\128

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

A.I. LABORATORY

Artificial Intelligence
Memo No. 398

December 1976
LOGO Memo No. 42

Capturing Intuitive Knowledge in Procedural Description

by

Jeanne Bamberger

Trying to capture intuitive knowledge is a little like trying to capture the moment between what just happened and what is about to happen. Or to quote a famous philosopher, "You can't put your foot in the same river once."[1] The problem is that you can only "capture" what stands still. Intuitive knowledge is not a static structure, but rather a continuing process of constructing coherence and meaning out of the sensory phenomena that come at you. To capture intuitive knowledge, then means: Given some phenomena, what are your spontaneous ways of selecting significant features or for choosing what constitutes an element; how do you determine what is the same and what is different; how do you aggregate or chunk the sensory data before you?

[1] Cratylus (5th Century B.C.) Paraphrased by Donald Schon, 1965

The work reported in this paper was supported in part by the National Science Foundation under grant number EC40708X and conducted at the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA.. The views and conclusion contained in this paper are those of the author and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the National Science Foundation or the United States Government.

Capturing Intuitive Knowledge in Procedural Descriptions

Trying to capture intuitive knowledge is a little like trying to capture the moment between what just happened and what is about to happen. Or to quote a famous philosopher, "You can't put your foot in the same river once."[1] The problem is that you can only "capture" what stands still. Intuitive knowledge is not a static structure, but rather a continuing process of constructing coherence and meaning out of the sensory phenomena that come at you. To capture intuitive knowledge, then, means: Given some phenomena, what are your spontaneous ways of selecting significant features or for choosing what constitutes an element; how do you determine what is the same and what is different; how do you aggregate or chunk the sensory data before you?

As for description--procedural or otherwise--these same internal processes for constructing coherence mediate naming and the whole range of possible other modes of description including their various symbols and media. So descriptions, like internal structuring processes, focus on some aspects of the phenomena and ignore others. Thus, a description will also be characterized by its particular selection of features as significant, by how it defines an element, how same and different are determined and by its implicit means for aggregating elements and building relations.

Interesting questions arise, then, in the interactions between internal "knowing" and out-loud descriptions--both as one tries to express his own knowledge and in the way one is influenced by another's descriptions.

[1] Cratylus (5th Century B.C.) Paraphrased by Donald A. Schon, 1965.

A computer music system can be a rich environment in which to explore these questions exactly because a description as input to the computer actually and reliably generates what is described. We can ask, then, what is the relation between the description as given to the computer--its meaning as locked into the implicit (or explicit) choice of elements, relations, level of detail, etc., and an individual's immediate apprehension, the sense he constructs--i.e., his choice of elements, relations and level of detail or aggregation?

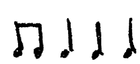
The question can be dealt with in various ways: You can accept the computer description as simply a vehicle for input and play with the code as code--manipulating the built-in, powerful potential for inventing formal structures, ignoring entirely the probable gap provoked by the differences between your intuitive structuring (your immediate apprehension) and the formal description. Or you can think, even, evidentally, come to hear, in terms of the structures imposed by the computer code, rejecting your earlier intuitions as primitive, unreliable. Or you can confront the incongruences between the two.

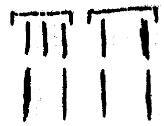
But if you do the latter, you may risk at least temporary cognitive disequilibrium. For if you confront the incongruences between coded description and immediate apprehension, you may risk shaking up, even giving up the intuitive means by which you have, up till now, found coherence and meaning in pitch-time relations. And besides, it's not easy! You must somehow try to hold still those evanescent actions of construction which I have called intuitive knowledge and also probe these same constructive processes implicit in the computer description.

But if you can learn to do that, you have, it seems to me, the possibility to gain new insight not just into the particular bit of stuff you have described and caused the computer to generate, but more important, into the very processes by which we find or construct musical meaning from sensory data.

