Piano Playing: A Model of Sight-Reading and Rhythmic Timing

Thu Minh Nguyen (tmn@drexel.edu)
Dario D. Salvucci (salvucci@cs.drexel.edu)
Department of Computer Science, Drexel University
3141 Chestnut St., Philadelphia, PA 19104 USA

Abstract

Musical performance involves both complex multitasking and, due to its rhythmic nature, strong interdependence among its various subtasks and temporal perception. We have developed a model of piano playing that integrates timing, playing, and sight-reading and demonstrates the relationship between these concurrent tasks. The model relies on an existing model of temporal perception to keep an accurate rhythm and maintain a (roughly) constant-time look-ahead for upcoming notes. The model successfully predicts several measures of pianist timing and eye movements as collected in two previous empirical studies. Through the model, we are also able to demonstrate that basic cognitive functions in ACT-R can be organized to optimize performance in complex and skilled tasks, specifically piano play.

Introduction

Musical performance in general, and piano playing in particular, involves managing a set of complex tasks. While much research has been done on various aspects of piano playing, such as sight-reading (e.g., Goolsby, 1994a, b) and sequence planning (e.g., Palmer & Pfordresher, 2003), none, to our knowledge, have extensively modeled multiple aspects together as multitasking in music. The interaction between them, however, should not be ignored. For instance, “look-ahead” — that is, reading ahead to upcoming notes — is limited by memory span; the pianist should only read as far as s/he can remember notes. However, it is considered is that look-ahead is also affected by the amount of time available between the pressings of each note. This lies in the fact sight-reading involves shifting attention to the note and encoding the information from each note into memory. The amount of time available between each beat determines the number of notes that can be read.

In our research, we have broken piano playing down into three major elements: rhythmic timing, sight-reading and finger determination. Due to the strong interplay between timing and sight-reading, our model concentrates more significantly on these two aspects. Timing in the model is based on the following model of temporal perception: as the pianist estimates the perceived passage of time, s/he uses this information to create a mental metronome of successive beats. These beats then form the basis of timing for individual notes, where each type of note is associated with a certain number of beats and counted out accordingly.

The sight-reading aspect of the model represents another significant component of piano playing. Sight-reading of course has a visual component in which a musician extracts information from the written musical notation, namely tone and duration, and translates these to physical movements, namely key presses. While encoding the information from an object the eyes stay fixed on that object. As a result, the fixation duration in sight-reading, just like in normal reading, can serve as an indicator of cognitive processing. With sight-reading, it is ideal to be able to look ahead as far as possible. However, look-ahead is limited by numerous factors. If a pianist sight-reads too far ahead s/he will start to forget the previous notes in the sequence. In this case, the effort is wasted and disruptions in performance are likely to occur. Furthermore, shifting attention to a note and encoding information for that note requires various cognitive processes that consume a significant amount of time and cognitive resources. Therefore, even if memory allows, the pianist only has time to read ahead a limited number of notes between each key press.

We are using the ACT-R cognitive architecture (Anderson et al., 2004) to create an integrated model of the tasks involved in piano playing. Keying of notes is simulated using simple finger “punch” motor commands (Byrne & Anderson, 1998). Rhythmic temporal perception is achieved through the formations of beats from the ticks produced by the temporal module (Taatgen, van Rijn & Anderson, 2004). The eye movements are made possible by the integration of EMMA (Eye Movements and Movement of Attention: Salvucci, 2001a). In our model, we are implementing time-dependent note encoding and retrieval, such that the pianist only encodes a note when there is time available. We fit this model to existing empirical data that quantifies piano-playing performance with two sets of measures, namely those pertaining to timing (e.g., actual vs. “ideal” beat times) and those pertaining to eye movements (e.g., average look-ahead while playing).

Model of Trained Piano Playing

Piano playing is a complex task that involves visual sensory and motor movements as well as various cognitive mechanisms. In order to achieve this, our ACT-R model utilizes routines from ACT-R/PM, EMMA, and the temporal module. The model focuses in particular on trained piano playing to study precise sight-reading and rhythmic timing. This implies the following:

1) The pianist sight-reads in a manner that facilitates the best possible performance.

