

MUSIC AND SCHEMA THEORY

EDST5303 Human Cognitive Architecture

Prof. Paul Chandler

Okko Buss, 3152903

June 2005

Understanding of human music perception and production as well as the workings and limitations of human memory have made significant advances in recent decades. This is largely due to more sophisticated theoretical frameworks reinforced by empirical evidence. The aim of this paper is to investigate some recent advances in both theory and research, with the hope that specific evidence from the latter, musicology, may help sketch-out crucial concepts such as schema-theory. In addition, I will conclude with a discussion of possible implications for music pedagogy, both in terms of music appreciation and production.

A comprehensive discussion of the role of cognition, schemata and memory in music/musicology was compiled by Marc Leman (1995). While applying recent findings in musicology, psycho-acoustics and computer science, it also echoes many findings described in Human Cognitive Architecture and the modal model of memory (Healy/McNamara, 1996). However before engaging in an application of Leman's work to Human Cognitive Architecture, critical concepts need explanation.

Human Cognitive Architecture (henceforth HCA), understands cognition to be a dynamic information-processing system comprised of different interacting types of memory and the processes and sub-structures contained within these memory structures. It is particularly concerned with Long-Term Memory (LTM) and Working Memory (WM), but also accounts for the role of Sensory Memory (SM).

One cornerstone of the interaction of these two types in particular is 'schema theory'. Understanding these structures and the processes involved, that is to see how HCA works in general (especially LTM, WM and schemata), is important before attempting an application of evidence from Leman's work in particular.

Underlying HCA is the observation that human WM is severely limited both in terms of its capacity (Miller, 1956) and duration (Greene, 1992). This observation

has been established for several decades since George E. Miller's seminal publication and remains undisputed in its basic form to this day although the exact nature of the limitation is notoriously difficult to quantify. Miller (1956) posited an approximate limitation of "seven, plus or minus two" elements present in WM at any given time, where an element is taken to be the most basic unit of information.

The problem of measuring WM is one of controlling for LTM. At any given time it is difficult to establish whether an element in WM didn't in fact come out of LTM where it was previously stored. Moreover, WM is known to decay quickly. The exact number of elements in WM storage is therefore intangible, and the real task of experiments is to conceive tasks to control for this dual trap of LTM interference and WM decay. As a result of this difficulty of controlling for the effect of LTM on WM and the nature of decay in WM, alternative quantifications have been suggested, however for the purposes of this paper and understanding HCA, the exact limitation is not important.

WM is furthermore comprised of at least two slave-systems, an auditory loop and a visual-spatial register. For the purposes of discussing the workings of music and cognition, the former is of greater interest, although a comprehensive study of musical cognition must account for the role of the visual-spatial register in musicians reading sheet music.

The auditory loop contains elements of aural input from the audio Sensory Memory (SM), an ephemeral memory register of about 1.5 seconds which pre-filters audio input for WM. Traditionally, the role of this loop is primarily linguistic, keeping in memory a store of streaming speech for interpretation. While language allows for theoretically infinite syntactic structures, this limitation of SM and the auditory loop in WM puts a natural limit on how complex a sentence humans can parse. In music,

this limitation applies equally, determining how complex a perceived musical structure can be.

Another critical element of HCA is the nature of LTM. Unlike WM, it is known to be stable over time and practically limitless. It serves as long-term store of various types of knowledge, which have roughly been categorized to be at the very least procedural ('how'), declarative ('what') and conditional ('why') types, though episodic memory ('when') is also known to exist independent of these. For music cognition, this particular aspect is of lesser interest than the general question of how WM and LTM interact, i.e. how LTM receives knowledge from WM to store it and how it returns it during retrieval.

The contents of LTM must first be processed in WM in order to be encoded and/or retrieved later. There has been some debate as to how this is achieved in HCA. This question of control was attempted by A. Baddeley (1992), who posited the existence of a 'control executive' to do this important task. This control executive was said to be a third sub-system of WM, existing alongside the auditory loop and the visual-spatial register. Its role was supposed to be to co-ordinate the encoding of information in LTM passing from SM through WM. In reverse, retrieval of information from LTM into WM is a second putative function.

This model has been reworked and disputed over the years. Baddeley himself reworked it such that the control executive only needs to account for processing of novel information. Alternatively, the control function can be done away with entirely, by attributing control to schemata in LTM. In this case, it is the schematic information stored in LTM that controls what goes into WM for processing by providing schematic context for new information to be encoded or previously stored information to be retrieved. Leman's work confirms this observation when he posits

what he calls 'self organizing maps', as we shall see later.