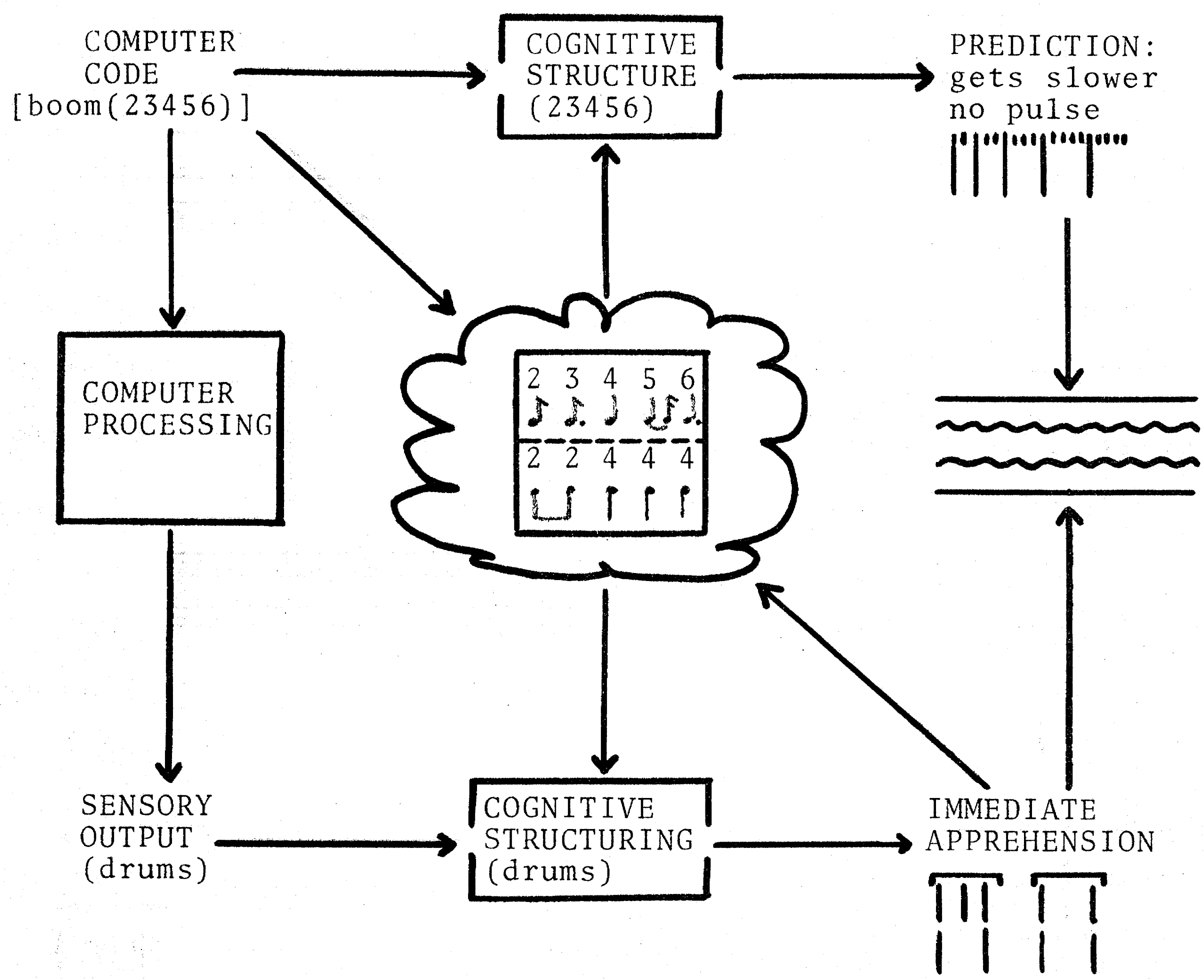
Let me illustrate with a very simple example which I will proceed to make probably unnecessarily complex. The example is a true one which happened some time ago in the LOGO LAB at M.I.T. that served to trigger the kind of confrontation I have just described.

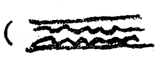
I had asked my students (mostly musically untrained M.I.T. undergraduates) to invent some figures which were free of any sense of underlying pulse. The figures were to be generated by the computer and played by one of the percussion sounds on the computer controlled "music box". The students know that BOOM is the command to the music box drum and that numbers indicate time values--i.e., time from one attack to the next--such that, for example, 4 is twice as long as 2, half as long as 8.

