? Exploring different characteristics of a visual-to-auditory sensory substitution device. Perception, 40(11), 10–11.

Canbek, N. G., & Mutlu, M. E. (2016). On the track of Artificial Intelligence: Learning with Intelligent Personal Assistants. Frontiers in Psychology, 7, 553). Doi: 10.3389/fpsyg.2016.00569

Cohen, L. G., Celnik, P., Pascual-Leone, A., Corwell, B., Faiz, L., Dambrosia, J., Honda, M., Sadato, N., Gerloff, C., Catala, M. D., & Hallett, M. (1997). Functional relevance of cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. Brain, 120(9), 1595–1606. https://doi.org/10.1093/brain/120.9.1595

Collignon, D., Champoux, F., Voss, P., & Lepore, F. (2018). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. Frontiers in Psychology, 9, 689. https://doi.org/10.3389/fpsyg.2018.00689

Collignon, D., Champoux, F., Voss, P., & Lepore, F. (2018). Sensory rehabilitation in the plastic brain. In Progress in brain research (Vol. 231, pp. 211–231). https://doi.org/10.1016/bs.pbr.2017.09.020

Collignon, D., Champoux, F., & Voss, P. (2018). Cross-modal plasticity for the spatial processing of sounds in visually deprived subjects. Experimental Brain Research, 235(6), 1475–1486. https://doi.org/10.1007/s00221-018-5135-3

Colnaghi, 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), 343–355. https://doi.org/10.1007/s00221-008-1553-5

Dawson, P. W., Mauger, S. J., & Hersbach, 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

Dawson, P. W., Mauger, S. J., & Hersbach, 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, 596. 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 synaesthesia, or artificial paradises? Frontiers in Psychology, 4, 742. https://doi.org/10.3389/fpsyg.2013.00660

Dewey, M., & Wilk, 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 networksa review. Pattern Recognition, 35(10), 2279–2301. https://doi.org/10.1016/S0031-3203(01)00178-9
Pinheiro Lima Neto, E., Martins da Costa, R., Silva Alves Fernandes, D., Pescetelli, N. (2021). A Brief Taxonomy of Hybrid Intelligence.

Neugebauer, A., Rifai, K., Getzlaff, M., & Wahl, S. (2020). Navigation aid for blind persons by visual-to-auditory sensory substitution: A pilot study.

Nagel, S. K., Carl, C., Kringe, T., Eppe, M., Kerzel, M., Strahl, E., Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing.

Hameed, J., Harrison, I., Gasson, M. N., Lorencin, I., Anell, S. M., Fenzlaff, S., Hartmann, D., Nagel, S. K., & Papiach, S. M. (2019). Differentiation between infected and non-infected wounds using an electronic nose.

Kutana, A. (2016). Practical Genetic Algorithms. John Wiley & Sons, Inc. Doi: 10.1002/0471671746.

Körsch, 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. Doi: 10.3390/ijerph18126216.

Kingma, D. P., & Welling, M. (2014). Auto-Encoding Variational Bayes. arxiv:1312.6114 [Cs, Stat]. Doi: 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. Doi: 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. Doi: 10.1017/9781316626378.

Kiverstein, J., Farina, M., & Clark, A. (2015). Substituting the senses. In The Oxford handbook of philosophy of perception (pp. 659–675). Oxford University Press.

Kinsella, V. J., & Peck, V. (2012). Sensory substitution devices: A comparison of their effectiveness. In Proceedings of the 2012 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. 1045–1046. Doi: 10.1145/2196821.2196858.

Keller, S., & Hutter, M. (2007). A Collection of Definitions of Intelligence. A Collection of Definitions of Intelligence.

Koch, C., & Karduck, E. (2016). A neural network detects human uncertainty. In Proceedings of the 2016 IEEE International Conference on Cybernetic Intelligent Systems, CIS 2016. Doi: 10.1109/CIS.2016.7762272.

Koch, C., & Karduck, E. (2016). A neural network detects human uncertainty. In Proceedings of the 2016 IEEE International Conference on Cybernetic Intelligent Systems, CIS 2016. Doi: 10.1109/CIS.2016.7762272.

Koch, C., & Karduck, E. (2016). A neural network detects human uncertainty. In Proceedings of the 2016 IEEE International Conference on Cybernetic Intelligent Systems, CIS 2016. Doi: 10.1109/CIS.2016.7762272.

Koch, C., & Karduck, E. (2016). A neural network detects human uncertainty. In Proceedings of the 2016 IEEE International Conference on Cybernetic Intelligent Systems, CIS 2016. Doi: 10.1109/CIS.2016.7762272.

Koch, C., & Karduck, E. (2016). A neural network detects human uncertainty. In Proceedings of the 2016 IEEE International Conference on Cybernetic Intelligent Systems, CIS 2016. Doi: 10.1109/CIS.2016.7762272.
