The fundamental frequency: Extending sound perception theory to extended-reality collaborative environments

GARNER, Tom A. <http://orcid.org/0000-0002-2432-0485>

Computers & Education: X Reality 5 (2024) 100080

Contents lists available at ScienceDirect

Computers & Education: X Reality

journal homepage: www.journals.elsevier.com/computers-and-education-x-reality

The fundamental frequency: Extending sound perception theory to extended-reality collaborative environments

Tom A. Garner

Sheffield Hallam University. Howard St, Sheffield City Centre, Sheffield, S1 1WB, UK

ARTICLE INFO

Keywords:
Emergent perception
Extended reality
Collaborative virtual learning environments
Auditory perception

ABSTRACT

It has been suggested that media and technology effectiveness in pedagogy is a myth. An intervention is not automatically effective simply by virtue of it being new. Yet, so often, initial hype leads to inflated expectations and subsequent disappointments. Virtual and augmented reality, the metaverse, and collaborative virtual learning environments that utilise an increasingly wide range of digital platforms have all made appearances within this narrative. However, initial failures to meet expectation, especially when value is taken for granted, should not condemn these technologies to being dismissed. With the burgeoning opportunity for heterogenic design (asynchronous and asymmetrical roles, tasks, interface platforms, user capabilities, etc.) their technical capabilities and pedagogic potential are too significant. The need is for deeper learning through meaningful experience, the latter facilitated through affective and cognitive engagement derived from user-experience factors that include presence, flow, and self-efficacy. The central assertion of this article is that the effectiveness of these technologies for learning can be greatly enhanced through user-centred software designs that focus upon evoking these factors. Hardware configurations and software designs should deliver training scenarios that are built upon research-informed interaction design. The twist here is that in this article, we look to the oft underappreciated field of auditory perception, specifically that pertaining to human interactions with (and through) digital technologies, to present a novel set of interaction design principles with the goal of enhancing extended reality collaborative learning.

1. Introduction

Is media effectiveness a myth? This sentiment is expressed strongly in Dillenbourg and Fischer, 2007 review of computer-supported collaborative learning. They point to several significant moments dating back to the 1980s in which a new form of media enters pedagogic discourse and is reflexively hailed as possessing intrinsic value to learning–and each time that assumption is shown to be false. Their position on this myth is not to make the equally presumptuous hypothesis that new media cannot improve learning outcomes, but that it can, provided we understand how such things influence learning.

When considering extended reality (XR) technologies (including but not limited to virtual, augmented and mixed reality), contemporary research points to significant problems with technical configuration and pedagogical grounding (Parmaxi, 2020). Studies have also shown that educational applications of XR present frequent problems in successfully evoking a sense of embodiment and presence for users, consequently impacting negatively upon learning outcomes (Pan & Hamilton, 2018). Lege and Bonner (2020) further observe that educational use of XR technologies is made significantly more challenging due to the constantly changing interaction capabilities of the systems, making it difficult to determine the affordances of the technology to build effective learning experiences upon. Kuleto et al. (2021) observe further difficulties with XR education related to a lack of user-centred design–specifically that those developing the learning systems are not understanding the bespoke needs of younger users.

These research challenges all feed into the overarching aim of facilitating deeper learning, revealing how simply deploying a novel technology will not by itself produce enhanced learning outcomes. Of course these are not basic challenges. Knowing that, for instance, weak presence negatively affects learning, only tells us that presence needs to be addressed. It does not tell us how to achieve that. Presence itself is not something that can be directly controlled, so how do we get from the aspects of a learning system that we can control to these elusive qualities and effects that we cannot? The following sections seek to trace a line between that ambition for deeper learning and the low-level contributing factors that it is possible for us as developers of CVLE systems to control. This is done first using a top-down approach by consulting the

E-mail address: t.garner@shu.ac.uk.

https://doi.org/10.1016/j.cexr.2024.100080
Received 10 July 2024; Received in revised form 14 August 2024; Accepted 31 August 2024
Available online 6 September 2024

relevant literature on computer-supported collaborative learning, then attempting to meet in the middle by applying a range of contemporary auditory perception research to suggest specific mechanisms by which desirable user-experiences can be reliably evoked that may, in turn, yield deeper learning.

2. Computers and collaboration

2.1. CVLEs and Computer Supported Collaborative Work

As largely social creatures, when we discover or develop a new human capability, one of our first instincts is to consider if and how we can experience it with others. As such, it is not surprising that the capability for multiple individuals to interact within some form of shared virtual space was quickly seized upon, in fields such as defence, academia, and computer games (San Chee, 2001). At the broadest conceptual level, a Distributed Interactive Application arguably best-encapsulates the subsequent terminology in that it captures any system that utilises networking technology to enable multiple users to interact with a software application (Delaney et al., 2006). More specific, a Networked Virtual Environment requires the presence of a virtual world, accessible to multiple individuals in real-time through any form of system input and sensory feedback. By the later 1980s, these networked systems had begun utilising real-time (interactive) graphics to create 3D virtual environments, most notably in defence applications including the Naval Postgraduate School Networked Vehicle Simulator (NPSNET – largely referred to in the relevant literature as a real-time networked simulator – see Storms, 1995) and SIMNET, a wide-area network military vehicle simulator. Although developed in the late 1980s, the term 'Collaborative Virtual Environment' (CVE) didn't appear prominently within literature until 1993, when exploratory research began to consider matters of system architecture (Benford, 1993) and models of interaction (Benford & Fahlén, 1993) with specific focus upon collaborative applications in a shared 3D virtual world. Though not expressly referred to as a CVE, SIMNET literature emphasises the function of cooperative interaction and team-based training (Lentz, Shaffer, Pratt, Falby, & Zyda, 1995, September; Miller & Thorpe, 1995). These systems evolved to merge with virtual learning environments (broadly defined as a computer and network-mediated environment facilitating communication and the exchange of materials relevant to learning [Wilson, 1996]) to become 'Collaborative Virtual Learning Environments' (CVLEs). Contemporary CVLEs require collision detection, gravity, kinematics, and behaviours as essential simulated components of a virtual environment, with autonomy, interaction, and presence posited as core user-affordances (Baladi et al., 2008), and their primary function is to realise the pedagogical benefits of computer supported cooperative work (CSCW).

In the first few years of the 1980s, the bringing together of the internet and computer networking with real-time interactive virtual environments precipitated a flurry of new concepts and exciting possibilities. Just as these technological capabilities were emerging, so too was the notion of CSCW, a then-new interdisciplinary field bringing together collaborative behavioural psychology and computer science technology. The acronym CSCW was coined in 1984, with what arguably stands as its first seminal publication, *Computer*-supported *cooperative work: a book of readings* (Grief, 1988), exploring a wide range of topics that remain highly relevant to the discourse nearly 40 years later – from artificial intelligence and simulation to human-centred and social computing. As the name would suggest, CSCW could refer to any usage of digital technology in support of cooperative work and the concept has underpinned developments in email, teleconferencing, file sharing and collaborative text/image production (Grudin & Poltrock, 2014). The 1990s saw the gradual adjustment of the term to Computer Supported Collaborative Work, with both terms seemingly used interchangeably (Hwang & Su, 2012).

As discussed by Bullinger-Hoffmann et al. (2021), key CSCW conceptual frameworks and theories include:

- The quadrant model of space and time: distinguishing multiuser systems in terms of heterogeneity (collocated or remote, synchronous or asynchronous)
- The 5C model: describing interaction forms within CSCW (coexistence, communication, coordination, consensus and collaboration)
- The Groupware paradigm: a move away from perceiving the computer as a tool for manipulating and sharing data, towards a human-centric perspective in which the computer is a shared space for collaboration
- Appropriation: the extent to which the system can integrate into existing professional and social practices
- Malleability: the ability for the system to accommodate a wide variety of work practices without extensive customisation or development of new sub-systems, tools or other technical content
- Benefit-orientation: offering a greater perceived reward for engagement than perceived effort
- Coordination Theory: coordination within a team must be implemented in the early stages of a project, facilitate direct communication between all individuals within a team, be continuous, and utilise situational leadership in which the approach to leadership adapts to the specific and changing needs of the project (Follett & Urwick, 1949)

For Grudin and Poltrock (2012), CSCW takes influence in terms of structure from McGrath's (1984) Typology of tasks – a framework that separates cooperative tasks into acts of planning, creativity, intellective (judgement/decision making), and performance/psychomotor. The Typology also includes 'conflict tasks' that, although distinguished from cooperation tasks, can have significant relevance to CVLEs. Conflict tasks include deciding issues with no clear correct answer, resolving conflicts of perspective, resolving differing motives and conflicts of interest, and competitive tasks for resolving conflicts of power. Dix' (1994) framework structures CSCW interactions across matters of understanding, control, feedback, feedthrough, and communication. Here, a system may enable both participants control over (and receive feedback from) the artefact, or it may limit that access to a single person. Direct communication can be afforded, but indirect communication can also occur by way of 'feedthrough', in which changes made by one person, itself communicates information to another. Removing direct communication in favour of feedthrough is referred to as 'anti-computer mediated communication'. It is also possible for each person to have direct control/feedback with a different artefact (or a discrete component of a single artefact) whilst maintaining direct communication.

As with CSCW, Computer Supported Collaborative Learning (CSCL) is better understood as a pedagogic approach than as a discrete platform for learning. A review by Dillenbourg and Fischer (2007) defines CSCL as an encapsulation of principles:

- Social interaction over individualisation of learning; scenarios and interactions over inherent features of a new technology
- Effort to construct shared knowledge must be minimised
- Greater resemblance to the 'real-world' is not automatically superior and 'realism' should not be chased at any cost
- Computer supported collaboration can influence both communication between learners and affect the way in which they reason about the material being taught
- CVLEs and similar systems cannot simply be created and deployed, their delivery must be scaffolded (e.g., around a traditional lecture/workshop curriculum structure)
- CSCL technologies offer automated interaction analysis, but full-automation offers too-limited insights
- group interaction can be captured and fed back to learners

– CSCL technologies should not have their definition limited to computers and should encapsulate any computer-connectable in/out devices
– CSCL can facilitate much larger cohorts than a classroom can hold and should exploit the potential for forming 'learning communities'

Benefitting from an academic discourse stretching back several decades, literature evidencing the learning outcome benefits of CVLEs includes numerous randomised-control trials, systematic reviews, and meta-analyses (Msonde & Aalst, 2017; Mystakidis et al., 2021; Raupach et al., 2009). In addition to performance-driven metrics, the literature also supports increased satisfaction and self-confidence in learning (Ryan et al., 2023), engagement, and low attrition (Reisoglu et al., 2017) as further educational benefits of CVLEs. As would be hoped for of a collaborative system, CVLEs have also demonstrated advantages in team-training contexts, with many studies evidencing their capability to improve learners' pair problem solving strategies (Fakomogbon & Bolaji, 2017), enhance collective creativity (Alahuhta et al., 2014), reduce intra-group conflict (Hsu & Chou, 2009) and, through greater accessibility, increase the diversity of a team (Morrison-Smith & Ruiz, 2020).

2.2. Extended reality and CVLEs

Within XR, academic literature characteristically focusses upon one of its three central forms: virtual (VR), augmented (AR), and mixed reality (MR). Along with VR, studies also repeatedly show meaningful benefits to the usage of AR and MR interfaces when integrated into a CVLE system (Garzón et al., 2019; Pellas et al., 2019; Wang et al., 2014). A review by Hamilton, McKechnie, Edgerton, and Wilson (2021) found that, of 29 relevant studies, head-mounted display VR (HMD-VR) produces some improvement over the control in 19 cases, showing a negative impact in 2 instances with the rest revealing no meaningful difference. It is also worth noting that the effect size across the positive results varied significantly, with some presenting very minor improvements and others claiming a substantial effect. Negative effects include greater cognitive load requirements (Makransky et al., 2019), though this has been shown to be an issue that can be rectified by way of conscious design and deployment. One example of mitigation involves the use of pre-training protocols, which has been shown to reduce the cognitive load attributed to VR-based training, mitigating the negative learning effects, and enabling the system to enhance learning outcomes overall (Meyer et al., 2019). Interest, motivation, and embodiment can be readily evoked using fully-immersive HMD-VR but cannot be taken for granted, whilst self-efficacy, cognitive load, and self-regulation also present opportunities to enhance learning if the system evokes them as positive effects. However, poor design and implementation can easily cause these effects to have a significant negative impact upon learning (Makransky & Petersen, 2021).

