TRYING TO KEEP UP: MANAGING COGNITIVE LOAD IN AUDIO SCORES

James Saunders
Open Scores Lab, Bath Spa University
j.saunders@bathspa.ac.uk

ABSTRACT

This paper examines how cognitive load theory can inform the design of audio scores that deliberately create unstable performance situations. Audio scores present temporal information that may be unknown in advance and which constantly disappears, creating unique cognitive demands compared to static visual notation. Drawing on cognitive load theory from instructional design, I identify how intrinsic, extraneous, and germane cognitive load—alongside specific effects including split-attention, modality, transient information, and task switching—shape performer responses to dense streams of verbal instruction.

Unlike instructional design, which aims to reduce extraneous cognitive load to optimize learning, my compositional approach deliberately manipulates these parameters to create productive difficulty. Through discussion of two recent pieces I consider how varying the density, simultaneity, and sequentiality of verbal cues generates distinct performance conditions. The discussion explores how player control over pacing fundamentally alters the way cognitive load manifests, shifting from individual overwhelm to distributed decision-making. For composers working with audio scores, this suggests practical strategies: manipulating verbal instruction parameters provides direct control over extraneous load while keeping individual task complexity low, and incorporating repeated instructional patterns helps players develop response schema.

1. INTRODUCTION

Audio scores present complex information in time, requiring users to manage their responses in an environment where this information may not be known in advance and may constantly disappear. The flow of information may not be within the control of the player, instead being managed by the speed and density of delivery of information as an external audio stream. Whereas in static visual scores the stream of notational information that is needed is mostly available in advance, with studies [1] showing the way a musician 'only rarely fixates a bar ahead of what is currently being played' [1, p. 18], in audio scores this is not generally the case. This categorical difference presents an opportunity to play with expectation and player responses to produce a more reflexive performance situation. In particular, the cognitive load placed on players when faced with a dense stream of instructions creates contingent situations where a host of strategies driven by heuristics might be employed to encourage certain musical and behavioural responses. The scenarios might engage with notions of failure, empathy and pragmatism which have the potential to reference situations we experience in everyday life, making connections with our lived experience and framing this in a playful and reflective manner.

In this paper I focus on *cognitive load theory* to understand how scores might be designed to manage user responses effectively. I begin by surveying recent scholarship from the field, isolating key aspects that might translate to music composition before presenting two examples of my own work that deploy these ideas to create unstable performance scenarios.

2. COGNITIVE LOAD THEORY AND AUDIO SCORES

The delivery of instructional information in an audio score can be thought of, in part, as a learning and comprehension task. The information is presented in a particular way with respect to its density, speed of delivery, simultaneity and sequentiality [2]. The configuration of these elements creates the score's different affordances. Users must be able to decode and understand this information in time in order to carry out the score's instructions. As such we might understand more about the way audio scores can be structured for preferred outcomes by looking at theory related to information learning and instructional design, specifically *cognitive load theory* and its framing of multimodal information delivery.

2.1. Cognitive Load Theory

Cognitive load theory is one of the principal frameworks for instructional design, which 'deals with learning and problem-solving difficulty that is artificial in that it can be manipulated by instructional design' [3, p. 295]. It is useful in relation to audio scores because it simultaneously considers 'the structure of information and the cognitive architecture that allows learners to process that information' [4, p. 1], taking into account the way in which information is presented to a user as well as the inherent nature of the information itself as variables that impact understanding.

2.1.1. Element Interactivity and Intrinsic Cognitive Load

Cognitive load is driven by *element interactivity* [3], [4], which determines the way material can be learned, either sequentially or simultaneously. Sweller notes that 'if elements can be learned successively rather than simultaneously because they do not interact, intrinsic cognitive load will be low' [3, p. 295]. For example, learning the name of a single person in a group has a low level of element interactivity as the person-name pair does not have any further reliance on other information. Although the task may be made harder by being required to memorize a lot of names, each of these is a single instance within a sequential set of tasks as the name of one person is not reliant on the name of another. In contrast, if simultaneous 'interactions between many elements must be learned, then intrinsic cognitive load will be high' [3, p. 295]. For example, learning the correct set of moves to solve the Rubik's Cube has high element interactivity as learning the result of a single move, while useful as a basic operation, is not sufficient to solve the complete cube which relies on being able to connect complex sequences of moves to achieve the desired result.