2) Rhythmic timing is produced with high accuracy in relation to beats.

3) Errors caused by memory or motor movements are very minimal (negligible).

The trained pianists in the reported studies have had at least two years of formal training; this training more than suffices for the fairly easy pieces used in the studies discussed here. For example, the pieces call for the pianists to use only one hand that remains at the same position above the keys during piano play.

Model Outline

For our piano-playing model, all parameters in ACT-R were kept at their default settings. The basic flow of the model is as follows:

- Before starting to play, the model reads the first few notes up to the look-ahead limit of 1.3 seconds in total note duration (Furneaux & Land, 1999).
- The model retrieves the first note from memory; once the retrieval is done, it starts the temporal module.
- The production to punch the key for the retrieved note is fired and the next note is retrieved.
- The model encodes new notes when there is enough time between beats and the look-ahead is still below the limit.
- Whenever the temporal module reaches the one beat duration of 395 ms (16 ticks), which constitute the length of one beat, the model resets the timer for the next beat.
- When the current note has been pressed for the required number of beats, the model punches the next key.

Toward the end of the music, when there are no more notes left to read, the model simply plays the last few notes in memory and times the beats until completion.

While the model involves sequential firing of productions, tasks by different modules are prepared and done in parallel. For example, after the production to punch a key is fired, the motor module prepares to execute this movement. While this is being done, the production to find and encode the next unread note is fired. This causes the vision module to shift attention at the same time the motor module is executing its task.

Although sound feedback is a factor in piano playing, we ignore it in our model for one reason: it has been shown that even complete sound deprivation does not have a significant effect on the pianist's performance (Repp, 1998). Therefore, we decided that it is not necessary to model auditory effects to study rhythmic timing and sight-reading.

Motor Movements

Finger movements are done with the assumption that the right hand is placed with five fingers on top of the five keys and remains in this place for the entire piece — the same assumption is used in the empirical studies examined below. In the model, since the hand remains stationary, pressing a key simply corresponds to issuing a "punch" movement for the correct finger. Even though this movement was developed for modeling typing, the speed at which a punch movement is executed closely fits with that of the piano key press in the condition where all five fingers are already placed above the keys. Thus, this movement is perfect for our purposes. The fact that the punch movement does not hold down the key is not important for this study since we only represent and analyze only the onset of each key press.

Rhythmic Temporal Perception

The subjects from Truitt et al. (1997) were asked to play at the speed corresponding to 395 ms per beat. Before the experimental trials began, the subjects were allowed to practice at this speed with the aid of a metronome to fine-tune their perception of the time interval before actually doing the trials.

This beat duration of 395 ms is equivalent to 16 ticks, without noise, in the ACT-R temporal module (Taatgen, van Rijn, & Anderson (2004). While this is the ideal speed, it is subject to noise and delays due to noise in both the temporal module and the production system. Similarly, the pianists will also experience noise and delays from distractions of multitasking and from motor movements.

The parameters for the temporal buffer were kept at their default values. The tick counting and noise built into this buffer are based on research done by Taatgen, van Rijn & Anderson (2004). In order to achieve rhythmic timing, the production to reset the temporal module was set to fire whenever the tick count reached 16, signifying a beat has occurred. In doing so we are theorizing that pianists utilize the same basic timing mechanism modeled by the temporal buffer.

Timing is delayed when retrieval of the next note cannot be completed at the time the previous note is over. Logically, the model would have to wait until this is done before moving on since it cannot play a note that it has not yet recalled. It is further delayed by the fact that the production to reset timing cannot fire while another production is being fired even if the tick count has reached 16. This could cause a delay of up to 50 ms for each beat.

Sight-Reading

Through EMMA, we were able to direct the eye movements from note to note in a left-to-right manner. In human sight-reading, the eyes will not always act optimally; from time to time they may wander ahead or lag behind. These saccades could simply be a result of curiosity or boredom. A more meaningful interpretation of the far look-ahead is getting the overall structure of the music being played. With the exception of the notes already played, regressive saccades could be used for reinforcement of the notes read in memory. In fact, most studies on saccadic eye movements in sight-reading ignore the regressive saccades because they cannot be easily explained. It should be noted that more skilled pianists have less saccadic eye movements in comparison to the less skilled ones (Goolsby, 1994a). Therefore, we can infer that as a pianist becomes more skilled, the sight-reading ability converges to optimality where the eyes look more at the notes for encoding and less on random locations. It is this skilled performance that we are attempting to represent in our model.