When considering XR for education and skills training, there is significant challenge in making systematic progress in our understanding. This difficulty comes from the considerably large number of moving parts that make up the landscape: domain of application; learning objectives, modes, theories and associated constructs; hardware and software capabilities; underlying design, human-computer interaction and perception theory; contextual and organisational requirements; and initial learner knowledge, skill, and motivation. Whilst rate and degree of change may vary, none of the above components remain static for long periods of time, with developments in one area often dictating a revisiting of our current understanding of another (Hammad et al., 2020). At present, there are relatively few published conceptual frameworks for CVLEs or collaborative extended-/mixed reality systems, with those available largely emphasising network configuration and either presenting very high-level system architectures or more specific designs, made to suit a more particular function (Hudák et al., 2020; Pereira et al., 2019; Wang et al., 2023). Though under researched, the landscape of multiuser XR platforms for education and skills training is substantial in both scale and complexity. It encapsulates multiple academic domains including computer science, psychology, physiology, pedagogy, and design–and this is before we consider the domain of application. A detailed taxonomy by Gerard (2011) structures aspects of the technology around a core distinction of inputs and outputs. In both cases, features are branched by way of sensory modality, broad approach to device interaction (determined as either handheld, wearable, or spatial), device type (e.g., microphone, data glove, inertia measurement unit, game controller, camera, etc.) and information type (e.g., object recognition, orientation tracking, language, gesture, etc.). Gerard's taxonomy highlights just how busy the XR technological landscape was now more than a decade ago. With numerous additional developments realised between then and the present, this landscape is now even more densely populated and complex.

One of the few review studies that explore XR technologies within CVLEs is Kostov and Wolfartsberger' (2022) framework for collaborative mixed reality. It places interaction (objects in the shared virtual space), representation (analogous to avatars), and laser pointer (a key mechanism for selection-based interactions) as essential components of a XR-CVLE system. De Back et al. (2023) present a dynamic logging model that captures user-actions (interactions, inputs, decisions, etc.) and body-related information (e.g., positional and rotational movements of the head and hands in standard VR, embedded eye-tracking or further physiological inputs, etc.). This captured data is then stored to facilitate offline in-depth analysis, but it is also fed into a real-time behaviour encoder module and a visual replay module to facilitate immediate feedback, sharing, and interpretation of the users' actions. For Plopski et al. (2022), the use of eye/gaze-tracking is expected to feature in future CVLE research and development across three areas: direct user-interaction (e.g., for selecting objects at a distance); implicit or adaptive environments and interfaces (real-time biofeedback in which the system adapts in response to its interpretation of user attention or affective state); and in collaboration with both other human users and in human-agent collaboration (in which the system is able to interpret meaningful information from the human user based on their eye data). Using the Citation Network (v.12, 2020), Nguyen and Bednarz (2020) visualise relationships between fields of study relevant to the topics of XR, user experience (UX), and communication. Their results included qualitative UX measures largely associated with forms of presence (situational, spatial, and contextual awareness); technology topics commonly denoting input/output hardware (VR/AR/MR, handheld, CAVE, immersive virtual environment) along with several references to 3D virtual environments; team-relevant terms (cooperative, collaborative, social, network); and means of communication (gesture, eye-/gaze-tracking, negotiation, facial expression, and non-verbal cues).

Although focussing upon AR-based collaborative systems, Marques et al., 2021 hierarchical taxonomy reveals the conceptual complexity that emerges when considering CVLEs and XR technologies together. Their taxonomy considers several core aspects that align with much of the conceptual work discussed above, emphasising matters of communication, team, input/output, task, tracking, actuation/interaction, shared context, and time. Of particular note, is their frequency analysis of published studies relevant to each characteristic within the above themes–providing a very helpful at-a-glance view of research gaps. Examples of this include: fully-asymmetric user-actuation; shared scene updates by way of static images, captured videos or live feeds; deployment in a mix of indoor and outdoor environments; sensory output modalities other than sound and vision; adaptive customisation of system inputs and outputs; task types relevant to management, negotiation or psycho-motor skills; making and committing to decisions as communication intentions; and asynchronous systems and matters of predictability (whether interactions within the system are scheduled or not). Furthermore, team features including size greater than three, collaborating over a long-term timespan, and physically distributed in a mixed reality context were also identified as significant knowledge gaps.

Task interdependence is separated into four sub-classes: pooled (each member can contribute independently); sequential (members contribute to a clear and well-established sequence of steps); reciprocal (the initial sequence of steps can be interrupted and adjusted based on real-time events); and intensive (all team members must work simultaneously with shared awareness of time pressure). Of these, the latter two were identified as knowledge gaps. Lastly, communication structure is divided into 'star' (every team member freely passes and receives information), 'chain' (communication follows a linear hierarchy), and 'hub and wheel' (all communication passes through a single team leader), with the former two modes identified as further knowledge gaps.

2.3. Heterogeneity, asymmetry, and asynchrony in CVLEs

The notion of asymmetric training relevant to XR largely appeared in studies published on or after 2019, making this a notably novel and emerging area of interest. At the widest conceptual level, the term 'asymmetric training' can denote any form of training protocol that features some form of non-symmetry in its design or delivery. It commonly features within physiology and sports science as a feature of a rehabilitation training protocol (Brown et al., 2017; Lee et al., 2014). In computer science, a relatively recent discipline to utilise the term, asymmetric training often describes an experimental feature of a machine learning algorithm or neural network that trains an artificial intelligence to fulfil a target function more effectively (Bao et al., 2017; Onen et al., 2022). Relevant to CVLEs, Kwon et al. (2019) differentiate between partial and fully asymmetric role collaboration. Described within an AR context in which one user is 'local' or based in-situ (interacting within a relevant physical environment) and another is remote, partial asymmetry would describe a situation in which the remote user could observe the physical environment and communicate with the in-situ user to guide them through the solution to a problem. A fully asymmetric system would afford the remote user the ability to directly influence the in-situ environment, to the extent that they could, if required, complete the task without the in-situ user being required to act. A common example of full asymmetry would be an IT service with remote desktop access, enabling the support advisor to directly fix an IT issue remotely rather than talk the customer through the process. It follows that this could be applied to a MR training scenario in which the system facilitates seamless interaction between virtual and physical aspects of the diegetic training space, with remote computer operation, teleoperation, robotics, and drone technologies all potential routes to full asymmetry.

At present, one of the most recent and relevant review articles is that produced by Mayer et al. (2023), who consider the current literature on heterogenic CVLEs, with a specific aim of uncovering research gaps. Here, asymmetry is clearly differentiated from asynchrony. The former describes a MR scenario, largely prescribed here by the technology or platform being used (e.g., all users in VR would be symmetrical, whilst some in VR and others using mobile AR would be an example of asymmetry). Asymmetry in CVLEs is often determined at a technological level, based on the selection of hardware being deployed within a single training protocol (Mayer et al., 2023; Plopski et al., 2022). Whilst this is certainly an appropriate means of identifying asymmetry, it could be asserted that this has some limitations and there is the option to also consider asymmetry at a software level by incorporating design, interaction, and feedback.

Asynchrony refers to difference in time, with asynchronous referring to situations in which users do not contribute to the task at the same time – the authors connect this to co-location as a means of effectively highlighting four quadrants of synchrony in CVLEs: synchronous distributed (e.g., video conferencing); asynchronous distributed (working on a shared document); synchronous co-located (in-person meetings); and asynchronous co-located (shift work). The review observed school-based/broader educational training to represent almost a third of CVLE use-cases, with military, medical, business and industry accounting for a further ten percent each. Differentiation by learner group revealed a significant lean towards vocational-based training and use of CVLEs for development of procedural and implicit knowledge, followed by deployment in HEIs and for the purpose of improving declarative and explicit knowledge. With regards to asymmetry and asynchrony, roughly twice the number of use-cases utilised a symmetrical approach when compared to asymmetrical, though the difference between synchronous and asynchronous was significantly more pronounced at a ratio of more than ten to one. Mayer and colleagues' identify three key gaps in our understanding pertinent to CVLE design as relevant to their findings: 1) embodiment and user-interaction; 2) remote and asynchronous setups; 3) technical issues such as multiple users interacting with shared virtual elements, scalability, and data security.

Considering the above and placing this in the wider context of technological developments in generative AI, platform engineering, and cloud native developing (to name a few), two things become clear. The first is that there is an abundance of technology with potential benefit to CVLEs, including automation (of assessment, scenarios, curriculum, and potentially even learning objectives), insight (via multifaceted performance analysis, engagement, and behavioural data, along with subjective user feedback), and flexibility (with firms increasingly able to engineer bespoke CVLE platforms with which non-technical experts can effectively and efficiently design and deliver content). The second, is that the huge array of choice on offer without direction from an evidence-based framework is leaving platform engineers and CVLE designers at the mercy of trial and error.

Within CVLEs, heterogenous design can effectively enhance learning outcomes by way of several affordances that include: enhanced accessibility, as individuals can use one of several hardware devices to access the CVLE (Ouverson & Gilbert, 2021); facilitation of new collaborative interactions and strategies (Clergeaud et al., 2017); and enhanced communication by enabling multiple viewpoints (Sugiura et al., 2018). Similar findings emerge querying the effects of heterogeneous design specifically upon XR systems, with accessibility and opportunity for a wider range of roles and simulated scenarios noted as key benefits (Burova et al., 2022). Contemporary CVLE ecosystems are no longer exclusive matters of human-computer collaboration, but through ongoing developments in telepresence/teleoperation, robotics and artificial intelligence-driven software agents, human-robot/agent and robot-robot/agent collaboration now exist as important topics for further investigation (Salehzadeh et al., 2022).

As the use of XR interfaces and heterogeneous configurations is a very recent topic of interest, there remains much opportunity to develop and evaluate such solutions, but it would follow that the importance of good design and deployment, just as it appears vital in earlier CVLE contexts, must be at the foundation of our approach to this new technology. This article presents emergent perception theory as a powerful means of understanding the complex and dynamic nature of contemporary CVLE ecosystems, informing developers to build more effective learning solutions. Emergent perception theory helps us to explain the relationship between factors that are tangible and controllable, the perceptual effects they can evoke, and how these effects collectively give rise to learning.

3. Emergence and learning

How do we improve learning? Whatever answer this question brings to mind it is highly unlikely anyone would suggest that there is a special button we can press on a person to enable them to learn. We can take the horse to water … But is it that simple, or that limiting? From where does learning emerge and what *could* we control that, if indirectly, enhance learning? The assertion presented here is that the answer can be found in the understanding and application of emergent perception.

3.1. Emergent perception

The notion that all observable phenomena are formed not only by the presence of component elements, but by the configuration, orientation, and interactions between these elements has been explored and validated across several disciplines including physics, neuroscience, and education (Tachihara & Goldberg, 2019). Within cognitive psychology, the term 'emergence' represents a very recent concept, but not in philosophy. As McCelland, (2010) observes in the Lewes (1877), ***Problems of Life and Mind***, we have long been considering the notion that there are certain aspects of ourselves and of our experience that result from cooperant elements and forces, and which cannot be reduced to the sum of their parts. More recent work by Penner (2000) describes an emergent system as a multilayered set of interactions between micro-level components through which macro-level components emerge and, in turn, interact with each other to realise further levels above. Here, micro components may be small, but they are not insignificant, and changing their state can potentially have a dramatic effect upon the emergent higher levels. Similarly, 'emergentism' posits that cognition is an interaction between organism and environment, and that the functions of the mind do not support predetermined or domain-specific understanding or capabilities (Gregg, 2003).

Early usages of the term 'emergent perception' include Chaplin (1973) in the context of present-day feelings emerging from childhood experiences, and McGraw (1995), who suggests our perception of alphabetical letters is an emergent process not solely based on the constituent parts of the letterform, but also through multiple top-down influences of letter concepts. Though the term is not explicitly used, Van Orden and Goldinger (1994) describe a perceptual framework that documents several theoretical components relevant to emergent perception. Specifically, that cognitive systems are best understood through a lens of experiential realism. This position emphasises the value of experience, both that which is actual (past experiences retained through memory) and that which is potential (possible future experience, evoked by way of the imagination). It also postulates that an objective external world does exist, and that the nature of that world places restraints upon what we can conceive. It suggests that cognition is embodied, and that our conceptual systems emerge from bodily experience, precisely for the purpose of understanding that experience. Lastly, it asserts and that classical categorisation of cognition is inadequate, as cognition operates through relative and 'fuzzy' mechanisms (Muma, 1991). Simpson (2011) considers in depth the notion of emergent perception, within the frame of Maurice Merleau-Ponty's perspectives upon phenomenology. A noted early 20th Century philosopher, Merleau-Ponty asserted that the physical world, including our own physical form, is a deeply integrated Gestalt construct that is formed of characteristic properties that emerge from an aggregate of integrated components to form a higher level of complexity. Simpson explains perception in Merleau-Ponty's emergent frame as a system of systems; a non-linear, self-organising process incorporating environmental components by way of the senses and bodily components by way of proprioception.