Element interactivity is therefore a defining factor of *intrinsic cognitive load*, which is 'concerned with the natural complexity of information that must be understood and material that must be learned' [5, p. 124]. In essence, some material is hard to learn and creates a high cognitive load due to its complex element interactivity, correspondingly making higher demands on working memory capacity. For audio scores, which may require players to recall complex information in real time while undertaking other actions, element interactivity and intrinsic cognitive loads are key indicators of the difficulty of the task.

2.1.2. Extraneous Cognitive Load

In contrast, *extraneous cognitive load* is imposed on this material by the instructional design. In the previous person-name example, learning one pair at a time in isolation at a comfortable rate has a low extrinsic load as each instance can be reinforced separately as an aid to retention. Consider though if the learning method was in a group of 20 people where each person gives their name in sequence once with the same requirement to learn person-name pairs. This would be a much harder task given the way the learning situation is presented, despite the task of remembering person-name pairs being the same. The intrinsic load is low, but the extrinsic load is much higher due to the instructional design.

Crucially, extraneous load is something that can be mitigated by careful instructional design, which is why it is of significance to making audio scores. As Sweller suggests in his original paper on cognitive load theory, 'extraneous cognitive load ... is artificial because it is imposed by instructional methods' while intrinsic cognitive load is something 'over which instructors have no control' [3, p. 307]. For audio scores, careful management of the instructional design is within the control of the score-maker, enabling aspects such as density, speed of delivery, simultaneity and sequentiality to impact appropriately on player responses.

2.1.3. Germane Cognitive Load

A third category, *germane cognitive load,* is also a significant and controllable factor in instructional design, but one which enhances learning through devoting working memory resources 'to schema acquisition and automation' [4, p. 2]. Recent developments in the definition of cognitive load theory have repositioned germane cognitive load as 'the cognitive load required to learn.' Specifically germane cognitive load 'redistributes working memory resources from extraneous activities to activities directly relevant to learning by dealing with information intrinsic to the learning task' [6, p. 264]. It helps learners focus on the intrinsic elements of the task by recalling and developing new schema to apply to the task. For example, when learning to solve the Rubik's Cube a network of heuristic observation and move techniques can form a schema to speed up user response to recognizable and previously experienced types of configuration. Instructional design can therefore leverage germane cognitive load to improve learning in future tasks. In audio scores, this might be through the use of common formulations of instructions which help players recall previous responses and strategies in an efficient manner, rather than treating each as novel information to process.

2.1. Cognitive Load Effects

Beyond the foundational categories of intrinsic, extrinsic and germane cognitive load and their impact on working memory, cognitive load research has isolated several distinct effects that instructional design creates [6]. Three of these—*split-attention effect*, *modality effect* and *transient information effect*—are of particular relevance to audio score design.

2.1.1. Split-attention Effect and Modality Effect

When learners are 'required to divide their attention among and mentally integrate multiple sources of information' [7, p. 319] split-attention effect can occur. For example, when looking at a diagram with spatially separate text explanations, there is a higher cognitive load on working memory as the learner must cross-relate and integrate two sets of information and determine their relationship to understand the material. If the same textual information is presented on the diagram as annotation, split attention effect is reduced as all the information is in the same place. As a parallel, consider the use of performance note prefaces in scores needing a lot of explanation and the impact on performer understanding when learning a new piece.

Split-attention effect is more prevalent when all the information is presented in the same domain, such as visual information (image-text, or text-text). Mousavi, Low and Sweller showed that when learners 'must split their attention between multiple sources of information that require mental integration, cognitive resources available for learning can be increased by presenting some of the verbal material in auditory rather than written form' [7, p. 333]. This *modality effect* suggests that presenting information in multimodal visual-auditory channels improves learning given the 'existence of separate working memory channels or processors for auditory (especially verbal) and visual (especially pictorial) information' [7, p. 319]. For audio scores, presenting some material in a visual mode may therefore reduce the load on working memory and improve the effectiveness of player responses.