Using the system, now, as a sophisticated sketch pad for design and testing, this student types: BOOM [2 3 4 5 6]. Since the numerical relations have no common factor, he expects to hear a figure with no underlying pulse, a figure that simply gets progressively slower. But to his surprise, as he listens, reconstructs the figure in his memory and claps it back, he finds a neat, almost metrical figure:  . The meaning he spontaneously constructs doesn't match the meaning suggested by the coded instruction. Events which look like they should sound different, sound the same. He describes the figure as in two groups--three events and then two. And there seems to be an underlying pulse, the two groups are nearly equal--two beats in each, four beats in all. He draws a picture of the figure as he hears it:



What's wrong? Has he mistyped? Is the system malfunctioning? He tries it again. This time he hears the figure as slightly tipsey. His apprehension is changing as it interacts with the coded description--but it is still not what he expected. Confronting the incongruences thrusts him into inquiry. Together we try to model the problem:



Looking at his coded instruction (NOTATION), he constructs its meaning (COGNITIVE STRUCTURING) as time relations (GET SLOWER, NO PULSE). He types the instruction to the computer which sends the results to the music box which translates it into sound (DRUMS--SENSORY OUTPUT). Listening, now, he structures the sensory data (COGNITIVE STRUCTURING) which results in his IMMEDIATE APPREHENSION. He finds that his prediction (GET SLOWER, NO PULSE) is incongruent with his IMMEDIATE APPREHENSION (). But in order to get hold of the incongruences between his immediate apprehension and his predictions, he must make an out-loud description--he claps, he talks about grouping and pulse. Using the two descriptions as evidence, now, he has the possibility to reflect on the differences in cognitive structuring--i.e., how he constructs meaning in response to the code in contrast to his construction of meaning from the sensory data.

His knowledge of numbers evidently leads him to take 1 as a basic unit of measure. Then comparing each number to the next, he sees an accumulating series: 1-11-111-1111, etc. Translating this into events in time, he compares each event to the next, implicitly counts up by 1's and imagines an accumulating series of attack times articulated by the drum sounds--each one now 1 time unit longer than the previous one and thus "going progressively slower".

How is this different from the cognitive structuring which leads to his memory of the figure, his clapping and his verbal description--two groups nearly symmetrical in time? Evidentally, his intuitive structuring includes a search for "nodes" (accents) in relation to which individual events cluster. At the same time, he constructs a temporal grid derived from the time relations between nodes into which he can fit the whole figure. Thus, 2 3 4 becomes a group bounded by and clustering in relation to the longer event, 4, the node of the first group. 2 and 3 are equalized and <u>4 is constructed as the unit time of the grid</u>. 5 and 6 form a group constructed at the still higher level of the group pulse with 6 as the node. 5 becomes a 4 by contextual association and 6, of course, can be anything since it has no subsequent event to delimit it. In this way, the student "bends" the absolute values (those described by the code) within the larger groups, regularizing them in relation to the temporal grid. This would explain his apprehension of the higher level relations--two groups, end accented and equal in total time:

2 3 4 5 6

2 2 4 4 4

No wonder he was surprised by what he heard! Translating the numbers into imagined sound, the means he used for constructing meaning--the implicit element, the level of aggregation and even the bases for aggretating--were all quite different. Instead of the 1, understood as unit time in dealing with the numbers, he constructed 4 as the unit time in immediate apprehension, derived from the figure's grouping and from the relations between nodes. In fact, he was aggregating at a much higher level. And instead of comparing each event to the next, consecutively, his intuitive structuring focused on the relation of low level events to the larger group in which they were embedded, the relation of one group to the next and all of it in relation to the constructed grid.

Testing his theory, now, he tries BOOM[2 2 4 4 4]. Of course he can hear the difference, but the higher level relations are right--it's a cleaned up version of his intuitive description--the grouping, the nodes, the grid match what he heard. Going on, he tries strings of random numbers for the values of drum sounds. The theory seems to hold. We were reminded of what St. Augustine had said long ago:

> For it is one thing to have the number, another to be able to sense the harmonious sound....

Maybe he can invent a language of higher level relations which would capture the on-line procedures that characterize his intuitive construction. But even short of that, he has the germ of a tentative theory of cognitive structuring of simple events in time which he can test. He has also shed some light on what might happen to descriptions of complex time relations in immediate apprehension--maybe that helps explain why beautifully designed formal structures, full of subtle rhythmic relations, often are heard as essentially simple structures with little rhythmic interest. But we need to ask further questions: What kinds of threshholds are significant to this process of grid-making and regularizing--tempo of the underlying pulse[2], for example; or the relation of rhythmic grouping to measuring or meter; what are intuitive levels of temporal discrimination and aggregation and does this change and develop with experience and learning? Will the theory hold when more parameters are added--e.g., how will pitch, texture, timbre influence grid construction and regularizing? There is a long way to go but the model of experimental procedure seems to be a productive one.