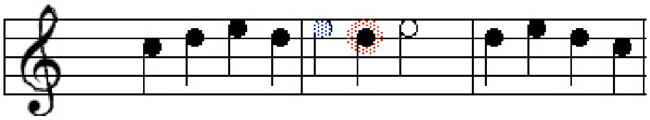

Figure 1: Sample model screen with indicators of visual attention (dotted ring on sixth note) and eye-gaze (dotted circle on fifth note).

EMMA performs eye movements from left to right as shown in Figure 1. The dotted ring indicates the location of attention and the dotted circle indicates the location of eye gaze. Visual attention is first directed to a note. Afterward, the actual eye-gaze moves in a saccadic manner toward that note as well.

The distance of look-ahead is limited by two major factors: memory and the amount of free time available for encoding the notes. A study of look-ahead in sight-reading by Furneaux & Land (1999) found that skilled pianists are capable of reading ahead by about 1.3 seconds in total note duration. However, this is under a slow playing speed approximately from 750-909 ms per beat. Under a fast playing speed of about 357-500 ms per beat, the look-ahead is about 0.7 second. From this, we can draw the conclusion that a pianist can effectively store the note information in memory for about 1.3 seconds. With a fast playing condition, this memory capability does not change, but the time allowed for encoding notes is more limited, thus leading to a decreased look-ahead. In accordance with this, we allow our model's read-ahead limit to the maximum duration of 1.3 seconds.

The terms fast or slow playing, in this case, is relative. It depends on the complexity/difficulty of the musical piece. The pianists from Furneaux & Land's study (1999) had to play two-handed music with two staves. The human data (Goolsby, 1994a; Truitt et al., 1997) and our piano play model are based on one-handed piano playing with just one staff. Hence, the playing speed of 395 ms per beat in our case is considered slow because of the simplicity of the musical piece.

Although the small differences in distance between notes have little effect on eye movements, we drew our music based on the conventions used in Truitt et al. (1997). Quarter notes are drawn at every 23 pixels on the staff. Half notes are drawn at every 46 pixels and so on. This is not necessary for the overall performance in piano play. However, it permits us to compare the read ahead distance of the model and the data from the research.

Model Simulation and Data Analysis

For the sake of consistency, we chose to use an adaptation of a musical piece from Bartok (1940), as Truitt had. This piece serves our model well in its one- (right-)handed piano playing with five fingers directly above five notes. Furthermore, in choosing real music, we do not have to compromise structure for simplicity. The model performed the whole music piece over 100 independent runs. The runs were kept independent to maintain consistency with the empirical studies of sight-reading. Essentially, the pianists either had not seen the music piece they are playing or only seen it once before (Truitt et al., 1997).

Performance

The piano performance is measured by the duration of the note presses. Since the subjects from Truitt's study (1997) were instructed to play at the speed of 395 ms per beat, our model was set to play at the same speed. In this case, note durations are measured by inter-onset intervals – the duration of a note press is calculated from the time that note was pressed to the time the next note is pressed. Therefore, the duration of the last note is ignored in calculating the note duration, and the last measure is ignored in playing time calculation.

Table 1 shows the playing time from our model and that of the pianists from experimental study (Truitt et al., 1997). Overall the model fits the human data very well. Both sets of data show that the speed at which the music played is slower than the set speed of 395 ms. We can infer from this that the playing speed for both the model and the pianists influenced by some number of delays. An important key to validating our model's productions is to establish whether or not the mechanisms delaying the model's playing speed are the same as those affecting the performance of pianists. It is reasonable to state that both the pianists and model's playing time is delayed by unfinished retrieval, since music not retrieved cannot be played. However, from the difference in duration of the quarter notes, we can see that the delays during the model's performance are not identical to that of pianists' performances. This difference could be due to the fact that a trained pianist might be able to execute finger movements faster than ACT-R would predict for typical human behavior.