2.2.2. Transient Information Effect

Given the temporal nature of audio scores, the *transient information effect* is highly relevant as it is concerned with 'information that is presented to learners but disappears after a few seconds, for example, in spoken text or in instructional video or animation' [6, p. 277]. This contrasts with non-transient information, such as a written text with pictures where 'all information is available to the learner at the same time and may be revisited when needed' [6, p. 277]. Having to store this disappearing information in working memory correspondingly increases the cognitive load, making the task harder to accomplish. Strategies to mitigate this problem include *self-pacing* where learners can manage the temporal flow of information, or *segmentation* where temporal information is presented in blocks with pauses in between [6, p. 278]. For audio scores, transient information is a fundamental reality but giving players opportunities to control the speed and density of instructions, or presenting breaks for them to settle, may also improve response. This may be contextually sensitive, however. Sweller, Van Merriënboer and Paas cite research by Leahy and Sweller which suggests 'short pieces of audio-visual information were more effective than visual information only (i.e. traditional modality effect), but longer pieces of audio-visual information were less effective than visual information only because of the abundance of transient information in the longer, auditory piece' [6, p. 278].

2.2. Task switching

Cognitive load can also be heightened by *task switching*, where learners must switch between different task types in logical, predictable, or randomized ways depending on the context. Where a task sequence requires learners to change what they are doing an interruption can occur, such that 'it costs more, typically measured in reaction time, to switch between task sets than to repeat a task of the same type' [8, p. 666]. This finding has been termed *switch costs*. As Kiesel et al. suggest, in 'a task-cuing paradigm with unpredictable sequences ... the order of the tasks and thus the order of task switches and repetitions are random' resulting in performance that is 'typically worse in switch trials than in repetition trials, revealing robust switch costs also in the task-cuing paradigm' [9, p. 852]. This may be particularly noticeable in generative scores where a novel sequence of cues is generated for each version of the piece. Interestingly, where there is similarity between cues which may still produce different results, such as the verbal cue 'g' referring to a pitch, a major or minor chord, or a fragment of music, 'switching between task sets with common elements such as shared stimuli or responses also imposes greater costs' [8, p. 666].

2.3. Emotion, stress and uncertainty

Beyond the inherent configuration of cognitive load in a set of tasks, the presentation context for audio scores also impacts on performance. A further emergent effect on the ability of a player to respond to cues is caused by emotion, stress and uncertainty, which Sweller, Van Merriënboer and Paas, citing Moran, suggest 'may restrict the capacity of working memory by competing with task-relevant processes; thus, they increase cognitive load, hamper learning and decrease transfer' [6, p. 285]. In emotionally heightened and stressful states, such as performing complex tasks in front of an audience with a high and visible prospect of failure, it is likely that uncertainty will lead to errors. They note, however, that this is an integral part of some activities, such that learners must develop 'professional competencies enabling them to perform professional tasks up to the standards, including the ability to deal with emotions, stress and uncertainty and to maintain overall wellbeing' [6, p. 285]. Rehearsal and conditioning naturally mitigate this effect, but there remains an uncertainty where the flow of information to which players respond is outside their control.

In combination, the three categories of cognitive load theory and its emergent effects have significance for audio score design given the need for users to hold complex multimodal information in their working memory and respond in real time. To evidence this I present examples from two of my pieces where management of cognitive load has been part of the compositional design.

3. WORKING WITH COGNITIVE LOAD

In my recent work I have been exploring different ways to manage cognitive load. In instructional theory, much of the narrative is around clarifying processes to remove extrinsic cognitive load so that learners can focus on the salient features of the task and improve learning. My interest partly aligns with this, at least regarding understanding the task, but is more concerned with the way cognitive load can create ambiguity, error and the unexpected. In doing so, my hope is that the experience of the player undergoing the task creates an empathic response for the audience, generally in relation to the way we experience overload in everyday life.

To do this, my recent pieces explore group decision-making and individual player agency in networks of interaction, thinking about how power structures emerge from groups working on a joint task. These pieces present players with many conflicting choices to create unstable dynamic situations that require constant attention. Individual preferences can be manifested but are always mitigated by the decisions of others in the group. These decisions are signalled orally, with players or artificial voices giving spoken cues that are also audible to the audience, making the decisions and their consequences explicit. The dynamically-produced audio score that emerges requires players to process different types of cue to complete the tasks as best they can, which can create different levels of cognitive load.

3.1. just trying to do what I'm told and keep up and do the best that I can (2025)

In *just trying to do what I'm told and keep up and do the best that I can* (2025) [10], a pianist responds to cues given by artificial voices, controlled by a bespoke app, attempting to follow all the instructions as they are given, as well as interacting verbally in conversation with an AI voice. In places the density of instructions may be too great to manage, but the aim is to try to follow as many cues as possible and adopt strategies to cope where this becomes difficult. The intrinsic cognitive load is relatively low as most of the cues are simple pairings with a required action. These include playing specific pitches or chords, cueing named samples, and using sound-making objects. When presented individually as discrete actions the task is relatively trivial, but the extrinsic cognitive load created by the design of the piece affects the level of difficulty due to the split-attention effect, modality effect, and transient information effect.