The role of the body in cognition, in stark contrast to prior dualistic perspectives, has been an important topic of study for many years.

Margaret Wilson's (2002) *Six Views of Embodied Cognition* arguably represents the breakthrough paper on embodied cognition. Although a single-statement summary is arguably a little reductive, embodied cognition posits that cognitive processes (what could be denoted as the 'mind') are inexorably linked to (or held within) the brain, which is itself within the body, which is within the world. This physical configuration means that the mind cannot be meaningfully understood as an independent entity, nor can it function as such. Cognition is situated in (and deeply influenced by) both space and time, and its fundamental function is to support interaction within the world that is of (at least perceived) benefit to the self. Contemporary explorations of perception reliably follow on from much of Wilson's assertions, including the Interface Theory of Perception, that argues for perception to be understood not as a means for realising truth, but for realising useful action–and that in many cases, access to the truth can deter optimal action (Hoffman et al., 2015). Importantly for this paper, approaching the mind as an emergent phenomenon inexorably tied to the brain, the body, and the physical environment, has significant implication for learning. Numerous studies have attested to the value of situated learning, of which our fundamental bodily experiences are an important aspect (Kosmas et al., 2018; McCelland et al., 2015; Núñez et al., 1999). A review of 44 papers by Duijzer et al. (2019) found that the wider literature evidenced students' bodily experience as an essential contributor to learning. Their work also identified the extent to which an embodied training experience reflected a real-world context, utilised multimodality of sensory feedback, provided clear and coherent semiotics, and provided multiple representations of information as key features of a system that would enhance learning outcomes. Further modulating factors included attention-capture, learner control, and cognitive dissonance.

So, if we are to accept the notions of emergence and embodied cognition, we are accepting that the body and the environment can be manipulated to affect the mind–in our particular interest, to facilitate deeper learning. Utilising a top-down strategy, we must first look to our aim of deeper learning as the uppermost emergent effect at ask what factors make up its aggregate. As theoretical understanding of CSCW and its practical application to learning within CVLEs grew, several factors identified as key contributors to learning emerged. These factors include presence, flow, interactivity, motivation and interest, self-efficacy, coherence, autonomy, and diexis (shared understanding).

3.2. Presence

Presence is something of a fuzzy term that can be broadly understood as a feeling of *being there*, though its precise definition and evoking mechanisms are not widely agreed upon (Skarbez et al., 2017a). Considering presence in a VR context, research by Riva (2009) suggests a causal relationship between presence (being there) and self-efficacy (being able). Presence requires vivid, representationally accurate, high resolution, and consistent multisensory experience, gained from a personal perspective, with interactions that are responsive, naturalistic, scalable, reliable, low-latency, and have persistent effects on the environment (Baladi et al., 2008). Järvinen's (2017) model of presence delineates four ways in which we can feel *there* within a virtual world. Active presence describes being drawn into a virtual world by interacting physically, using our bodies, affecting our physiological state through physical activity or exercise, or simply the feeling of being physically proximal to virtual entities. Embodied presence refers to a sense of connection to a virtual form or avatar through coherent mapping of user-body to avatar-body. Such mapping could include behaviours, viewing perspective, or a sense of purpose or role. Emotional presence captures intellectual and affective connections to the virtual world. This could refer to activities such as problem solving or interpretive or intuitive tasks but could also include narrative or even cinematic aspects of the world that evoke an emotional response. Lastly, social presence describes connections with other characters within the virtual world. Such characters could be other human users or computer-controlled non-player characters, with the connections drawn by way of interpersonal communication, engaging in socially centred activities, or through roles or tasks that are carried out within a group to evoke a sense of shared responsibility and belonging.

Presence has a history of association with technology-enhanced education. In CVLEs, embodied presence has also been shown to improve learner performance (Chen & Wang, 2018; Mikropoulos, 2006). Research also posits that XR interfaces can enhance CVLEs by way of evoking presence, which it often characterises as an emergent experience linked to the immersion facilitated by devices such as extra-large/projection displays or head-mounted displays (Cummings & Bailenson, 2016; Reisoglu et al., 2017). The benefits of presence to learning are most frequently identified when considering social presence specifically, both between learners and, if relevant, their instructor (Richardson et al., 2017). The *Community of Inquiry* model (Garrison et al., 2010) locates educational experience as central to an overlap of social presence (the self and sense of personal characteristics is projected into the learning activity), cognitive presence (construction of meaning is supported through opportunity for reflection and discourse), and teaching presence (high visibility of instruction and a culture of inquiry). Nguyen and Bednarz (2020) consider co-experience in collaborative XR platforms. Although they acknowledge that the transdisciplinary nature of the topic presents a substantial range of future research opportunities, they emphasise gaps in our understanding of social presence and copresence, asynchronous collaboration, long-term usage, increased merging of physical and virtual content, user-representation/avatars, and group dynamics and collaboration patterns. This suggests that although presence in a multi-user/CVLE context has substantial potential to modulate learning outcomes, significant further work is needed to understand it fully.

3.3. Flow and interactivity

Flow describes a sense of balance between challenge and capability that can be characterised by high concentration, low stress/anxiety, high relaxation and high task satisfaction (Pereira et al., 2022). A prevalent topic of study within sports science, systematic reviews of the literature reliably point to flow state contributing to enhanced athletic performance (Jackman, Dargue, Johnston, & Hawkins, 2021), though similar results have also been presented in music, languages, computer-engineering contexts (Pereira et al., 2022) and replicated further in digital game-based learning (Perttula et al., 2017). Flow states have been observed to elicit positive emotions and enhance academic performance within computer-enhanced learning (Rodríguez-Ardura & Meseguer-Artola, 2017). Flow in multiuser virtual environments has been associated with multiple other learning factors including presence, motivation, and interest (Faiola et al., 2013).

Interactivity reflects a foundational principle of constructivist pedagogy, that the best way to learn is to *do*. Meta analyses on the subject reveal that interactivity can enhance learning outcomes, but not simply by virtue of being present. Basic website interactivity for example, has been shown to increase engagement but not enhance learning outcomes such as comprehension, memory, or knowledge gain (Yang & Shen, 2019). By contrast, when well-designed, implemented, and supported, interactivity has been shown to reliably improve learning (Maor & Volet, 2007). Schäfer et al. (2022) position interactivity as one of three pillars of CVEs, alongside environment and avatars. One of the earliest theoretical works to be published at the time CVEs were formally entering academic discourse, Sims (1994) provides us with a taxonomy of interactivity relevant to technology-enhanced learning. With 7 levels, Sims presents a continuum of low-to-high interactivity forms, which includes: passive (linear scrolling through a sequence of materials), hierarchical (non-linear set of predefined materials), update (system provides predefined problems to which the user must respond), construct (user manipulates objects within a system to complete a goal), simulation (procedural training requiring the user to complete a task by executing a series of steps), free (expansion of hierarchical to the extend the user meaningfully explores the information), and situated (a complete virtual environment that attempts to accurately mirror a real-world situation and/or task).

3.4. Self-efficacy and coherence

Self-efficacy describes a belief in an individual's self-perceived capability to accomplish a goal. Conceptualised by Albert Bandura (1977) as part of a unified theory of behavioural change, self-efficacy is defined as the result of affective state (broadly positive emotional experience and psychological wellness), verbal persuasion (extrinsic motivation, encouragement, support), vicarious experience (modelled behaviour with feedback linking to positive or negative outcomes), and mastery experience (reflection upon past personal successes and failures). Self-efficacy can be understood as a positive, four-step learning journey in which the learner is at first cautious, aware of their newness to the topic or skill, unsure in decision making, and concerned about how they are perceived by their peers and tutor. Through continued engagement, with extrinsic motivation and social presence, the learner builds relationships within the learning cohort to reconcile with their current low-capability and restore any diminished self-esteem. Further engagement progresses the learner to the third step, characterised by a swing from extrinsic to intrinsic motivation and increasing self-confidence before they reach the fourth state of self-efficacy, at which point they have manufactured perseverance in their continued learning, resilience against future failures, and willingness to take risks in the personal goals they set themselves (Bray & McClaskey, 2016).

Self-efficacy has been evidenced to be a significant predictor of increased learning by way of enhancing motivation, persistence, and breadth of engagement in a topic (Zimmerman, 2000), with similar assertions evidenced in online virtual learning environment contexts (Hodges, 2008). CVLE literature reinforces this mechanism, arguing that the need for self-efficacy is increased substantially in complex, team-based training scenarios (Schaffer et al., 2012). Systematic reviews on the subject have linked the use of virtual simulations to self-efficacy (Penalo & San, 2020) and evidenced that self-efficacy, driven by the higher levels of user control offered in such environments, contributes significantly to transfer of learning (Gegenfurtner et al., 2014). Designing a CVLE system with self-efficacy as a key requirement for learning has been evidenced to enhance social presence and knowledge exchange (Kubo et al., 2002) whilst also forming a positive feedback loop in which the social presence afforded by collaborative learning systems contributes to self-efficacy (Hatami, 2015).

The coherence of a CVLE refers to the extent to which the system facilitates a learner's ability to acquire, interpret, and apply information (Segedy et al., 2015). Numerous studies have shown coherence to be an important contributor to positive learning outcomes (Lowyck & Pöysä, 2001; Smeby & Heggen, 2014). Research has also evidenced a relationship between coherence and interest in which if either coherence or interest is high, it can help mitigate negative the impact of the other being low (Muller et al., 2008). Research in CVLEs has expressly promoted coherence as a means of supporting interaction between learner groups and their tutors (Kubo et al., 2002). Tsiatsos et al., 2010 assert that coherence in CVLEs can help realise learning objectives by way of co-presence and easy familiarisation. Basing some of their framework upon the cognitive theory of multimedia learning (Mayer, 2014), De Back et al. (2023), identify several contributing factors to strong CVLE-coherence that include carefully segmenting information, providing pre-training system-familiarisation, adopting intuitive interaction and interface design, and providing consistent and direct feedback. With specific regard to collaborative learning, the authors further advocate for balanced accessibility between learners, leveraging platform-specific affordances whilst also mitigating their limitations, minimising the effort (and also time and cost) to both learners and tutors, and maximising collaborative learning through interdependent and active participation and instruction.

3.5. Autonomy and shared knowledge/understanding

For Fierro-Saltos et al. (2019), autonomous learning is synonymous with self-regulation and describes scenarios in which the learner takes some form of control over their learning activity. A recent systematic review by Raitskaya and colleagues (2021) positions autonomy as a vital pre-condition for self-efficacy and asserts that virtual learning systems can help foster this autonomy by providing accessible, flexible, repeatable, and asynchronous learning opportunities not traditionally feasible in a classroom environment. It is unsurprising them, that CVLEs have been repeatedly identified as useful means of promoting autonomy (Pursio et al., 2021). Multiple reviews have been conducted on the matter of autonomy in learning, linking it to engagement (Nii & Yunus, 2022), confidence and professional identity (Allen et al., 2019), openness to change, and enhanced personal responsibility (Oļesika & Rubene, 2023). Autonomous learning is, however, as much an end as it is a means. Its realisation is dependent upon the fostering of commitment and positive attitude towards the topic of learning (Hernández et al., 2021). Without these assets, approaches to learning that require autonomy and self-regulation can suffer from high drop-out rates and can have negative impact upon overall learning outcomes (Fierro-Saltos et al., 2019).

Autonomy can itself be the subject of training, as evidenced in a meta-analysis by Su and Reeve (2011) that revealed autonomous learning can be nurtured by way of non-controlling language, evoking learner interest, acknowledging the validity of multiple perspectives and approaches, and by providing rationales for the curriculum and pedagogy. Interestingly, affording learners choice between multiple activities produced the lowest improvements to autonomy. Their work also showed that the learner demographic significantly modulated intervention efficacy, with teachers (compared against practitioners and non-professionals), inexperienced trainees (against experienced practitioners), and individuals with a personal orientation towards independent learning all substantially more responsive to autonomy training.