This leads to the last important concept in HCA, a discussion of how knowledge (procedural, episodic etc.) is stored in memory. It may be the strongest commonality between Leman's work and HCA: schema-theory. Both musicological findings and HCA accept the notion of schemata as "building blocks of knowledge" (Rumelhardt, 1981). Schemata in this sense serve two specific functions. They guide learning by providing context for new material and they assume a control function for elements stored in WM.

The former is a well-established fact from a century of human developmental psychology, though it's essential claim goes back to Immanuel Kant, who in 1755 published his *Universal Natural History* and developed his view of human knowledge in terms of schematic representation over later works. Kant believed that only a very limited amount and basic type of knowledge was innate: the concepts of causality, space and time. All further knowledge is acquired, although humans have the propensity to build knowledge structures that develop systematically over time and with experience. There are strong echoes of this view in the observation that schemata are hierarchically organized and in what Leman calls 'self-organization' in his model, as will become apparent later in this paper. However this basic observation from Kant underlies the modern notion of a knowledge schema.

In the early 20th century, Jean Piaget (1928) developed the concept of a schema further by systematically studying how children contextualize new knowledge in terms of existing knowledge. To Piaget, schemata assume a guiding role for contextualization of new knowledge through three processes: assimilation, accommodation and equilibration. Some definition and examples help illustrate these processes.

New knowledge can be assimilated to a closely matched existing schema. When one encounters a black horse, previous experience only with brown horses nevertheless allows the assimilation of the black horse into the a schema for horses. If no such close schema exists when processing novel information, that information is usually accommodated into existing schemata. For example, knowledge of horses but not zebras allows classification the zebra as a relative of horses. Eventually, schemata are said to seek balance through equilibration, that is stable schemata for zebras and horses will form and co-exist.

This approach is useful when considering how completely novel information is processed and the kind of errors (of classification, interpretation or otherwise) occur as a result. Europeans equipped with knowledge about the world from natural history struggled to properly classify and schematize the platypus. F.C. Bartlett showed that cultural context has a similar schematic basis, as discussed later on in this paper.

Aside from providing context for novel information, schemata provide a second important function, allowing LTM to control what information passes in and out of WM. While there are other candidate structures for this important task, such as mental models, which are “knowledge structures ... used in performing specific tasks (Leman, page 41)” and scripts, which describe different kinds of knowledge (e.g. What do we know when we sit, order and eat at a restaurant), the schematic approach will be discussed here since it is compatible with findings in musicology.

By providing context for new knowledge and the interpretation of incoming stimuli in WM, a schema determines what can go into WM. Without a structural, schematic notion of knowledge, it would be impossible to account for orderly and systematic information processing in WM. To rephrase a popular phrase about the Buddha on

the mountaintop, whom one cannot find there unless he is already within one's self, the only knowledge found in WM is the knowledge brought there by LTM, and for that to occur, we need schemata.

The existence of schemata that fulfill these functions has been suggested in studies by F.C. Bartlett (1932). He demonstrated that Piaget's characterization of learning and dealing with new and unknown information as schema assimilation, accommodation and equilibration was correct. In his famous War of the Ghost experiment he presented American university students with a mythological native American story. When asked to summarize and answer basic questions about the story, the students showed that they were deeply entrenched in their western notions of literature that they largely misunderstood the myth. In other words, their schema for plot, characters, drama etc only allowed them to read this new story in terms of those existing schemata.

From the results of the study, Bartlett introduced 'flattening', 'sharpening' and 'rationalization' to the terminology of schema theory to explain how schemata are constructed. However what may be more important, the experiment demonstrated that cultural context has a schematic basis. The interpretation of stories told from outside a familiar cultural context, is difficult and readers often fell back on familiar elements in order to make sense of the material. Leman, as we will see, will offer important insights to make a similar case for music cognition.

In summary, HCA, or the modal model in particular, is a model for cognition as information processing by making a tripartite distinction of SM, WM and LTM, elaborating substructures and processes in each, and by accepting schema theory as a basis for knowledge structures. Psychologists and philosophers such as Kant, Piaget and Bartlett have worked-out an established description of what a schema

and its role in cognition is, and how it serves to underpin a description of the learning process in terms of HCA. However there's relatively little knowledge of how a schema is manifested in the human thinking apparatus, the brain. Leman's research on music cognition will offer some insight into this what the ramifications of an implementation of schemata in the brain are for HCA. However a few musicological and related concept first need clarification.