The deployment of cues can be varied when a version of the piece is set up, and typically this might involve a tutorial approach, where a pair of categories, such as major and minor chords, are used in isolation before adding in more categories. This helps the pianist, and also the audience, understand what is happening. Each category is assigned a different voice which helps differentiate between the categories, while also emphasizing the split attention of different cues operating within the same mode. Mousavi, Low and Sweller report on a 1972 experiment by Allport, Antonis and Reynolds which found that 'people could attend to and repeat (shadow) continuous auditory speech while processing unrelated visual scenes or sightreading piano music' [7] which suggests that the modality effect is helpful for audio scores that instigate instrumental performance. In contrast, however, the same experiment evidences the problem of receiving multiple verbal instructions, noting that 'if participants were presented with a series of words to remember, recall was worse if the words were presented in auditory rather than visual mode...Hearing words while listening to speech interfered with memory of those words more than did reading them.' [7, p. 321] While not directly analogous with the situation the pianist faces in my piece, it may be significant in increasing the extraneous cognitive load.

As this is an audio score, transient information effect is also apparent as the cues are delivered in time requiring the player to store new information while simultaneously acting on previous information. The flow is outside the player's control so the cognitive load cannot be mitigated by self-pacing or segmentation, although this may be a feature of the score design. The transient information effect therefore impacts on player efficacy, especially in sections where there is a very dense, overlaid sequence of cues given across multiple categories.

The other significant layer in the piece—holding conversations with the AI—further taxes the player's working memory by introducing an additional activity to attend to. There is a distinct difference involved in responding mechanically to simple word-action cue pairs and holding a free-form conversation with a fairly unresponsive AI persona. The difficulty here may be explained by the cost associated with task switching, especially as engaging in the creative thinking necessary in conversation is set against listening and responding to instructions and performing a range of disparate physical performance actions. The player has some control over what they choose to do—they may decide or forget to talk to the AI for instance—but this too may be a necessary way to manage the cognitive load when there is too much to do.

3.2. it is really hard to focus when there are too many things to do (2023)

Managing tasks in this way occurs slightly differently in pieces where players do have more control over the pacing of cues. In my ensemble piece *it is really hard to focus when there are too many things to do* (2023) [11] players are presented with a collection of tasks which involve completing different series of actions, including playing melodic patterns, sequences of object sounds and samples. Individual players may cue these sequences, affecting the behaviour of the group. Once a sequence has been started, players must complete it while trying to manage the completion of any unfinished sequences. Simultaneously, other instructions modify the way they complete these tasks, necessitating decisions as to how best to proceed to do the best they can.

Here the instructions are in part delivered by other players, so are in effect extraneous for each individual, but players are also able to deliver their own instructions so the transient information effect can be partially reduced through self-pacing. In this piece the instructions are also delivered live by human performers who may respond to the current situation, which contrasts with the piano piece where the instructions are regulated in a controlled random manner by the app. This may introduce segmentation as the ensemble tacitly agrees to pause instruction giving to reduce the cognitive load (or for other reasons, such as satisfaction with the current musical material).

Other noted effects are apparent. Split-attention effect and task switching are highly relevant as players playfully cue new actions and musical material. In parallel a rising and falling numerical count determines when the piece will finish, with players shouting out numbers one higher or lower than the previous number, and the piece ends when the count gets to zero. Consequently, players are required to track multiple cues related to different types of action. Frequently these will interrupt each other, such as beginning a new pattern while finishing playing a sequence of samples. The high task switching costs create additional cognitive load as players need to make executive decisions about how to resolve incompatible instructions where they are required to do two or more things at once.

4. CONCLUSION

The discussion of these two pieces demonstrates how cognitive load theory provides a productive framework for understanding and designing audio scores that deliberately create unstable performance situations. In *just trying to do what I'm told and keep up and do the best that I can*, the combination of split-attention effect, transient information, and task switching costs creates moments where the pianist must selectively attend to some cues while necessarily abandoning others, developing real-time coping strategies such as forgetting to engage with the AI conversation when other demands become overwhelming. In *it is really hard to focus when there are too many things to do*, the ensemble's collective agency introduces different dynamics: players can self-pace and segment instruction delivery through tacit coordination, partially mitigating transient information effect, though this comes at the cost of heightened task switching as they navigate multiple concurrent sequences and the numerical countdown. The contrast between these pieces suggests that player control over pacing fundamentally alters how cognitive load manifests, shifting from individual overwhelm in the solo piece to distributed decision-making in the ensemble context.