Table 1: Performance measures, model and human.

	Model	Human
Playing time/measure (ms)	1812	1727
Duration of quarter notes (ms)	453	434
Duration of half notes (ms)	906	931
Standard deviation of quarter notes (ms)	68	69
Standard deviation of half notes (ms)	129	116

It would be naive to assume that rhythmic timing, even in trained individuals, could be perfect. For lack of more in-depth research, this is the current assumption of our model. It can be clearly seen from the table that the pianists' half note durations averaged at more than twice that of the quarter notes. Furthermore, the half note has greater mean duration with smaller standard deviation relative to that of the model. This indicates that there is not just a simple linear increase in noise but there also is time stretching for longer periods in rhythmic tasks. This stretching is unique in that the standard deviation does at the rate with the increased time interval.

The standard deviation of the note durations from the model's performance is produced by the noise from the temporal module as well as by delays in production firing. The correspondence between the standard deviations of the quarter notes is gives a positive correlation between the noise in the model and that observed in the performance of the subjects. While we cannot truly be sure if the noises are caused by the same factors for both, this is still a good indicator that the noise for timing for pianists may come from delayed production firing and noise from temporal perception similar to that of the model.

Saccadic Eye Movements

Sight-reading is analyzed through eye movements. Table 2 shows the data from our model and human data from Goolsby (1994a) for fixation duration, range, and standard deviation, and Truitt et al. (1997) for eye-hand span.

Table 2: Eye-movement measures, model and human.

	Model	Human
Fixation duration (ms)	526	377
Fixation range (ms)	171 - 1325	119 -1012
Fixation standard deviation (ms)	243	176
Eye-hand span (pixels)	67	42

Overall, the model fits the human eye-movement measures reasonably well, though not quite as well as the motor performance measures presented above. As expected, the mean fixation duration from the model is greater than that of the pianists' performance. Unlike the model, which makes only progressive fixations in a sequential manner, pianists tend to make both regressive and progressive fixations. These fixations are sometimes to random undirected locations with no apparent significance. For simple musical pieces like those in the research (Goolsby, 1994a,b; Truitt et al., 1997), one possible reason for the many saccadic eye movements may simply be due to boredom.

The distance difference between the location of the note being pressed and the note the eye is fixating on is referred to as the eye-hand span. By analyzing the eye-hand span, we get a picture of the look-ahead, or more generally "plan–ahead," of the pianists. The eye-hand span of the model is about three beats, which is consistent with the 1.3 ms read ahead of slow piano playing found by Furneaux & Land (1999). If the model were made to read notes from two staves, the eye-hand span would certainly be much smaller due to the limited amount of time between beats.

The eye-hand span of the model is about one beat larger than that found in Truitt et al. (1997). It is difficult to speculate about the reason for this difference since there were no details for the data from Truitt's study. One possible contributor is the fact that Truitt also included regressive saccadic eye-movements in her study, resulting in a lower mean value. Although, there are many studies being done on sight-reading, regressive saccadic eye-movements are still not well understood cognitively. However, with just the analysis of forward saccadic eye-movements alone, gives us a great deal of useful information such as profiles of music note processing, and plan-ahead in sight-reading.

Discussion

From the sight-reading data by Goolsby (1994a, b), we can see that a pianist generally fixates on a single note at a time. Occasionally, the gaze is directed to a small group of notes with certain basic structures, such as chords or notes are beamed together for easy readability. Like a word, each individual musical note contains essential information. However, unlike words, the vertical location of the note is also essential. Therefore, encoding of musical notes requires a different level of attention in comparison to words. This is especially true due to the fact that sight-reading is coupled with other tasks in piano play, so the level of attention is probably higher. Also, the rhythmic nature of piano playing is likely to affect saccadic eye movements.

A major finding in our work is that, using only the temporal module's default parameters with no parameter fitting, the model produced behavior that correlated well with human pianists' behavior. This is especially true of the standard deviation in duration of note presses, which is indicative of the amount of "noise" inherent by the human system. However, even when formally trained pianists perform the rhythmic task of piano playing, there seems to exist a time-stretching effect over long intervals. We will need to re-examine this effect to understand the mechanisms behind it.