This rather too long discussion of a rather too short example serves to illustrate one way of exploring the question raised earlier--i.e., what sorts of interactions may occur between internal "knowing" and out-loud

---

[2] The tempo of the present example as generated by the computer is about ♩ = 120 where ♩ = 4 in the computer code. Clearly a slower tempo would have a very different effect, one closer to the predicted result.

descriptions? In the preceding example, incongruence between a given description (computer code) and the intuitive construction of sensory data (immediate apprehension) led to confrontation when the student made an out-loud description of his "knowing" that he could then compare with the given description. But what if we start with a set of given descriptions, each different, but all referring to the identical sensory output?

The following computer procedures each produce the same result when given as commands to the music box drum. In standard music notation, the figure would look like: BONGO ♩♩♩♩ ♩♩ --i.e., six equidistant hits on the electronic drum. The question now becomes: How does each description influence your internal knowing? Is the figure indeed "the same" when constructed through the "filter" of each description?

I	II
NOTE -27 4	BOOM 4
NOTE -27 4	BOOM 4
NOTE -27 4	BOOM 4
NOTE -27 4	BOOM 4
NOTE -27 4	BOOM 4
NOTE -27 4	BOOM 4

III	IV
REPEAT [BOOM 4] 6	BOOM [4 4 4 4 4 4]

V

BEAT 4 6

Example I is low level computer code: NOTE is the single music-playing primitive; -27 is the code word for the drum sound; 4 is its time value. For the computer-type this description has the most information--it tells him how the system works. Example II aggregates and speaks English--4 is an input to the single command, BOOM. III aggregates at a still higher level and introduces the function, REPEAT--it makes you think of the figure as one thing instead of 6 and it suggests cyclic action. IV is most like standard music notation--an instrument and what you play on it--it aggregates the time values into a thing which functions as a single input to the command or the instrument, BOOM. But V takes the biggest leap--it creates a new function called BEAT: A BEAT is some constant duration (4) regenerated some number of times (6). The concept is generalized and, as a procedure, captures the human process of BEATing: When you BEAT, you recursively regenerate some unit of time--to BEAT is to regenerate, not just to repeat! The procedure, itself, defines a meaning which reflects this process:

```
TO BEAT :DUR :TIMES
IF :TIMES = 0 STOP
BOOM 4
BEAT :DUR :TIMES-1
END
```

Through this new description, we find new meaning for our own experience. And yet each of the procedures is valid; each captures certain features and relations of the figure and ignores others. Together they provide a multi-dimensional, multi-faceted view of this little world.

This kind of multiple description-making often reveals unsuspected potential of even the simplest musical configurations--especially the potential for grasping new relations which, in turn provide the student composer with the potential for making transformations of some initial configuration. If a computer language gives its user the capacity to define and re-define elements and relations, it also encourages him to make structure-specific descriptions in terms of which he can think and hear in the course of developing a particular motivic idea or even the larger relations of a piece.[3]

My final examples show a small sample of this last process at work. The experiment began by playing with another general procedure when a student varied the inputs and got unexpected results:

	UP	:START	:INTERVAL	:DURATION	:TIMES[4]
1.	UP	0	1	2	24
2.	UP	0	4	2	7

To our surprise, Example 2 is described as faster than Example 1, even though the duration given to each event is the same in both. Why? Well, if you focus on the boundaries of each figure, the distance from bottom to top, then, indeed, Example 2 is faster than Example 1--it covers the same "distance" in about a quarter the time. The spatial metaphor comes alive!

---

[3] I was alerted to the importance of this capability in our LOGO music system when David Lewin recounted that he often feels like making a new "program" specific to each new piece he is working on. But, he added, this was, of course, foolish since he wasn't about to "start punching cards". While our system is truly minimal in terms of sound generation, its power as an information processing language built in the context of Artificial Intelligence research, makes it particularly useful for developing and testing "knowledge structures" including those inherent in the structuring of a musical composition. In this sense the system is, as I have hinted, a tool for designing and modeling, the ideas perhaps later to be further implemented and worked out on a system with sophisticated potential for timbre, dynamic range, etc.

[4] Example 1 starts at pitch 0 (:START), goes up by an interval of 1 (:INTERVAL) giving each event a duration of 2 (:DURATION) and does this 25 times in all (:TIMES). The result is a chromatic scale starting at middle C and ascending for two octaves. Example 2 goes up by an interval of 4 (a major third). Each event also has a duration of 2 and there are 7 events in all.