Lastly, shared knowledge and understanding is, unsurprisingly, a recurring theme in CVLE literature. Dix (1994) uses the term 'diexis' to refer to the various configurations of communication that can give rise to shared understanding relevant to the subject or object of knowledge ('artefact'). For example, two users may have direct communication with each other and equal ability to interact with a single artefact, or each user may have a separate artefact that they can coordinate their interactions with by way of direct communication. Alternatively, one user can use their unique perspective of the artefact to directly guide the actions of the second user, and so on. The literature broadly endorses shared understanding as beneficial, particularly where learner equivalence or teamworking skills are important priorities (Bittner & Leimeister, 2014; Weinberger et al., 2007). In a professional context, Ahmad and Karim (2019) link knowledge sharing activities to enhanced creativity, individual learning outcomes, and future team cohesion. A systematic review by Sensuse et al. (2021) evidences trust, appreciation, management support and clear organisational goals as key to successful knowledge sharing in a CVLE context.

The above factors do not represent an exhaustive list and their precise contribution to learning is unclear. Each factor reveals multiple inter-relational effects with multiple others, and their effectiveness in enhancing learning outcomes is arguably heavily dependent upon a further array of contextual circumstances. What is clear is that, as a general rule, evoking presence, flow, a sense of cohesion, and so on, can enable us to enhance learning. However, the consistent challenge across every one of these factors is that they cannot be directly manipulated themselves. There is no button one can press to increase their sense of self-efficacy, nor any of these learning factors. They are themselves emergent experiences. Therefore, in the development of effective virtual learning technologies, we must identify the tacit (i.e., directly controllable) components of the XR-CVLE ecosystem and understand how they can be positioned, oriented, and configured to facilitate the emergence of learning factors.

4. The fundamental frequency

4.1. Applying emergent perception theory to collaborative learning

Attempting to consolidate much of the above theory and apply it to the context of auditory perception, Grimshaw and Garner (2015) propose a model of sound as an emergent perception. The model asserts that, when we listen, countless physical and psychological factors collectively form an 'aggregate'. Here, physiological factors are referred to as 'exosonic' and include soundwaves, non-auditory stimuli, physiological state, behaviour, and spatial/material properties of the environment. Psychological factors are known as 'endosonic'. They include memory, imagination, expectation, and belief. In any given moment of experience, the components that make up the aggregate, but also their individual state, configuration, and orientation, give rise to the emergent sound that we perceive. This moment of experience that reveals the emergent perception from the aggregate is referred to as 'actualisation'. There is some conceptual resonance with that of a musical note in that, within the above model, perception has a theoretical base level, from which it is not possible to reduce further. This we can equate to the fundamental frequency of a note. The emergent property of that lowest aggregate may then contribute to the formation of a new aggregate, that itself can be actualised to form a 'harmonic', adding complexity and texture to the perception. Just as more harmonics will produce a richer musical note, the more layers of emergence, the greater the complexity and the richer the experience.

If we are to accept emergence as a general principle for understanding complex systems, of which the learner experience in CVLEs arguably qualifies, we must seek to identify: 1) the top-level emergent effect of the phenomenon; 2) the macro-level components that collectively actualise that effect; 3) the micro/sub-components that determine the macro-level–continuing until we reach the fundamental frequency (the aggregate containing tacit components that we can directly manipulate); 4) the configuration and orientation of the tacit components that form the aggregate, and; 5) the dynamics and interactions between components that collectively form the actualisation. From such a model, we may then look to determine which components of the fundamental aggregate are within our power to control and what the extent of that control is. Fig. 1 (above) presents a theoretical solution to four of the five requirements:

The first step is straightforward as the uppermost emergent effect should reflect the ultimate deliverable of a CVLE–the *learning aggregate*. For the second step, we can return to the discussion in Section 2 of this paper, in which the literature broadly identified motivation, flow, interest, agency/self-efficacy, presence, autonomy, and understanding, as the primary factors contributing to learning. This provides us with the first aggregate of factors that have been well-established across the literature to be direct contributors to learning. Containing only the effects that actualise from within the user, and not features or qualities of the system, the learning aggregate cannot be directly manipulated. Therefore, we move to the third step to consider the factors that contribute to it. Here, the *experience aggregate* describes the culmination of all relevant factors that evoke learning but do not emerge from within the user. These factors still do not define features of the system directly, but they do describe the experiential qualities that can be attributed to it–namely its quality (depth, accuracy, consistency) of interactivity, feedback, immersion, and coherence. Of course, such qualities can be realised with a button-press. Furthermore, they are not determined entirely by the system itself, but by the interactions between the *system aggregate* and the *user aggregate* (also incorporating the *collaborative aggregate* if the system is multiplayer).

Here, the user aggregate can be delineated in alignment with embodiment theory to reveal the *physiological aggregate* (directly describing the physical characteristics/states of the body), *proprioceptive aggregate* (relative to the perception of characteristics and dynamics of the body), and *psychological aggregate* (cognitive and affective states, traits, and processes).

On the system-side, the *virtual aggregate* describes the digital aspects of an intervention and the core features of a virtual world that are presented through responsive graphics and sound. It also encapsulates the behind-the-scenes data systems that govern the world as well as mechanics such as modes of interaction, physics, weather, lighting, and other environmental effects. The *diegetic aggregate* is primarily concerned with matters of situational learning and roleplay. It incorporates

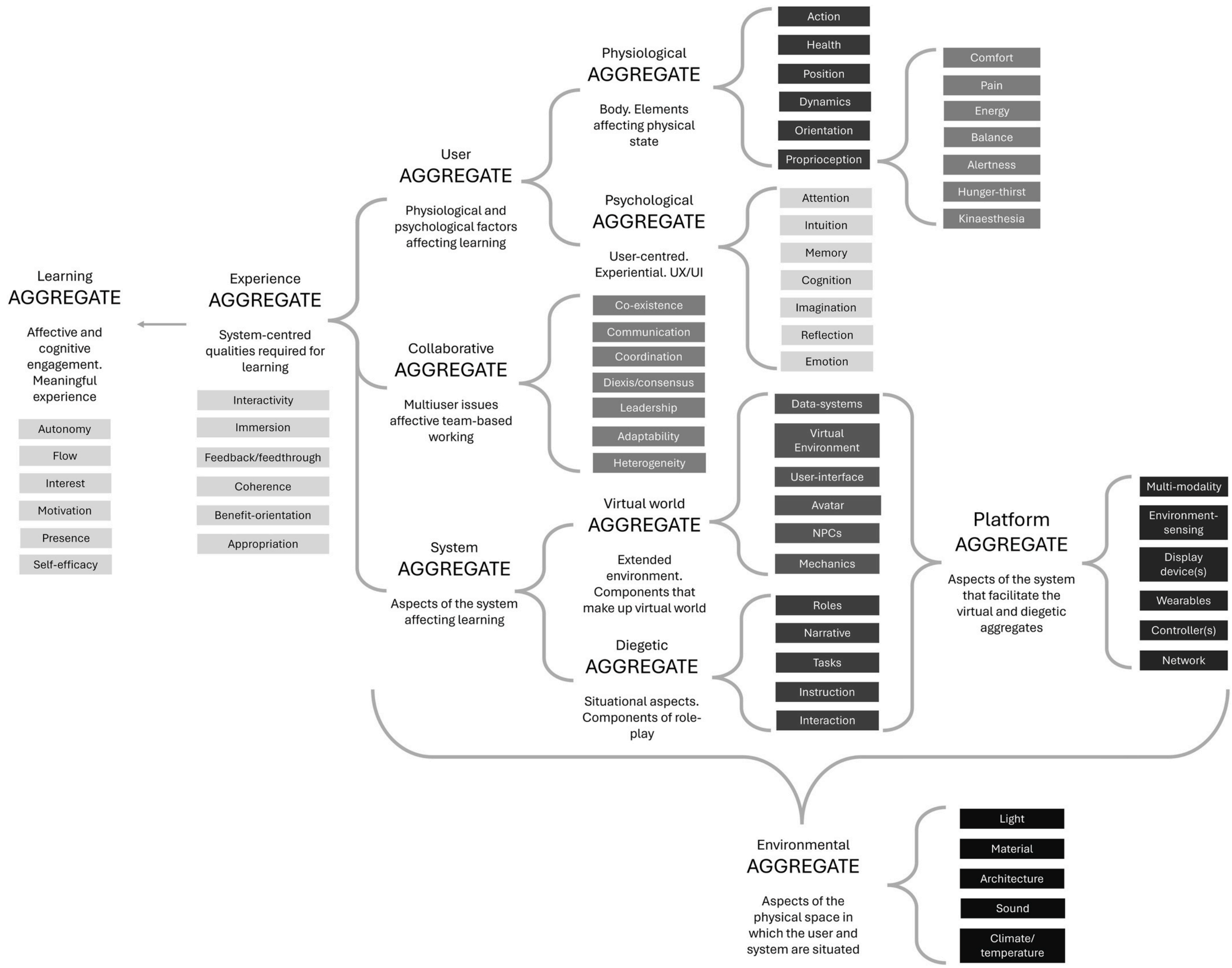

Fig. 1. XR-CVLE: emergent learning experience model.

factors related to the narrative of the fictive or simulated environment and can includes cinematic events, characterisation, tasks, instruction, and interaction. The *platform aggregate* incorporates the physical aspects of an intervention, inclusive of overt user-interface hardware (input/controller, output/display) and background/hidden hardware such as environmental and physiological sensors, network technologies, and pre-processors.

Lastly, the *environmental aggregate effectively* encapsulates everything in a physical sense, impacting upon the learning by way of the physical characteristics of the immediate space within which all other factors are situated. From the perspective of a developer, the 'fundamental frequency' of a CVLE is that which we can directly control. This is what can potentially separate XR-CVLEs from traditional systems. Whilst the latter can only manipulate the system, XR presents opportunities to meaningfully control the wider environment, significantly increasing our ability to influence the user. This affords us more power, but not yet clear understanding of how to utilise it effectively. The challenge for XR-CVLE design is that to effectively reach the learning aggregate, we must understand as comprehensively as possible the configuration and orientation of everything that feeds into it, so that we may reliably determine the emergent effects. This is the fifth and final requirement of an emergent model, and it represents the scope of the proposed research agenda at the end of this paper.

4.2. Multilayered and looping effects

Although it would be fair to suggest that the emergent theory as described is already a little complicated, the framework presented in Fig. 1 implies that emergent effects operate in a single direction–the lower-level elements collecting to form aggregates, then emergent effects give rise to higher and more abstract experiences. However, returning to Grimshaw and Garner's (2015) model, Fig. 2 (below) visualises a looping, multilayered principle of emergence, in which it is possible for an actualised emergent property to be integrated into that same aggregate to potentially influence continuing emergence (described in the model as 'reaggregation'). This model takes influence from research into the phenomenon of perceptual priming, itself based on observations of both behavioural and physiological differences when a given stimulus is experienced repeatedly (see Wiggs & Martin, 1998). Our own earlier work into the effects of video game sound upon a user suggested that perceptual priming could be modulated when considerate of the effects an initial sound could have upon the perception of a subsequent sound (Garner & Grimshaw, 2011). An example of this looping effect could be looking at a painting in an otherwise static environment and discovering that your perception of that painting shifts with extended viewing as your initial emergent experience is retained and fed forward into the next moment, influencing your subsequent perception and thereby affecting the overarching experience. You

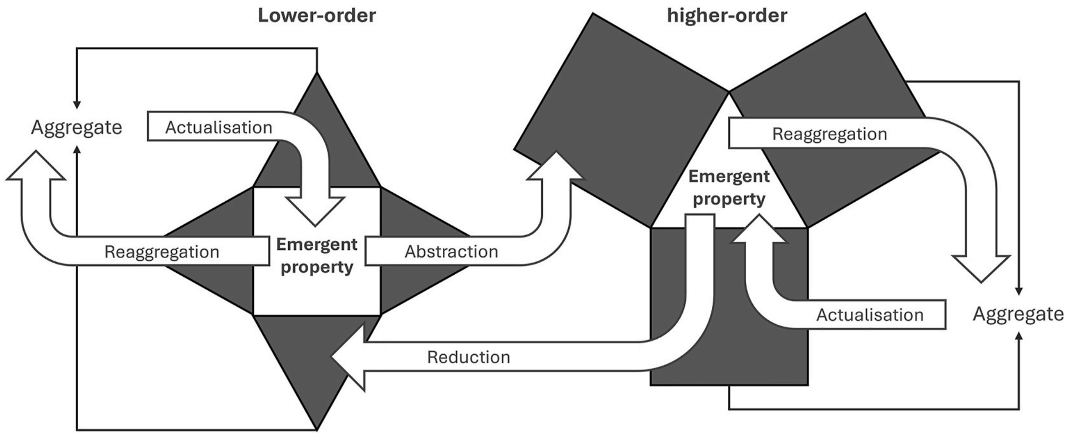

Fig. 2. Multilayered model of emergence.

observe the vivid colours that evoke a positive emotional response and an emergent experience of aesthetics. This response persists as you continue considering the colours, making them seemingly appear even more vivid. The term 'abstraction' describes the potential for an emergent property to become part of a higher-order aggregate that can itself be actualised to form a further emergent property, whilst 'reduction' denotes a higher-order emergent property being retained and influencing the aggregate and subsequent emergent experience of a lower-order emergent property. Returning to our painting, imagine the positive emotional state and sense that the painting is beautiful, as evoked by the colours, encourages you to examine the painting further and you turn your attention to the texture of the brush strokes–your interpretation of this lower-order aspect of the painting is now being influenced by your prior higher-order sense of aesthetic.