The most critical concept needed to understand how the human mind perceives, produces and rehearses music, is a notion of 'tone semantics'. Tone semantics has several distinguishing features. It defines how and why a listener will perceive a particular piece of music as harmonic, melodious or generally 'pleasing' and be able to distinguish dissonance. It has to satisfy the fact that listeners are apt judges of these things and that such judgments are not random.

Secondly, a theory of meanings of tones has to account for the fact that "the meaning of a tone (or chord) can be determined by its context, while the context itself is determine by the constituent tones (chords)" so "what is determined by the context is itself part of that context. (Leman, page 4)" Both of these features of tone semantics foreshadow implications for schema theory.

There have been several enterprises to account for these two aspects historically. The oldest known is that of the Greek mathematician Pythagoras, who "discovered that tone intervals could be represented by simple ratios between lengths of strings. (Leman, page 4)" Since these rations were thought to be isomorphic with those between bodies in the sky, it lead him to the conclusion that harmony and melodiousness are a result of mirroring the perfection of mathematical relationships between celestial bodies.

Pythagoras's work, while failing to account for the true nature of how tones

acquire meaning, paved the way for many later enterprises. In the 18th century L. Euler provided a strikingly accurate mathematical account based on G. Leibniz's mathematics and Pythagoras's interval theory, however it failed to account for context-sensitivity (Leman, page 5).

Other musicians (Zarlino, Rameau) and musicologists (Helmholtz et al.) have attempted similar explanations over the years but none remedied the essential shortcoming that mathematical and sensory accounts couldn't get their heads around: "[M]usical contexts indeed involve learning processes which introduce a cultural factor. (Leman, page 8)" In other words, musical perception and production is founded on the experience one has with music.

Knowledge about HCA and modern musicology both tell us that this experience has an effect on schema construction. In terms of HCA, this process of concerted learning of musical structures (whether intuitive ones about 'what sounds right' or the in-depth and explicit knowledge of theory in professional musicians) taking place over long-periods of time and requiring sufficient musical input. Musicology, or at least some musicologists like Leman, who take schema theory to be the underlying model of musical representation share this point of view (Leman, page 41). The logical conclusion is that the result of exposure to music is a schematic representation of 'tonality' in the human mind. This is the motivation of the investigation that follows: to gain an understanding of tone semantics in terms of schema theory.

While HCA and musicology share a belief in a schematized mind (or at the very least schematic LTM storage), they do not, however, have a common understanding of how a schema may be represented in the human brain. To answer this question, Leman employs tools from AI and computer science. Advances in connectionism

have offered him a plausible way to test how schemata may be represented in the brain. A quick discussion of connectionism and “neural net” will help understand this approach.

Connectionism is a movement within Cognitive Science and Artificial Intelligence research that attempts to explain cognition in terms of neurology and the study of the brain. While there are skeptics who do not believe that a satisfactory approach to cognition can be a reduction to neurology (for some arguments see Fodor (1975) and Pylyshyn (1985)), the connectionist approach has made key breakthroughs in pattern recognition, establishing itself among other cognitive disciplines. Its fundamental process and significance is briefly discussed below.

Its most notable contribution to the AI toolkit and methodology of modeling the brain have been neural nets, computer models that simulate input, middle and output layers of 'neurons' (i.e. computer implementations of real neurons). Neurons are connected with each other (hence connectionism) and each connection has a weight. If the aggregate of weights from firing neurons connected to a neuron is greater than that neuron's activation thresholds, it in turn fires. The net is trained by presenting it with iterations of input data (such as text, chords, pictures) during which the connection weights automatically adjust themselves. The result is a system that can represent practically limitless and novel inputs in terms of a systematic firing of output neurons. This is its main breakthrough in pattern recognition.

By training a special neural net, a so-called 'self-organizing map' (henceforth SOM, Kohonen, 1984) Lemn demonstrates how musical experience translates into schemata such as tonality and melodiousness. Interestingly, while it is not yet possible to build and test complicated schemata such as scripts (e.g. the oft-cited 'eat at a restaurant' script/schema) on a neural net, perception of musical tone offers

a wonderful opportunity here. There are two reasons for this.

Firstly, perception of tonality in music is relatively autonomous of other mental schemata. For a schema of a complex script, like 'eating at a restaurant', to be complete (i.e. for one to have such a script/schema), many different sub-schemata are required. Concepts such as 'chair', 'door', 'waiter' etc are all part of the script, which requires an intricate interplay of huge number of sub-schemata. This is not the case in perceiving tonality, harmony or melodiousness in music. One can determine tonality (or 'tone center' as it is more appropriately called) of a series of chords played on a piano just the same if the same series is played by a violin section. Semantics in music, for which Leman's neural net is intended to hold as a schema, is largely autonomous of the many other schematic relations that riddle effective testing for schemata pertaining to most other types of complex knowledge.