On the other hand, if we play this:

UP	:START	:INTERVAL	:DURATION	:TIMES
UP	0	1	2	25
UP	0	4	8	7

the same people respond that the second is slower than the first. Notice that their focus flips from outer boundaries to the duration of individual events within these boundaries. The event time is, indeed, slower than the event time in the first example (a duration of 8 as compared with 2)--but if the individuals had continued to focus on the relative time taken to traverse the distance from bottom to top (i.e., on boundaries), then they would need to respond that the two examples are the same--they cover the same distance in the same time.

Finally, in this last comparison:

UP	:START	:INTERVAL	:DURATION	:TIMES
UP	0	1	2	25
UP	0	1	1	25

everyone agrees that the second is faster than the first. In fact, this is the prototype which leads to the response, faster, in the initial comparison. Everything else being equal, you do indeed go faster if you cover the same distance in less time. Pitch-distance and time-distance can and must be distinguished but they have an extraordinary influence on one another--not a new idea for the sophisticated musician, but one that comes strikingly alive for the neophyte student in this experimental environment where procedural description and immediate apprehension are in constant interaction.

One student, intrigued with these possibilities decided to explore them further my making a whole piece. Starting out with scribbles on paper which captured his heard pitch-time relations in a spatial analogue, he translated these visual-spatial designs directly into computer procedures. Procedurally interrelated modules coordinate his "heard-in-head" scheme with what the music box actually plays.

His piece is made up of modules. The initial structural module (UP 0 2 2 5) goes through a series of transformations to make up a larger module which he calls LINE1:

```
TO LINE1
UP 0 2 2 5
REST 6
UP 1 2 2 5
REST 6
UP 2 2 2 4
DOWN 10 2 2 4
DOWN  2 2 2 4
REST 6
END
```

The procedure, LINE1, then, is a set of instructions for transforming a motive--it tells the computer in this way how TO LINE1. Listening to the result, lines of the procedure describe perceived structural "chunks", each one bounded or articulated by silence (REST).

Further transformations of the initial module create LINE2 and LINE3:

```
TO LINE2
REST 8
DOWN  16 2 2 5
REST 6
DOWN 15 2 2 5
REST 6
DOWN 14 2 2 4
UP 6 1 2 5
REST 6
END
```

```
TO LINE3
DOWN 8 3 4 4
DOWN 6 3 4 4
DOWN 4 3 4 4
DOWN 2 3 4 4
END
```

And finally each of the larger modules (LINE1, LINE2, LINE3) becomes, itself, a module in a superprocedure which he calls MYPIECE. Superimposing the larger modules in various combinations, his superprocedure, as description, captures our immediate apprehension of increasing density and activity of texture as well as the particular relations among separate voices. He and you can _hear_ what his notation describes.

```
TO MYPIECE
LINE1
CHORUS [LINE1 LINE2]
CHORUS [LINE1 LINE3]
CHORUS [LINE1 LINE2 LINE3]
CHORUS [LINE2 LINE3]
LINE1
END
```

We seem to have come full circle. Starting with a description in computer code (BOOM [2 3 4 5 6]), our student risked confronting the incongruences between its apparent meaning and the meaning he intuitively constructed in immediate apprehension. Examination of the nature of the mis-match triggered further explorations leading to insights into the constructions which I have called intuitive knowledge. Comparing a variety of possible descriptions of a single figure together with comparisons of varied spontaneous responses to another, provided further insight. Specifically, these comparisons reinforced the notion that intuitive knowing and out-loud description can be a powerful influence on one another as they interact, since each includes an implicit choice of salient features, level of aggregation, definition of same and different, etc. And finally, we saw a small attempt on the part of one student to coordinate his intuitive knowing and hearing with out-loud procedural description--his instructions to the computer came close to capturing the spontaneous structuring of immediate apprehension.

These stories suggest only a bare beginning for research which, if joined by others, could have implications in several directions. Teaching and learning music could be quite transformed. Our students might learn <u>as musicians</u> (for example, as A. Schnabel once said, "By experiment rather than drill") instead of being taught <u>about</u> music. In turn, further insight into the cognitive aspects of our musical intuitions, might quite dramatically transform traditional analysis. Finally, we can foresee the possibility for developing high-level procedural languages which will be close enough to the active, procedural constructing of intuitive knowledge that naiscent composers will be able to think in them--the computer could then provide an environment where thinking makes it so.