WORKING MEMORY AND THE PERCEPTION OF HIERARCHICAL TONAL STRUCTURES

Morwaread Farbood

Music and Audio Research Laboratory (MARL)
New York University

ABSTRACT

This paper examines how the limitations of working memory affect the perception of hierarchical tonal structures. Within this context, it proposes some modifications to Lerdahl's tonal tension model in order to better explain certain experimental data. Data from a study on the perception of musical tension were analyzed using regression analysis that took into account various parameters including harmonic tension, melodic contour, and onset frequency. Descriptions of how these features change over different time spans ranging from 0.25 to 20 seconds were used in an attempt to identify the best predictors of the general tension curve. The results indicate that change in harmony best fit the tension data when the time differential was between 10-12s, while other features best fit the data at a time differential of around 3s. This suggests that the memory of tonal regions is retained for a considerably longer period of time than is the case for other musical structures such as rhythm and melodic contour.

1. BACKGROUND

The presence of hierarchical structures in music has long been observed by both music theorists and cognitive psychologists. While there is general agreement as well as supporting empirical evidence indicating that these structures exist, the extent to which listeners perceive them is still under investigation. Hierarchical structures in music cover a range of different musical features such as rhythm and meter, grouping structures, and tonal structures. The presence of these structures allows for greater chunking of musical information, thus enabling an increase in short-term memory capacity.

There are a number of factors that weaken or strengthen how hierarchical tonal structures are perceived in time. These include the influence of veridical and schematic expectancies (Meyer, 1973; Jones, 1976; Boltz, 1993; Huron, 2006), the relative stability of an established tonal center and its distance from previous keys (Toiviainen & Krumhansl, 2003), and the limitations of memory in recalling key changes (Cook, 1987). This work focuses primarily on the third factor—how well listeners recall or retain the memory of a previous key. The goal is to examine the real-time perception of tonal hierarchical structures and offer a perspective that incorporates the limitations of working memory. Within this context, it proposes some modifications to an existing quantitative model—Lerdahl's tonal tension model—in order to better explain certain experimental data.

Lerdahl's model defines a formula for computing quantitative predictions of tension and attraction for events in tonal music. There are four required components needed to calculate this formula (Lerdahl, 2001):

- A representation of hierarchical event structure
- A model of tonal pitch space and the distances between chords within it
- A treatment of surface dissonance (largely psychoacoustic)
- A model of melodic/voice-leading attractions

The hierarchical component of the formula is based on the prolongational reduction described in the Generative Theory of Tonal Music (Lerdahl & Jackendoff, 1983).

The quantitative nature of Lerdahl's model has made it a convenient vehicle for music cognition experiments, many of which have provided evidence confirming the efficacy of the model (Bigand et al., 1996; Lerdahl & Krumhansl, 2007). On the other hand, the results of some studies have questioned the unqualified application of the hierarchical aspects of the model—in particular, Bigand & Parncutt (1999) concluded that musical tension was only weakly influenced by global harmonic structure and was determined more directly by local cadences. Although the influence of tonal hierarchies was essential in describing shorter excerpts in their study, the results suggested that a strict application of hierarchical structure for calculating tonal tension values does not accurately predict listeners' responses to key changes. For example, Lerdahl's model predicted that listeners would hear an entire section in a new key at an elevated tension level from the previous key while the experimental data indicated that the tension level dropped quickly after a new key was established.

The findings of Bigand and Parncutt imply that there is in essence a "reset" of sorts when a new key is established. They argue that musical events are perceived through a short perceptual sliding window where events perceived at a given time are negligibly influenced by events outside the window. In particular, short-term memory retains events 3-5 seconds in the past (Synder, 2000) and seems to correspond with this sliding perceptual window. Despite their conclusions that hierarchical structures are not significantly influential at a global level, Bigand and Parncutt still acknowledge that Lerdahl's model was the most effective of the several models they were testing. There seems to be little doubt that the influence of tonal hierarchies is essential in describing harmonic tension in shorter excerpts, particularly within a single key.

Lerdahl himself (2001) states that his model is constructed with an idealized listener in mind. In other words, his theory does not provide structural descriptions for how the music unfolds in time, but for the final state of the listener's understanding. Therefore attempting to apply his model directly to empirical data might naturally result in some discrepancies since the model is not intended to describe real-time cognitive processing of music. Yet despite this issue, the formula still effectively models harmonic tension within relatively short time spans that do not contain lasting key changes.