Secondly, the neurology of music perception is well-understood (Leman, page 30). The mechanisms used by Leman to train his SOM closely resemble those that occur in real-world situations when real people experience real music (Leman, page 66). It would be difficult to convey to a neural net the intricacies of scripts or mental models for knowledge such as eating at a restaurant. This is largely a question of scope, since many faculties of the cortex and other parts of the brain and body are involved in the representation of such complex scripts and models. But the most fundamental cortical properties of music experience have been worked out by psycho-acoustics. Accepting the autonomy of recognizing harmonic structures from other tasks such as walking and talking and attributing them to specific neuronal activity underly Leman's approach.

With these two disclaimers about neural nets as an appropriate indication of how a schema for tone semantics might be represented in the brain, it is possible to take

a look at how an appropriately trained SOM works and how it fulfills this function. While a SOM (or neural nets in general) are not the same as the brain, they are the closest thing to it available to experimentation, especially with respect to training/learning. Before seeing how one might use a SOM for illustrating how schemata of tonality are formed, a quick discussion of what a SOM is and how it works is in order.

As mentioned when introducing the concept of neural nets, a SOM is a type of neural net and as such it has two characteristic functions of all neural nets: (1) it allows for training a simplified model of neuronal activity by setting it to an initial, unordered state and presenting it with iterations of input data and (2) once trained it can be tested by presenting it with previously unknown input which will create a typical neuronal output response. This makes schema construction a "long-term data-driven" process.

A further special feature of SOMs is that, as the name points out, "self-organizational". In detail, this essentially means that the process of going from its initially unordered state (all neurons are attributed random weights) to later ordered states (specific neurons are closely linked to particular harmonic relations, e.g. neuron X always fires if chord Y is struck) depends entirely on the type of input it is presented with. For example, if during training the SOM is presented with iterations of harmonic progressions from J.S. Bach's *Brandenburg Concertos* and F. Handel's *Fire-Music*, it will become most apt to interpret tone centers typical for Baroque music, more so than perhaps those found in Western classical music in general, and much more so than if it were asked to interpret harmonies of Balinese Gamelan music.

In his experiments, Leman trained a SOM neural net with a series of chords

composed of so-called "Shepard tones", which are essentially very pure tones devoid of timbre (for example a clean whistle blow, as opposed to the sound produced by a cello or a trumpet). This approach controls for clean unequivocal input and avoids that the neural net 'accidentally' learned to recognize features of its input that are irrelevant to tonality (e.g. Learn to distinguish trumpets from cellos rather than c-major from f-minor chords). The result was astonishing. After 300 iterations of training, the map had organized itself so that particular tones were closely associated with specific neuronal activity.

Moreover, it was possible to express familiarity with tonal relations (i.e. chord progressions) in terms of distances and connections weights between neurons on the map. This is said to be analogous to how the neurology of the cortex schematically represents familiar contexts and their relation to each other: "The response structure of the the long-term data-driven adaptation is similar to the response structures found in mental representations. (Leman, page 88)". (For some graphical representations of the development of this map, illustrating these observations, please see Appendix A.)

While Leman's work stands as a compelling argument for connectionism as a model for how the human brain might perform relatively isolated cognitive tasks, it can be evaluated to shed light in a novel way on some of the core issues and observations of HCA. Three in particular, can be addressed: The relationship between culture and schema (cf. Bartlett), the nature of learning as a concerted long-term process and the control-issue (cf. Baddeley).

The first observation about a schematic basis for what are commonly thought of as cultural phenomena has strong parallels in Leman's and Bartlett's works. While Bartlett demonstrated that students interpret how interpretation of literary elements

is largely due to the schemata that a reader possesses, Leman's work on musical SOMs underlines how a culturally determined idea of tone center and harmony is manifested neurologically as well as schematically. As discussed, musical knowledge such as intuitions about common chord progressions were interpretable from his trained SOM. The fact that people from different cultural, and therefore musical backgrounds have different but equally pronounced intuitions about consonance and dissonance presumably can be sufficiently represented by an SOM trained with data from different background. While work on this is in no way conclusive, Leman's work offers what might turn into a fertile testing ground for more in-depth investigations into the relationship between culture and schema.

Secondly, conclusions from HCA about the learning process are, at least in part, confirmed in Leman's work. According to HCA, learning cannot occur other than as a concerted long-term effort, during which complex schemata are constructed from simpler ones. Learning by osmosis, as various learn-while-you-sleep language tapes have tried to claim in the past, is impossible.