In a CVLE context, this provides an explanation for why vicious and virtuous cycles can be so powerful in learning (Wäschle et al., 2014; Wong et al., 2022) and presents a clear challenge to CVLE developers in that they cannot presume that a configuration of factors that yielded a desirable emergent outcome in one instance will have the same impact in the next, even upon the same learner.

In the above discussion, we have considered the effects that an emergent perspective can have upon learning. We have identified elements that make up the learning aggregate from which, when actualised, emerges learning. We considered some of the more fundamental (and controllable) aspects of the body, the intervention system, and the environment that indirectly contribute to the learning aggregate by way of proprioceptive, cognitive, social, and affective properties. As a means of demonstrating the potential application of this model, the following discussion presents two system design frameworks including a XR-CVLE technology ecosystem for heterogeneous collaborative training and an adaptation of emergent theory, previously applied to auditory perception, here expanded into a framework of emergent experience. The central assertion raised here is that learning aggregate factors can all be indirectly manipulated in CVLE design by careful control of the virtual, diegetic, platform, physiological and environmental aggregates–something which is uniquely possible in an XR-CVLE ecosystem.

5. Putting the fundamental frequency into practice

5.1. Modes of perception

As noted earlier, the approach to this work is to attempt to draw a line between deeper learning and controllable aspects of an extended CVLE through both top-down and bottom-up processes. So far, the focus has been upon the former. Here, we consider more practically relevant matters in which specific understandings on human interaction with sound are examined for their applicability to XR-CVLE development. Whilst these theories and taxonomies do overlap in places and differentiate themselves in others, one position they all share is that our relationship with sound is fundamentally concerning perception and action. Sound is *for something*. It is intertwined with behaviour and understanding of the world around us. The nature of that understanding and the behaviours it affects is largely what separates the following ideas. To elucidate the key differences, Fig. 3 (below) maps the various concepts along three dimensions of emergence. Location describes the perceived point at which our perception is focussed. If our experience is directing our perceptual focus outwards and towards objects in space, then the experience is distal. If we are encouraged to consider the space between us and the external events or objects, then our experience is medial, whilst an experience that makes us perceive inwardly (upon our mind or body) would be described as proximal or perceptual. Emergent form is closely related to location and largely helps to clarify the meaning of the latter (medial location describes focus on an event, proximal upon user-sensation, etc.). Lastly, emergent mode emphasises what an experience reveals to us based on our interactions with it.

Considering the various listening modes, many (but arguably not all) can be expanded into a multisensory context relevant to user experience within CVLEs. Truax's (1984) listening in search and listening in readiness translate respectively to attentively scanning the virtual environment for a cue of any sensory form that can be experienced within the extended reality platform, or intentionally dissipating attentional focus to allow sudden changes in the environment to be detected more reliably and immediately. Background listening can also be expanded to describe any scenario in which a user is concentrating on a task but is, if possible, intentionally keeping at least one sensory input stream 'available'. One example of this could be utilising peripheral vision to passively attend to the wider environment whilst concentrating upon a task within their central vision.

Chion's (2012) causal, semantic, and reduced listening modes can be readily expanded to functions of perception and interaction within a CVLE, with object-nominal identification, interpretation of instruction, and analysis of sensory characteristics all cognitive processes that can be exemplified in multi-sensory, virtual contexts. Bijsterveld's (2019) notions of interactive and synthetic listening both address dynamic matters of sound but can be applied more generally to interactions involving change within a virtual environment, whilst analytical listening is arguably comparable to reduced listening. Likewise, Rebelo et al., 2008 taxonomy of theatre, museum, and city listening can be readily expanded to CVLE interactions more broadly; the three forms distinguishing between a virtual environment that at a given moment facilitates only cognitive/affective interaction (theatre) compared with one that enables direct/behavioural interaction within a fixed, likely linear, space (museum) and, an open world type of environment supporting non-linear exploration and open interactions (city).

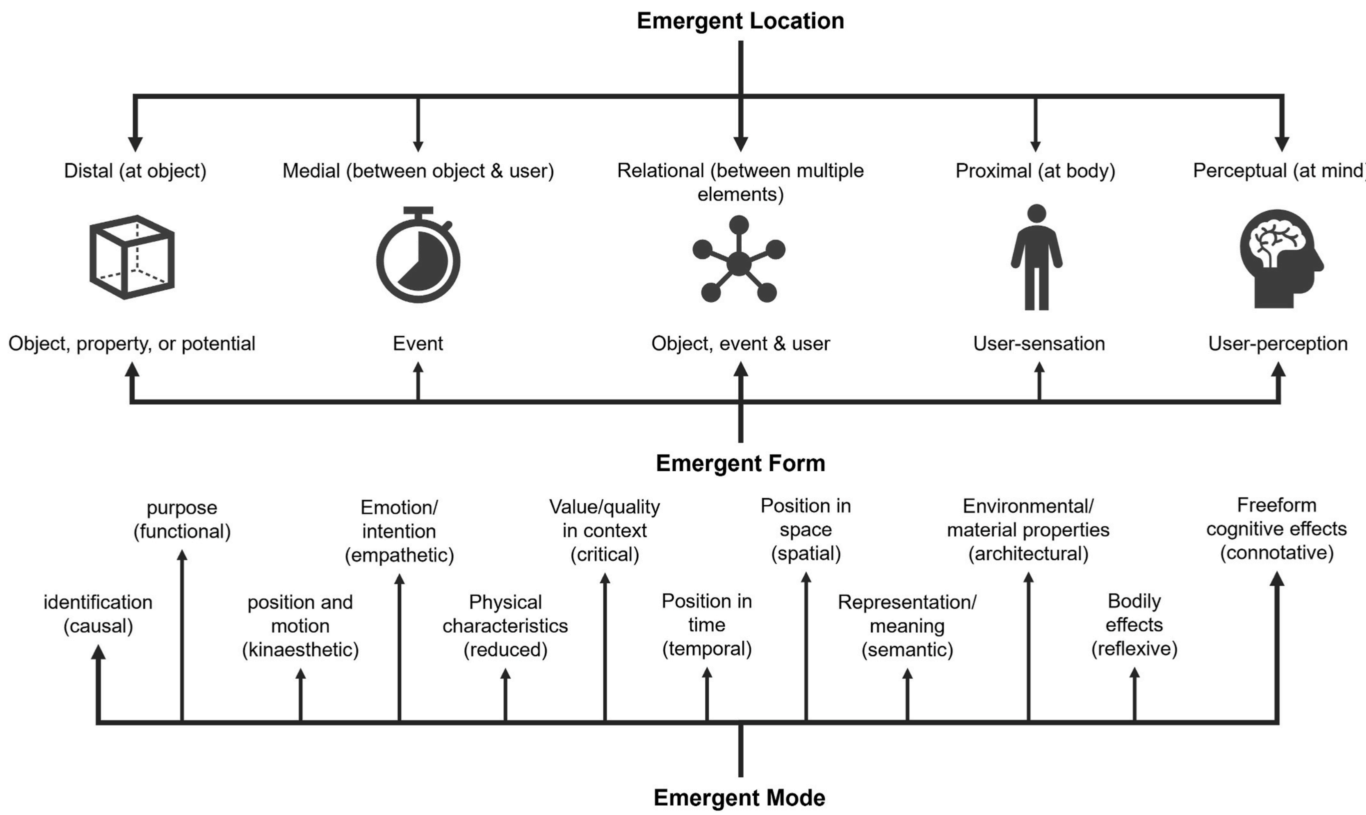

Fig. 3. Mapping of emergent location, form, and mode.

Determined specifically with virtual environments in mind, Grimshaw (2007) presents several unique functions of listening that have relevance to multisensory interaction. Considering aspects of a virtual environment for the purposes of getting from one point to another (navigational listening), interpreting terrain and architectural features and qualities (choroplast listening), and establishing a sense of situational (chronoplast listening) and setting-based (aionoplast listening) temporality are all cognitive functions that support user interaction and that make sense in CVLE perception and action more broadly. Furthermore, Grimshaw's Topoplast function arguably has even greater relevance to mixed reality scenarios, in which the presentation of sensory information to users may be both a blend of physical and virtual, but also of diegetic and non-diegetic stimuli.

Tuuri and Eerola's (2012) listening taxonomy delineates its modes along a continuum of cognitive engagement ranging from the most immediate (reflexive, pre-attentive) to the most considered (critical) means of evoking meaning from sound. As with the other modes discussed above, all of Tuuri and Eerola's taxonomy can be extrapolated quite neatly onto a wider, multisensory CVLE context. Rather than focussing exclusively upon auditory perception, these modes can describe the perceptual and physical effects that both virtual and actual stimuli can have upon a user (what we could call reflexive, kinaesthetic, and connotative *experience*), and the forms of information that can emerge from the cognitive engagement with CVLE stimuli (empathetic, functional and critical experience).

Table 1 (below) is an initial attempt to consolidate all of the above discussion by mapping the full range of listening modes, first by translating them into broader multisensory descriptions, then by connecting them to the most applicable emergent locations, forms, and modes. Lastly, each mode is mapped to the emergent learning experience model (presented earlier in Fig. 1) to highlight potential areas of focus for XR-CVLE developers who wish to evoke specific experiential modes.

To provide a brief example, a developer may wish to evoke learning through deeper presence. They decide to utilise embodied cognition theory to leverage kinaesthetic sensation to give the user a sense of proprioception and bodily awareness relative to the virtual environment. They subsequently identify reflexive and kinaesthetic experience as viable routes to enhance proprioception–for example, cues that evoke sensations of tempo and rhythm, encouraging a pattern of bodily movement, and a task that requires the user to manipulate their bodily movement in order to realise a gamified objective.

5.2. *Emergent perception to create a heterogenic XR-CVLE ecosystem*

Building on both the heterogenic ideas outlined in section 2, and drawing heavily from the emergent experience model, the XR-CVLE ecosystem (Fig. 4 below) is designed to exploit the capabilities of XR technology and provide control for CVLE developers over all facets of the accessible aggregate.

In what is arguably a rather decentralised system, no individual component has overbearing prominence and there is little explicit hierarchy. Except for the system manager acting as administrator, every element of the framework exists within a singular diegetic space. This represents the fictive world in which the roles of the learners and the narrative are established and within which the entire training scenario is enacted. The diegetic space contains all learners, all the hardware that will facilitate the digital components of the ecosystem, but also additional physical props that facilitate the roles and narrative, but which may or may not directly interact with the digital space. Within the diegetic space we find the physical space, a single (though potentially compartmentalised) tacit environment that separates co-located users from those accessing the system remotely. The physical space is a controlled environment in which the architectural and material properties can be built or adapted to suit the requirements of the system and control ambient light, sound and temperature. The hardware that provides access to the digital space (including the full array of XR control, display, and tracking systems) exists in both the physical space and to remote users, though the specific hardware provided may differ between

Table 1
Adapting emergent listening modes to emergent experience modes.