One interesting potential avenue for the study of piano playing lies in predicting hand and finger movements for arbitrary complex tasks. For example, when a driver reaches to tune a radio, s/he may use any combination of fingers to press the radio control depending on their layout, reachability, etc. Whereas current cognitive models of such tasks (e.g., Salvucci, 2001b, 2005) simply assume that the driver uses the index finger for all button presses, it is more likely that s/he positions the hand to facilitate use of all fingers (or, say, the middle three fingers) as a more efficient and comfortable method of navigating the buttons. As our piano-playing model expands to include more complex hand movements and fingering, we expect that this work can also lead to more general theories of hand and finger movement that can be applied in a host of other domains.

Another interesting avenue for future work lies in the visual processing involved in sight-reading. Thus far in our work, we have assumed that the model can visually attend to and encode only a single note at a time. However, clearly pianists (and musicians in general) develop more complex visual skills that enable them to encode larger patterns of music — for instance, a chord as a vertical grouping of notes, or an arpeggio representing a sequence of individual notes. To enable our model to capture such patterns, the visual module would require extension to define these patterns, or perhaps the module could learn the patterns through repeated practice. Like the case of modeling hand and finger movements, such a model would have implications far beyond sight-reading and musical performance, and could be generalized to a variety of visual tasks with complex object patterns.

Given its complexity and multitasking nature, piano playing is an interesting and challenging domain for modeling using a general cognitive architecture. The model presented here demonstrates that the ACT-R architecture along with its associated temporal module helps to formulate a reasonable theory of this complexity and, with default parameters, provide a good fit to human data. This research also shows that, in cases such as that of piano playing, people (and models) can be trained to organize basic cognitive functions in specific ways that optimize performance in the respective tasks.

Acknowledgments

This work is sponsored primarily by National Science Foundation grant #IIS-0426674 to the second author. We thank Yelena Kushleyeva and Walt Mankowski for helpful ideas and comments.

References

Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*, 1036-1060.

Bartok, B. (1940). *Mikrokosmos: progressive piano pieces: piano solo* (Vol. 1). New York: Boosey & Hawkes.

Byrne, M. D., & Anderson, J. R. (1998). Perception and action. In J. R. Anderson & C. Lebiere (Eds.). *The Atomic Components of Thought* (pp. 167-200). Hillsdale, NJ: Lawrence Erlbaum Associates.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceeding of the Royal Society of London Series B*, 266, 2435-2440.

Goolsby, T. W. (1994a) Eye movement in music reading: effects of reading ability, notational complexity, and encounters. *Music Perception*, 12(1), 77-96.

Goolsby, T. W. (1994b). Profiles of processing: eye movements during sightreading. *Music Perception*, 12(1), 97-123.

Palmer, C. & Pfordresher, P. Q. (2003). Incremental planning in sequence production. *Psychological Review*, 110(4), 683-712.

Repp, B. H. (1998). Effects of auditory feedback deprivation on parameters of expressive piano performance. *Proceedings of the Fifth International Technological Directions in Music Learning Conference* (pp. 96-101). San Antonio: Institute for Music Research.

Salvucci, D. D. (2001a). An integrated model of eye movements and visual encoding. *Cognitive Systems Research*, 1(4), 201-220.

Salvucci, D.D. (2001b). Predicting the effects of in-car interface use on driver performance: An integrated model approach. *International Journal of Human-Computer Studies*, *55*, 85-107.

Salvucci, D. D. (2005). A multitasking general executive for compound continuous tasks. *Cognitive Science*, *29*, 457-492.

Taatgen, N. A., van Rijn, H. & Anderson, J. R. (2004). Time perception: beyond simple interval estimation. *Proceedings of the Sixth International Conference on Cognitive Modeling* (pp. 296-301). Mahwah, NJ: Lawrence Erlbaum Associates.

Truitt, F. E. et al. (1997). The perceptual span and the eye-hand span in sight reading music. *Visual Cognition*, 4(2), 143-161.