Likewise, as the interpretation of the musical SOM after one, 60, 150 and finally 300 iterations demonstrates, even on a simple neural net learning cannot be a one-shot deal. An untrained neural net has at best a random shot of interpreting a chord to be consonant or dissonant, or guessing at the notes comprised in a harmony that it is presented with. After the full cycle of 300 iterations of training, the tuned SOM was able to distinguish chords by having a *characteristic neuron (CN)* attributed to particular chords/harmonies (Leman, page 77). The process of training and tuning a neural net thus serves to underpin the claim from HCA that long-term commitment to data and effort necessary in schema-construction.

The last of the three common observation between the findings from HCA and

Leman's neural net is related to the question of control in cognition. More precisely, the question asks how input from SM is interpreted in WM and how does that relate to retrieval from LTM. As discussed, Baddeley has offered an explanation in terms of a 'control executive' and this has been contested.

Leman agrees that this extra structure in WM is unnecessary. Furthermore, he assumes that part of the definition of a schema is “a control structure in the human brain that is sensitive to some frequently occurring pattern. (Leman, page 40)” Since schemata also describe the way in which knowledge in LTM is said to be organized (note the parallel with “self-organization” in SOMs), we can arrive at a synthesis of two ideas from HCA and Connectionism, claiming that the seat of control is best attributed to schemata, residing in LTM. Because Leman's work shows that a schema has the ability to control interpretation of input, his work offers a unique piece of evidence that schemata control what enters into WM and how that information is interpreted.

Finally, there are two brief observations about music cognition in general that are supported by the experimental work with neural nets and music, which have bearing on our understanding of the learning process. One is concerned with what is best described as music appreciation, the second is more practically related to music production (composition) and performance.

Music appreciation, insofar as the understanding of novel harmonic structures is concerned, requires that the listener is acquainted with the forms he or she is listening to. Much like Bartlett's students who failed to get their heads around the essence of a story, whose format they were unfamiliar with, harmonic forms and tonalities can deter from the experience of listening to a piece if they are unknown. However schematization of a broad range of musical contexts requires at the very

least continued exposure to such contexts.

One thing that may be worth noting at this point is that the nature of our understanding of music is something that does not require the same level of conscious commitment as many other analytical tasks, such as mathematics. Humans seem to develop a well-founded intuition about harmony and tonality without explicit training, rather than by long-term exposure to music. While this sounds dangerously close to 'learning by osmosis', it seems music is much more akin to language acquisition (Pinker (1995)) in this respect, which also happens naturally during early developmental psychology. In that light, it would seem advisable to submit children to a large range of musical influences at an early age in order to broaden their intuitive understanding of tonality, musical forms etc.. Whether there is some basis to this claim would have to be empirically verified, which may be difficult due to the long periods of time during which subjects would have to be kept under supervision. However this may prove to be fertile ground for research.

Secondly, this observation applies to composers and performing artists or anyone involved in the creative process of making music. Here, however, the 'intuitive' acquisition argument probably does not hold. To be able to employ musical forms in creative ways when writing/performing/improvising music, a concerted study of musical forms is required. Nevertheless, continued exposure to novel musical forms, harmonies, tonalities can have a significant impact on a composer seeking inspiration. Schematization of such novel forms allows for their retrieval during composition at a later point. Again, it would probably be difficult to find a safe metric for the effect of novel musical influences on musician's problem-solving abilities concerning tonality (or any aspect of music). At the very least, it can't hurt.

REFERENCES

- Baddeley, A. "Working memory" from *Science*, 255, 556-559, 1992
- Bartlett, F.C. *Remembering: A study in experimental and social psychology*, 1932, CUP, London
- Fodor, Jerry. "The Language of Thought", Thomas Y. Crowell Co., 1975
- Kohonen, T. *Self-Organization and Associative Memory*, Springer Verlag, Berlin, 1984
- Leman, Marc (1995) *Music and Schema Theory: Cognitive foundations of a Systematic Musicology*, Springer Verlag, Berlin
- Miller, G.A. "The magical number of seven, plus or minus two: Some limits on our capacity for processing information" from *Psychological Review*, 63, 81-97, 1956
- Piaget, J. *Judgment and reasoning in the child*, 1928, Harcourt, New York
- Pinker, Steven. "The Language Instinct", Penguin Books, London, England, 1994
- Pylyshyn, Zenon. "Computation and Cognition", MIT Press, Cambridge, MA, USA, 1984
- Rumelhardt, D.E. "Schemata: The building blocks of cognition" from *Comprehension and Teaching; Research Reviews*, International Reading Association, Newark, 1981