Mode	Emergent Sound (original definition)	Emergent Experience	Dominant location, form, and mode	Dominant system aggregate components
In search	Actively analysing the soundscape or scanning for a particular cue	Actively analysing the environment or scanning for a particular sensory cue	Distal; causal; spatial	Virtual environment; tasks; avatar
In readiness	Ready to respond to a sound cue but not actively scanning	Ready to respond to a cue but not actively scanning	Proximal; reflexive	tasks (focus away from virtual environment)
In background	Passive listening with some potential to recall aspects of soundscape	Keeping cues in perceptual periphery, able to recall some aspects of environment	Perceptual; connotative	Virtual environment (multi-focus/shifting focus)
Navigational	To use sound cues to localise oneself and navigate around a space	To use sensory cues to localise oneself and navigate around a space	Relational; spatial; temporal	Virtual environment; avatar; mechanics; user-interface;
Choroplast	To determine spatial properties of a virtual environment	To determine spatial properties of a virtual environment	Relational; spatial	Virtual environment; user-interface
Topoplast	Interpretation of real-world auditory qualities whilst in a virtual world	Interpretation of real-world physical qualities whilst in a virtual world	Non-specific/distributed	Virtual environment; mechanics (physics/lighting); data systems
Chronoplast	Using sound to determine temporal information relevant to a situation	Using environmental changes to determine temporal information in a situation	Relational; temporal	Virtual environment; data systems; user-interface
Aionoplast	Establishing temporal information relevant to setting (e.g., historical period)	Establishing temporal information relevant to setting (e.g., historical period)	Perceptual; semantic; architectural	Virtual environment; avatar; NPCs; roles; narrative
Theatre	Active interpretation of sound but no agency to interact directly	Active interpretation of environment but no agency to interact directly	Distal; medial; kinaesthetic; empathetic; semantic	Virtual environment; NPCs; narrative
Museum	Some agency to interact with sound within a controlled and fixed space	Some agency to interact with aspects of environment within a controlled and fixed space	Distal; causal; functional; reduced; critical	User-interface; NPCs; tasks; mechanics (limited); interaction (limited)
City	Greater agency to interact with sound within an uncontrolled space	Greater agency to interact with environment within an uncontrolled space	Relational; spatial; semantic; architectural	User-interface; NPCs; tasks; mechanics; interaction
Causal	To identify the sound source object and/or event	To identify an object or event from a sensory cue	Distal; causal	Virtual environment; NPCs; tasks; interaction
Semantic	To interpret discrete meaning (e.g., an instruction)	To interpret discrete meaning (e.g., an instruction) from a cue	Medial; relational; semantic	Instruction; NPCs; narrative; task
Reduced	To analyse the fundamental characteristics of the sound itself	To analyse the fundamental characteristics of the cue itself	Distal; reduced	Virtual environment; interaction
Reflexive	Pre-attentive bodily response (e.g., jump in response to sudden sound)	Pre-attentive bodily response (e.g., jump in response to sudden change)	Proximal; reflexive	Virtual environment; NPCs; task
Kinaesthetic	Pre-attentive sense of motion evoked by sound	Pre-attentive sense of motion evoked by environment/cue	Distal; proximal; kinaesthetic	Virtual environment; NPCs; task
Connotative	Free-form associations immediately associated with sound	Free-form associations immediately associated with cue	Perceptual; connotative	Virtual environment; NPCs; task
Empathetic	To infer aspects of the emotional state of the source	To infer aspects of the emotional state of the source	Distal; medial; empathetic	NPCs; roles; narrative; task
Functional	To interpret a sense of a sound's meaning/purpose/function	To interpret a sense of a cue's meaning/purpose/function	Relational; functional	NPCs; virtual environment; roles; narrative; task
Critical	To apply a value judgement to the quality/appropriateness of a sound	To apply a value judgement to the quality/appropriateness of a cue	Relational; critical	User-interface; virtual environment; task
Analytic	To analyse discrete properties of a sound within a focussed point in time	To analyse discrete properties of a cue within a focussed point in time	Distal; reduced; critical	User-interface; virtual environment; task; interaction
Synthetic	To analyse the general properties of a sound over a wider period of time	To analyse the general properties of a cue over a wider period of time	Distal; medial; reduced; critical	User-interface; virtual environment; task; interaction
Interactive	To interact with the source and/or environment then analyse the response	To interact with the source and/or environment then analyse the response	Distal; medial; reduced; critical	User-interface; virtual environment; task; interaction

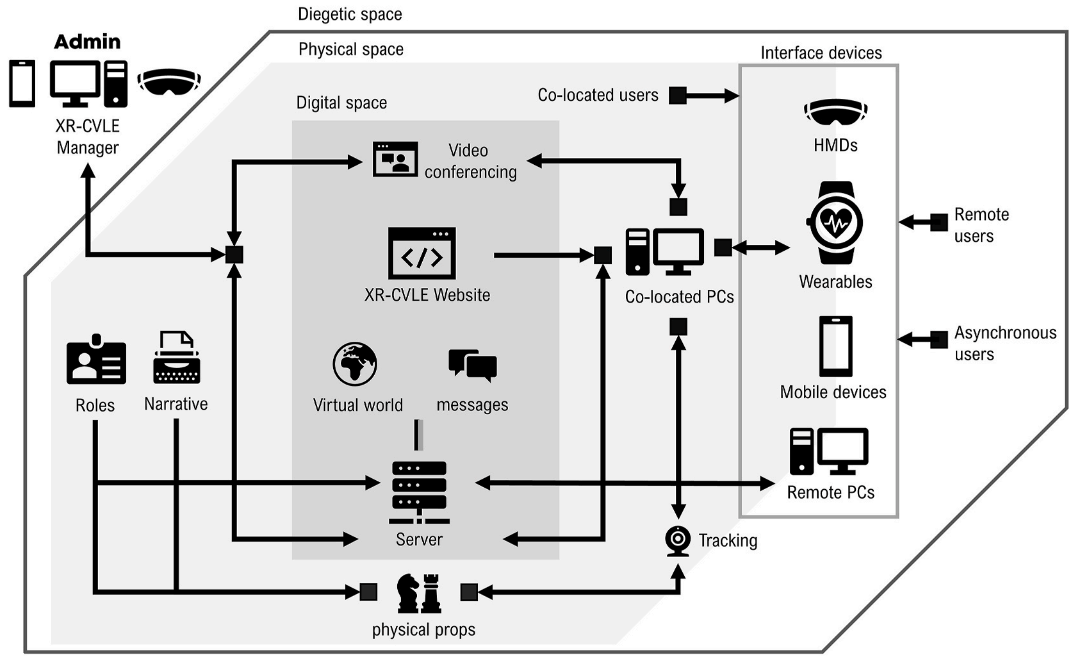

Fig. 4. Heterogenic XR-CVLE ecosystem.

the two. Of course, remote users are themselves situated within their own physical spaces, but the model separates them from the grouping of physical space as it is anticipated that meaningful control over the physical environment of remote users is unfeasible. Lastly, the digital (or virtual) space exists within the diegetic space and is accessible both from without and outside the physical space. By way of the XR hardware, the digital space presents the software aspects of the scenario, encapsulating the components of the virtual aggregate (virtual environment, avatar, non-player characters, etc.) and extending the diegetic aggregate (roles, narrative, instruction, etc.) into the digital domain to present a meaningfully mixed reality experience. Whilst the physiological aggregate cannot be forced, it can be heavily influenced using such a highly controlled ecosystem by presenting the user with tasks that dictate specific physical behaviours and that can have direct impact upon a user's position, orientation, and dynamic movements–with indirect potential to influence cardiovascular and respiratory activity).

6. Guidelines for applying the emergent experience model

Considering the above discussion, the following is a set of key practical principles to help guide XR designers and developers. The intention is to directly address the challenges referred to in the introduction (namely pedagogical grounding, technical configuration, embodiment and presence, changing interaction capabilities, and lack of user-centred design). Practical suggestions relevant to matters of cognitive overload, coherence, and the multilayered model of emergence are also presented below. With a focus largely upon auditory perception, it is acknowledged that matters of how different sensory modalities interact and integrate (and what impact this has upon learning outcomes) are a limitation of this work, and that subsequent study should seek to apply existing understanding of multisensory effects to the emergent experience model.

Because the emergent experience model is at a preliminary stage of development, recommendations relevant to its application are currently research-informed hypotheses and they describe the future research agenda for the Sheffield Extended Reality Audio laboratory, within the Industry and Innovation Research Institute at Sheffield Hallam university.

1. Consider all content and the users' experience of it as interactive. Engagement with any content, be it physical, cognitive, or affective, leads to emergent experience and is therefore an interaction between content and user. It is important to be mindful of precisely what form of interaction is desirable at any given moment and why this would contribute to deeper learning.
2. To address problems with pedagogical grounding, XR designers and developers must work closely with subject expert educators to ensure the learning objectives are fully understood. As a crucial part of this process, emergent experience should be mapped to learning objectives.
   a. It should be clear in the design precisely what emergent experience modes are being evoked and utilised at any given moment, with a rationale for this grounded in the learning objectives.
3. Difficulties with technical configuration can be reduced by questioning what hardware interfaces are necessary to evoke the desired emergent experiences and learning outcomes.
   a. It is tempting to assume more high-tech equipment will yield inherent benefit but the hardware itself is not the solution. The emergent experience is the solution, and the hardware offers a potential route to that solution. Designers and developers should always first define the requirements of the emergent experience, then work backwards to consider what is the most efficient technical configuration for delivering these requirements.
4. To manage challenges with limited embodiment and presence, aim to engage the user with as many modes of experience as possible.
   a. There is no hierarchy of engagement. For example, a reflexive experience should not be considered more superficial or reduced experience, more meaningful.
   b. Depth of experience is curated by utilising a full range of modes. Therefore, look for opportunities to encourage users to engage in different modes, staying mindful of the risks to cognitive overload.
   c. Content should evoke more reflexive and kinaesthetic modes of experience whilst encouraging users to develop a conscious

awareness of these modes and observe the links between bodily experience, cognition, and emotion.
   d. Biofeedback through wearable technologies can evoke embodied experience, both by way of direct feedback (e.g. display of user heart rate upon a heads-up display) or a relationship between biosensing data and systems within the digital environment (e.g. adaptive difficulty, where increased user stress is linked to more time to complete a given task).
5. To help overcome issues with constantly changing interaction capabilities, look to contemporary games and their established design systems and mechanics.
   a Be wary of creating a 'two-tier' user experience in which those familiar with video games will be able to perform significantly better than those without.
   b Designing more intuitive interactions can be a valid approach, but this is very difficult to achieve universally, as users bring a wide range of experience with different forms of hardware and software. Therefore, user-interactions should prioritise simplicity and ease of learning over intuitiveness, with sufficient 'onboarding' (tutorials or free exploration) prior to launching the main learning experience.
6. Be aware of the risks of cognitive overload when content requires attending to in multiple modes, and if this is task-relevant, time-pressured, or when the content is of the same sensory modality. One of these three features is low risk, two is moderate, and three is likely to be highly problematic. For example:
   a. A VR experience requires a user to navigate an unfamiliar space, survey the environment to identify cues indicative of a hazard–the cues themselves can be sudden and appear at unexpected locations relative to the user. Presenting this as a task under time-limitations would be very challenging. Allowing the user unlimited time to complete the task would reduce the risk, whilst also varying the modality (e.g. if auditory cues are the subject of analysis, then visual cues are used to aid navigation and reinforce relative location of hazards) would lower the risk more substantially.
7. Think longitudinally when designing an experience. The multi-layered and looping nature of emergence means that one moment will influence the next and so you must consider how the users' state may change in response to a moment and what effect that could have upon the next. Ideally, you are looking for opportunities to 'prime' a user with initial content to better-encourage them towards a desired state when they experience subsequent content. For example:
   a. Reflexive experience, evoked by (as an example) sudden and unexpected content, could help to prime 'in readiness' or 'in search' engagement.
   b. Clear navigational cues to encourage 'museum'-type engagement.
   c. Intentionally obscuring the identity of an agent producing cues to discourage causal engagement.
8. Be mindful of multimodal coherence. Consider your content, during both initial design and throughout implementation, to determine what modes a user is likely to use to interact with the content.
   a. For example, if an audio cue is intended to aid in navigation, is this 'encouraged mode' obvious? Does the cue in context present any conflicting information that could cause a user to misinterpret your intended function for the sound?
   b. Making sure concurrent cues between multiple sensory modalities evoke the same response can dramatically increase the clarity of user experience and direct users more reliably towards desired modes of engagement.
   c. Conversely, multimodal dissonance can also be useful when deployed with care, when the intention is to disorientate the user to make interactions within the environment more challenging.
9. Many of the emergent experience modes do themselves present opportunities for skill development and personal insight.
   a. Many users, unless directly trained to do so, can find (especially in an auditory perception context) modes such as reduced, analytical, critical, and in-background very difficult to do effectively, yet these can be valuable skills and can yield critical information about a situation.
   b. Understanding that such a range of modes exist and appreciating their capacity to help us interpret unique insights from an experience, can improve situational awareness, understanding of ourselves within the learning environment, and help us to communicate our experience more effectively to others.
10. Consider asymmetric multi-user engagement and diexis with relevance to modes of experience.
   a. This could involve each user being required to utilise a different mode (within a shared environment or across multiple environments) and to communicate their insights between the group to arrive at a shared understanding.
   b. Alternatively, each user could experience alternate versions of the same scenario, with cues crafted in different ways that encourage different modes of engagement.