This application of cognitive load theory inverts its conventional purpose. Where instructional design typically aims to reduce extraneous cognitive load to optimize learning, my compositional approach deliberately manipulates these same parameters to create productive difficulty. This inversion is artistically meaningful rather than simply perverse because it makes visible the processes of coping, failure, and adaptation that characterize contemporary experience more broadly. The audience hears not only the musical results but also the explicit verbal cues and performer responses, creating what I hope becomes an empathic engagement with the experience of managing competing demands under pressure. The ethical dimension of deliberately imposing overload on performers is mitigated by transparency—players know what they are entering, the difficulty is part of the artistic proposition, and the piece's title and structure make explicit that 'doing the best that I can' is the goal, not flawless execution.

For composers working with audio scores, this suggests several practical considerations. Manipulating the density, simultaneity, and sequentiality of verbal instructions provides direct control over extraneous cognitive load, while the intrinsic complexity of individual tasks can remain relatively low. Incorporating germane cognitive load through repeated instructional patterns helps players develop schema for response, making the overall task more navigable even when density increases. The modality effect suggests that combining visual and auditory information may reduce overall load, though in some contexts—as with the piano piece—maintaining everything in the auditory domain may be preferable for aesthetic reasons. Most significantly, decisions about player agency over pacing emerge as crucial: giving performers control fundamentally changes the experience from endurance to negotiation.

Cognitive load theory and its resultant effects present a useful lens through which to view audio score design. While some of these observations about audio scores are perhaps unsurprising as dealing with lots of dense simultaneous spoken instruction is obviously hard, the nuance presented by the theory may prove to be useful in structuring some elements of a score. Given that this field serves as a theoretical basis for instructional design, drawing on its strategies for managing cognitive load in learning tasks is in many cases directly transferable to scoring. In particular, considering how extrinsic cognitive load can be balanced to create required performance outcomes and germane cognitive load can help develop schema to support players' ability to complete tasks effectively, a better understanding of how players navigate this form of notation might emerge.

5. REFERENCES

[1] J. Perra *et al.*, "Review on eye-hand span in sight-reading of music," *Journal of Eye Movement Research*, vol. 14, no. 4, 2021.

[2] J. Saunders, "Perfect information: scores as self-evident processes," in *Ninth International Conference on Technologies for Music Notation and Representation*, 2024, pp. 53–57. Available: https://www. tenor-conference.org/proceedings.html.

[3] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," *Learning and Instruction*, vol. 4, no. 4, pp. 295–312, 1994.

[4] F. Paas, A. Renkl, and J. Sweller, "Cognitive load theory and instructional design: Recent developments," *Educational Psychologist*, vol. 38, no. 1, pp. 1–4, 2003.

[5] J. Sweller, "Element interactivity and intrinsic, extraneous, and germane cognitive load," *Educational*

Psychology Review, vol. 22, no. 2, pp. 123–138, 2010.

[6] J. Sweller, J. J. G. Van Merriënboer, and F. Paas, “Cognitive architecture and instructional design: 20 years later,” *Educational Psychology Review*, vol. 31, no. 2, pp. 261–292, 2019.

[7] S. Y. Mousavi, R. Low, and J. Sweller, “Reducing cognitive load by mixing auditory and visual presentation modes,” *Journal of Educational Psychology*, vol. 87, no. 2, pp. 319–334, 1995.

[8] M. Chis et al., “The cognitive load – productivity tradeoff in task switching,” in Proceedings of the Human Factors and Ergonomics Society Annual Meeting, vol. 67, no. 1, 2023, pp. 666–671.

[9] A. Kiesel et al., “Control and interference in task switching—a review,” Psychological Bulletin, vol. 136, no. 5, pp. 849–874, 2010.

[10] J. Saunders, “just trying to do what I’m told and keep up and do the best that I can” [musical score], 2025. Available: https://researchspace.bathspa.ac.uk/17643.

[11] J. Saunders, “it is really hard to focus when there are too many things to do” [musical score], 2023. Available: https://researchspace.bathspa.ac.uk/15305.