The hope is that this work provides a convincing rationale for being mindful of the difference between factors of emergent experience that can be directly manipulated and those that cannot, and also that it encourages XR designers and developers to carefully consider matters of flow, interest, motivation, self-efficacy, autonomy, presence, and understanding as core qualities of any learning experience. Understanding the importance of the physiological and environmental factors within which a learning intervention and its users are situated is too often neglected. This is creating barriers to learning and missing opportunities to deepen learning even further. Using the emergent experience model, deeper learning can be engaged through multiple emergent routes, whilst using multiple routes within a single experience could be exponentially more powerful. As mentioned previously, the efficacy of the above guidelines remains largely hypothetical and require systematic experimentation to help quantify and qualify their effect upon deep learning–particularly within an XR context. Further progress in this topic requires deeper understanding of the configuration and orientation of the virtual, diegetic, environmental, physiological, and platform aggregates. As a future research agenda, we need to understand more precisely how the emergent effect of learning actualises, not only from the nature of each of the elements within a particular aggregate, but also how the inter-relationships between the elements collectively give rise to the effect.

CRediT authorship contribution statement

Tom A. Garner: Writing – review & editing, Writing – original draft, Project administration, Investigation, Conceptualization.

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.

References

Ahmad, F., & Karim, M. (2019). Impacts of knowledge sharing: A review and directions for future research. *Journal of Workplace Learning, 31*(3), 207–230.
Alahuhta, P., Nordbäck, E., Sivunen, A., & Surakka, T. (2014). Fostering team creativity in virtual worlds. *The Journal of Virtual Worlds Research, 7*, 1–22.

Allen, M., Gawad, N., Park, L., & Raiche, I. (2019). The educational role of autonomy in medical training: A scoping review. *Journal of Surgical Research, 240*, 1–16.
Baladi, M., Vitali, H., Fadel, G., Summers, J., & Duchowski, A. (2008). A taxonomy for the design and evaluation of networked virtual environments: Its application to collaborative design. *International Journal on Interactive Design and Manufacturing, 2*, 17–32.
Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*(2), 191.
Bao, J., Chen, D., Wen, F., Li, H., & Hua, G. (2017). CVAE-GAN: Fine-grained image generation through asymmetric training. In *Proceedings of the IEEE international conference on computer vision* (pp. 2745–2754).
Benford, S. (1993). A distributed architecture for large collaborative virtual environments. In *IEE colloquium on distributed virtual reality* (p. 9). IET, 1.
Benford, S., & Fahlén, L. (1993). A spatial model of interaction in large virtual environments. In *Proceedings of the third European conference on computer-supported cooperative work 13–17 september 1993, milan, Italy ECSCW'93* (pp. 109–124). Dordrecht: Springer. Netherlands.
Bijsterveld, K. (2019). Sonic skills: Listening for knowledge in science. *Medicine and engineering (1920s-Present)*. USA: Springer Nature.
Bittner, E., & Leimeister, J. (2014). Creating shared understanding in heterogeneous work groups: Why it matters and how to achieve it. *Journal of Management Information Systems, 31*, 111–144.
Bray, B., & McClaskey, K. (2016). *How to personalize learning: A practical guide for getting started and going deeper*. Corwin Press.
Brown, S. R., Feldman, E. R., Cross, M. R., Helms, E. R., Marrier, B., Samozino, P., & Morin, J. B. (2017). The potential for a targeted strength-training program to decrease asymmetry and increase performance: A proof of concept in sprinting. *International Journal of Sports Physiology and Performance, 12*(10), 1392–1395.
Bullinger-Hoffmann, A., Koch, M., Möslein, K., & Richter, A. (2021). Computer-supported cooperative work–revisited. *I-Com, 20*(3), 215–228.
Burova, A., Opas, V., Mäkelä, J., Hakulinen, J., Lindqvist, T., Siltanen, S., … Turunen, M. (2022). Enhancing remote industrial training experiences with asymmetric virtual reality: Experiences, tools and guidelines. *Applied Sciences, 12*(15), 7745.
Chaplin, W. H. (1973). Identity and spirit in the recent poetry of john logan. *American Poetry Review, 2*(3), 19–24.
Chen, Y., & Wang, C. (2018). Learner presence, perception, and learning achievements in augmented–reality–mediated learning environments. *Interactive Learning Environments, 26*, 695–708.
Chion, M. (2012). The three listening modes. *The sound studies reader* (pp. 48–53).
Clergeaud, D., Roo, J., Hachet, M., & Guitton, P. (2017). Towards seamless interaction between physical and virtual locations for asymmetric collaboration. *Proceedings of the 23rd ACM symposium on virtual reality software and technology*.
Cummings, J., & Bailenson, J. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology, 19*, 272–309.
De Back, T. T., Tinga, A. M., & Louwerse, M. M. (2023). Learning in immersed collaborative virtual environments: Design and implementation. *Interactive Learning Environments, 31*(8), 5364–5382.
Delaney, D., Ward, T., & McLoone, S. (2006). On consistency and network latency in distributed interactive applications: A survey—Part I. *Presence, 15*(2), 218–234.
Dillenbourg, P., & Fischer, F. (2007). Computer-supported collaborative learning: The basics. *Zeitschrift für Berufs-und Wirtschaftspädagogik, 21*, 111–130.
Dix, A. (1994). Computer supported cooperative work: A framework. In *Design issues in CSCW* (pp. 9–26). London: Springer. London.
Duijzer, C., Heuvel-Panhuizen, M., Veldhuis, M., Doorman, M., & Leseman, P. (2019). Embodied learning environments for graphing motion: A systematic literature review. *Educational Psychology Review*, 1–33.
Faiola, A., Newlon, C., Pfaff, M., & Smyslova, O. (2013). Correlating the effects of flow and telepresence in virtual worlds: Enhancing our understanding of user behavior in game-based learning. *Computers in Human Behavior, 29*, 1113–1121.
Fakomogbon, M., & Bolaji, H. (2017). Effects of collaborative learning styles on performance of students in a ubiquitous collaborative mobile learning environment. *Contemporary Educational Technology, 8*, 268–279.
Fierro-Saltos, W., Sanz, C., Zangara, A., Guevara, C., Arias-Flores, H., Salazar, D., Aldás, J., Borja-Galeas, C., Rivera, R., Hidalgo-Guijarro, J., & Yandún-Velasteguí, M. (2019). *Autonomous learning mediated by digital technology processes in higher education: A systematic review* (pp. 65–71).
Follett, M. D., & Urwick, L. F. (1949). Freedom & Co-ordination. Lectures in business organisation. In L. Urwick (Ed.), *Edited, with an introduction*. Management Publications Trust.
Garner, T., & Grimshaw, M. (2011). A climate of fear: Considerations for designing a virtual acoustic ecology of fear. In *Proceedings of the 6th audio mostly conference: A conference on interaction with sound* (pp. 31–38).
Garrison, D. R., Anderson, T., & Archer, W. (2010). The first decade of the community of inquiry framework: A retrospective. *The Internet and Higher Education, 13*(1–2), 5–9.
Garzón, J., Pavón, J., & Baldiris, S. (2019). Systematic review and meta-analysis of augmented reality in educational settings. *Virtual Reality, 23*, 447–459.
Gegenfurtner, A., Quesada-Pallarés, C., & Knogler, M. (2014). Digital simulation-based training: A meta-analysis. *British Journal of Educational Technology, 45*, 1097–1114.
Gerard, B. (2011). Taxonomy revisited. The hitchhiker's guide to computer science. *Blog*. https://bartgerard.wordpress.com/2011/10/. (Accessed 26 January 2024).
Gregg, K. (2003). The state of emergentism in second language acquisition. *Second Language Research, 19*, 128–195.
Grief, I. (1988). *Computer-supported cooperative work: A book of readings*. UK: Elsevier Science.
Grimshaw, M. N. (2007). The acoustic ecology of the first-person shooter. In *(Doctoral dissertation*. New Zealand: The University of Waikato).
Grimshaw, M., & Garner, T. A. (2015). *Sonic virtuality: Sound as emergent perception*. USA: Oxford University Press.
Grudin, J., & Poltrock, S. (2012). Taxonomy and theory in computer supported cooperative work. In S. W. J. Kozlowski (Ed.), *The oxford handbook of organizational psychology* (Vol. 2). UK: Oxford Academic. Oxford Library of Psychology.
Grudin, J., & Poltrock, S. (2014). Computer supported cooperative work. *Interaction Design Foundation*. Online article: https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/computer-supported-cooperative-work. (Accessed 1 February 2024).
Hamilton, D., McKechnie, J., Edgerton, E., & Wilson, C. (2021). Immersive virtual reality as a pedagogical tool in education: A systematic literature review of quantitative learning outcomes and experimental design. *Journal of Computers in Education, 8*(1), 1–32.
Hammad, R., Khan, Z., Safieddine, F., & Ahmed, A. (2020). A review of learning theories and models underpinning technology-enhanced learning artefacts. *World Journal of Science Technology and Sustainable Development, 17*(4), 341–354.
Hatami, A. (2015). The effect of collaborative learning and self-assessment on self-regulation. *Educational Research Review, 10*, 2164–2167.
Hernández, Y., Regalado, O., Rosa, M., Cueva, C., & Muro, J. (2021). Autonomous learning in the educational context: A psychological perspective. *ASEAN Journal of Psychiatry, 22*(9).
Hodges, C. (2008). Self-efficacy in the context of online learning environments. *Performance Improvement Quarterly, 20*.
Hoffman, D. D., Singh, M., & Prakash, C. (2015). The interface theory of perception. *Psychonomic Bulletin & Review, 22*, 1480–1506.
Hsu, J., & Chou, H. (2009). The effects of communicative genres on intra-group conflict in virtual student teams. *International Journal of Distance Education Technologies, 7*, 1–22.
Hudák, M., Korečko, Š., & Sobota, B. (2020). Advanced user interaction for web-based collaborative virtual reality. In *2020 11th IEEE international conference on cognitive infocommunications (CogInfoCom)* (pp. 343–348). IEEE.
Hwang, W. Y., & Su, J. H. (2012). The study of surface computer supported cooperative work and its design, efficiency, and challenges. *Interactive Learning Environments, 20*(2), 177–198.
Jackman, P. C., Dargue, E. J., Johnston, J. P., & Hawkins, R. M. (2021). Flow in youth sport, physical activity, and physical education: A systematic review. *Psychology of Sport and Exercise, 53*, Article 101852.
Järvinen, A. (2017). Design for presence in VR, Part 2: Towards an applied model. *Virtual Reality Pop*. Online article: https://virtualrealitypop.com/designing-for-presence-in-vr-part-2-towards-an-applied-model-2784bf16a01. (Accessed 28 January 2024).
Kosmas, P., Ioannou, A., & Zaphiris, P. (2018). Implementing embodied learning in the classroom: Effects on children's memory and language skills. *Educational Media International, 56*, 59–74.
Kostov, G., & Wolfartsberger, J. (2022). Designing a framework for collaborative mixed reality training. *Procedia Computer Science, 200*, 896–903.
Kubo, M., Tori, R., & Kirner, C. (2002). Interaction in collaborative educational virtual environments. *CyberPsychology and Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society, 5*(5), 399–407.
Kuleto, V., P, M. I., Stanescu, M., Ranković, M., Šević, N. P., Păun, D., & Teodorescu, S. (2021). Extended reality in higher education, a responsible innovation approach for generation y and generation z. *Sustainability, 13*(21), Article 11814.
Kwon, J. U., Hwang, J. I., Park, J., & Ahn, S. C. (2019). Fully asymmetric remote collaboration system. *IEEE Access, 7*, 54155–54166.
Lee, D., Lee, M., Lee, K., & Song, C. (2014). Asymmetric training using virtual reality reflection equipment and the enhancement of upper limb function in stroke patients: A randomized controlled trial. *Journal of Stroke and Cerebrovascular Diseases, 23*(6), 1319–1326.
Lege, R., & Bonner, E. (2020). Virtual reality in education: The promise, progress, and challenge. *Jalt Call Journal, 16*(3), 167–180.
Lentz, F., Shaffer, A., Pratt, D. R., Falby, J. S., & Zyda, M. J. (1995). Npsnet: Naval training integration. *In the proceedings of the 13th DIS workshop* (pp. 18–22).
Lewes, G. H. (1877). *Problems of Life and mind*. London, UK: Trubner & Co.
Lowyck, J., & Pöysä, J. (2001). Design of collaborative learning environments. *Computers in Human Behavior, 17*(5–6), 507–516.
Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (camil): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review*, 1–22.
Makransky, G., Terkildsen, T. S., & Mayer, R. E. (2019). Adding immersive virtual reality to a science lab simulation causes more presence but less learning. *Learning and Instruction, 60*, 225–236.
Maor, D., & Volet, S. (2007). Interactivity in professional online learning: A review of research based studies. *Australasian Journal of Educational Technology, 23*, 269–290.
Marques, B., Silva, S., Alves, J., Araujo, T., Dias, P., & Santos, B. S. (2021). A conceptual model and taxonomy for collaborative augmented reality. *IEEE Transactions on Visualization and Computer Graphics, 28*(12), 5113–5133.
Mayer, R. E. (2014). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 43–71). Cambridge University Press.
Mayer, A., Kilinc, I., Sprügel, K., & Häfner, P. (2023). Towards research gaps in collaborative virtual reality environments for education: A literature review. In *International conference on remote engineering and virtual instrumentation* (pp. 513–525). Cham: Springer.
McClelland, D. C. (2010). Assessing individual differences in achievement motivation with the implicit association test: Predictive validity of a chronometric measure of the self-concept ''Me= successful''. *Implicit motives, 151*.
McClelland, E., Pitt, A., & Stein, J. (2015). Enhanced academic performance using a novel classroom physical activity intervention to increase awareness, attention and

self-control: Putting embodied cognition into practice. *Improving Schools, 18*, 100–183.
McGrath, J. E. (1984). A typology of tasks. *Group: Interaction and performance*, 53–66.
McGraw, G. E. (1995). *Letter Spirit (part one): Emergent high-level perception of letters using fluid concepts*. USA: Indiana University.
Meyer, O. A., Omdahl, M. K., & Makransky, G. (2019). Investigating the effect of pre-training when learning through immersive virtual reality and video: A media and methods experiment. *Computers & Education, 140*, Article 103603.
Mikropoulos, T. (2006). Presence: A unique characteristic in educational virtual environments. *Virtual Reality, 10*, 197–206.
Miller, D. C., & Thorpe, J. A. (1995). Simnet: The advent of simulator networking. *Proceedings of the IEEE, 83*(8), 1114–1123.
Morrison-Smith, S., & Ruiz, J. (2020). Challenges and barriers in virtual teams: A literature review. *SN Applied Sciences, 2*, 1–33.
Msonde, S., & Aalst, J. (2017). Designing for interaction, thinking and academic achievement in a Tanzanian undergraduate chemistry course. *Educational Technology Research & Development, 65*, 1389–1413.
Muller, D., Lee, K., & Sharma, M. (2008). Coherence or interest: Which is most important in online multimedia learning? *Australasian Journal of Educational Technology, 24*, 211–221.
Muma, J. R. (1991). Experiential realism: Clinical implications. In *Pragmatics of language: Clinical practice issues* (pp. 229–247). Boston, MA: Springer US.
Mystakidis, S., Berki, E., & Valtanen, J. P. (2021). Deep and meaningful e-learning with social virtual reality environments in higher education: A systematic literature review. *Applied Sciences, 11*(5), 2412.
Nguyen, H., & Bednarz, T. (2020). User experience in collaborative extended reality: Overview study. In *, Vol. 17. Virtual reality and augmented reality: 17th EuroVR international conference, EuroVR 2020, valencia, Spain, november 25–27, 2020* (pp. 41–70). Springer International Publishing. Proceedings.
Nii, A. T., & Yunus, M. M. (2022). Teachers' factors in autonomous learning during language learning: A systematic literature review. *International Journal of Academic Research in Business and Social Sciences, 12*(5), 1408–1427.
Núñez, R., Edwards, L., & Matos, J. (1999). Embodied cognition as grounding for situatedness and context in mathematics education. *Educational Studies in Mathematics, 39*, 45–65.
Oļesika, A., & Rubene, Z. (2023). *Professional autonomy as a cornerstone for effective professional and social activity* (pp. 319–329). To Be or Not to Be a Great Educator.
Onen, M., Gokmen, T., Todorov, T. K., Nowicki, T., Del Alamo, J. A., Rozen, J., … Kim, S. (2022). Neural network training with asymmetric crosspoint elements. *Frontiers in Artificial Intelligence, 5*, Article 891624.
Ouverson, K., & Gilbert, S. (2021). A composite framework of Co-located asymmetric virtual reality. *Proceedings of the ACM on Human-Computer Interaction, 5*, 1–20.
Pan, X., & Hamilton, A. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology, 109*, 395–417.
Parmaxi, A. (2020). Virtual reality in language learning: A systematic review and implications for research and practice. *Interactive Learning Environments, 31*, 172–184.
Pellas, N., Kazanidis, I., & Palaigeorgiou, G. (2019). A systematic literature review of mixed reality environments in K-12 education. *Education and Information Technologies, 25*, 2481–2520.
Penalo, L., & San, E. (2020). Potential influences of virtual simulation prebriefing and debriefing on learners' self-efficacy. *Nurse Educator, 46*, 195–197.
Penner, D. (2000). Explaining systems: Investigating middle school students' understanding of emergent phenomena. *Journal of Research in Science Teaching, 37*, 784–806.
Pereira, A. M. F., Fernandes, S., Bittencourt, I. I., & Félix, A. (2022). Flow theory and learning in the Brazilian context: A systematic literature review. *Educação e Pesquisa, 48*, Article e237870.
Pereira, V., Matos, T., Rodrigues, R., Nóbrega, R., & Jacob, J. (2019). Extended reality framework for remote collaborative interactions in virtual environments. In *2019 international conference on graphics and interaction (ICGI)* (pp. 17–24). IEEE.
Perttula, A., Kiili, K., Lindstedt, A., & Tuomi, P. (2017). Flow experience in game based learning - a systematic literature review. *Int. J. Serious Games, 4*.
Plopski, A., Hirzle, T., Norouzi, N., Qian, L., Bruder, G., & Langlotz, T. (2022). The eye in extended reality: A survey on gaze interaction and eye tracking in head-worn extended reality. *ACM Computing Surveys, 55*(3), 1–39.
Pursio, K., Kankkunen, P., Sanner-Stiehr, E., & Kvist, T. (2021). Professional autonomy in nursing: An integrative review. *Journal of Nursing Management, 29*(6), 1565–1577.
Raupach, T., Muenscher, C., Anders, S., Steinbach, R., Pukrop, T., Hege, I., & Tullius, M. (2009). Web-based collaborative training of clinical reasoning: A randomized trial. *Medical Teacher, 31*, 431–437.
Rebelo, P., Green, M., & Hollerweger, F. (2008). A typology for listening in place. In *Proceedings of the 5th international mobile music workshop* (pp. 15–18).
Reisoglu, I., Topu, B., Yilmaz, R., Yılmaz, T., & Göktas, Y. (2017). 3D virtual learning environments in education: A meta-review. *Asia Pacific Education Review, 18*, 81–100.
Richardson, J., Maeda, Y., Lv, J., & Caskurlu, S. (2017). Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis. *Computers in Human Behavior, 71*, 402–417.
Riva, G. (2009). Is presence a technology issue? Some insights from cognitive sciences. *Virtual Reality, 13*, 159–169.
Rodríguez-Ardura, I., & Meseguer-Artola, A. (2017). Flow in e-learning: What drives it and why it matters. *British Journal of Educational Technology, 48*, 899–915.
Ryan, G., Rafferty, A., Murphy, J., Higgins, M. F., Mangina, E., & McAuliffe, F. M. (2023). Virtual reality learning: A randomized controlled trial assessing medical student knowledge of fetal development. *International Journal of Gynecology & Obstetrics, 162*(1), 292–299.
Salehzadeh, R., Gong, J., & Jalili, N. (2022). Purposeful communication in human–robot collaboration: A review of modern approaches in manufacturing. *IEEE Access, 10*, 129344–129361.
San Chee, Y. (2001). Networked virtual environments for collaborative learning. In *Proceedings of ICCE/SchoolNet 2001Ninth international conference on computers in education* (Vol. 311). Seoul, S. Korea.
Schäfer, A., Reis, G., & Stricker, D. (2022). A survey on synchronous augmented, virtual, and mixed reality remote collaboration systems. *ACM Computing Surveys, 55*(6), 1–27.
Schaffer, S., Chen, X., Zhu, X., & Oakes, W. (2012). Self-efficacy for cross-disciplinary learning in project-based teams. *Journal of Engineering Education, 101*.
Segedy, J. R., Kinnebrew, J. S., & Biswas, G. (2015). Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal of Learning Analytics, 2*(1), 13–48.
Sensuse, D. I., Lestari, P. I., & Al Hakim, S. (2021). Exploring factors influencing knowledge sharing mechanisms and technology to support the collaboration ecosystem: A review. *DESIDOC Journal of Library & Information Technology, 41*(3).
Simpson, Z. (2011). Merleau-Ponty and emergent perception. *Journal of the British Society for Phenomenology, 42*(3), 290–304.
Sims, R. (1994). Creative multimedia design: Virtual learning environments for multiplatform applications. In *Proceedings of the second international multimedia symposium* (pp. 504–509).
Skarbez, R., Brooks, J., F, P., & Whitton, M. C. (2017). A survey of presence and related concepts. *ACM Computing Surveys, 50*(6), 1–39.
Smeby, J., & Heggen, K. (2014). Coherence and the development of professional knowledge and skills. *Journal of Education and Work, 27*, 71–91.
Storms, R. L. (1995). NPSNET-3D sound server: An effective use of the auditory channel. In *(Doctoral dissertation*. Monterey, California. Naval Postgraduate School.
Su, Y. L., & Reeve, J. (2011). A meta-analysis of the effectiveness of intervention programs designed to support autonomy. *Educational Psychology Review, 23*, 159–188.
Sugiura, Y., Ibayashi, H., Chong, T., Sakamoto, D., Miyata, N., Tada, M., Okuma, T., Kurata, T., Shinmura, T., Mochimaru, M., & Igarashi, T. (2018). An asymmetric collaborative system for architectural-scale space design. *Proceedings of the 16th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry* (pp. 1–6).
Tachihara, K., & Goldberg, A. (2019). Emergentism in neuroscience and beyond. *Journal of Neurolinguistics, 49*, 237–239.
Truax, B. (1984). *Acoustic communication*. USA: Praeger.
Tsiatsos, T., Andreas, K., & Pomportsis, A. (2010). Evaluation framework for collaborative educational virtual environments. *Journal of Educational Technology & Society, 13*(2), 65–77.
Tuuri, K., & Eerola, T. (2012). Formulating a revised taxonomy for modes of listening. *Journal of New Music Research, 41*(2), 137–152.
Van Orden, G. C., & Goldinger, S. D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human Perception and Performance, 20*(6), 1269.
Wang, H., Duh, H., Li, N., Lin, T., & Tsai, C. (2014). An investigation of university students' collaborative inquiry learning behaviors in an augmented reality simulation and a traditional simulation. *Journal of Science Education and Technology, 23*, 682–691.
Wang, P., Wang, Y., Billinghurst, M., Yang, H., Xu, P., & Li, Y. (2023). *BeHere: A VR/SAR remote collaboration system based on virtual replicas sharing gesture and avatar in a procedural task* (pp. 1–22). Virtual Reality.
Wäschle, K., Allgaier, A., Lachner, A., Fink, S., & Nückles, M. (2014). Procrastination and self-efficacy: Tracing vicious and virtuous circles in self-regulated learning. *Learning and Instruction, 29*, 103–114.
Weinberger, A., Stegmann, K., & Fischer, F. (2007). Knowledge convergence in collaborative learning: Concepts and assessment. *Learning and Instruction, 17*, 416–426.
Wiggs, C. L., & Martin, A. (1998). Properties and mechanisms of perceptual priming. *Current Opinion in Neurobiology, 8*(2), 227–233.
Wilson, B. G. (1996). Constructivist learning environments: Case studies in instructional design. *Educational Technology*. NJ, USA.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review, 9*, 625–636.
Wong, M. D., Quartz, K. H., Saunders, M., Meza, B. P., Childress, S., Seeman, T. E., & Dudovitz, R. N. (2022). Turning vicious cycles into virtuous ones: The potential for schools to improve the life course. *Pediatrics, 149*(Supplement 5).
Yang, F., & Shen, F. (2019). Involvement without knowledge gain: A meta-analysis of the cognitive effects of website interactivity. *Journal of Broadcasting & Electronic Media, 63*, 211–230.
Zimmerman, B. (2000). Self-Efficacy: An essential motive to learn. *Contemporary Educational Psychology, 25*(1), 82–91.