This paper proposes a simple modification to the theory that would bring the model more in line with real-time listening: adding a decay factor for the inherited hierarchical values. In practical terms, this means the addition of a decay factor for each upper-level branch in the prolongational reduction tree to taper the effect of these values.[1]

2. METHOD

In order to determine the possible longevity of a decay function, as well as the long-range effect of tonal hierarchies, empirical data gathered in a previous experiment measuring continuous listener responses to tension (Farbood, 2008) was analyzed to determine time-based effects of various musical features in their contributions to tension. The data consisted of the real-time responses of 33 subjects to an excerpt from a J. S. Bach organ transcription of a Vivaldi concerto (Figure 1). The excerpt was chosen for the purposes of this study because it was longer (1'03") than the other excerpts from the previous study and had prominent harmonic motion throughout. Four of the excerpt's salient musical features—harmonic tension, pitch height of the bass and soprano lines, and onset frequency—were quantified. The harmonic tension values were obtained using Lerdahl's original tonal tension model excluding the melodic attraction component. This latter component was deemed superfluous given that the pitch height of the melody and bass lines were already included in the analysis.

Figure 2 shows the prolongational reduction required to calculate the hierarchical tension values for the Bach-Vivaldi excerpt. In addition, mathematical derivatives were calculated for each musical feature—that is, a description of how each feature was *changing in time,* where the difference in time ranged from 0.25 to 20 seconds. In other words, the difference in value between each point and a corresponding point *dt* seconds in the past was computed.

As a first step, all of the feature graphs (shown in Figure 3) along with derivatives of a single time differential were used as input variables to regression analysis that attempted to fit the empirical tension judgments. This step was repeated for all of the time differentials (0.25, 0.5, 1, 2, 3, 4, 5, … 20s). The results indicated that while adding a derivative of a feature increased the overall goodness-of-fit, the best fit resulted when derivative graphs with time differentials of 3 to 4 seconds were used. Figure 4 illustrates the results for each of the time differential in terms of mean-square error.

Figure 1: Excerpt from J. S. Bach's organ transcription of Vivaldi's D Major concerto, RV 208 (BWV 594). See Figure 2 for the prolongational reduction of the excerpt as well as the calculated hierarchical tension values.

Breaking the problem down further, the regression analysis was repeated for all individual features in isolation (e.g. only harmonic tension and change in harmonic tension; only onset frequency and change in onset frequency, etc.). The results indicated that for every feature except harmony, the derivatives that were most

[1] Along similar lines, Thompson & Parncutt (1988) proposed a decay factor for echoic memory, which is at a lower perceptual level than short-term memory (< 1 second).

effective were those that described changes taking place under 3 seconds. Harmony, on the other hand, best fit the tension data when the change in time was 10-12 seconds in length (Figure 5).

3. IMPLICATIONS

The results of this preliminary study show a striking difference between optimal time differentials for harmony versus onset frequency and pitch height. This suggests that tonal hierarchies are processed in a different manner than other structures such as pitch contour and rhythm. If these results can be empirical verified with further studies, this 10-12s window also suggests that a decay factor imposed on inherited hierarchical value in Lerdahl's prolongational reduction should be applied after a minimum of 10 seconds of elapsed time. Furthermore, the decay factor should taper the value to zero at around 17 seconds in time, where the error values (see Figure 5) flatten out at a maximum.

Since these numbers are based on data for a single musical excerpt, more data is needed to verify the findings. Currently ongoing is a study that explores the memory of tonal centers when disrupted by modulations to other keys. The stimuli for this experiment are composed in a block-chord format with isochronous onsets and are designed to eliminate voice-leading and melodic connections as much as possible. This type of musical stimulus, as opposed to a complex musical excerpt, is expected to be more effective for isolating the direct effects of harmony perception.

While it is possible that constant coefficients for tonal memory decay functions can be discovered, it is also likely that there are other factors influencing the memory retention of tonal centers that are unrelated to pitch. Factors related to rhythmic and grouping structures might lessen or increase the retention of a previous key after a modulation. Such relationships would need to be explored as well.

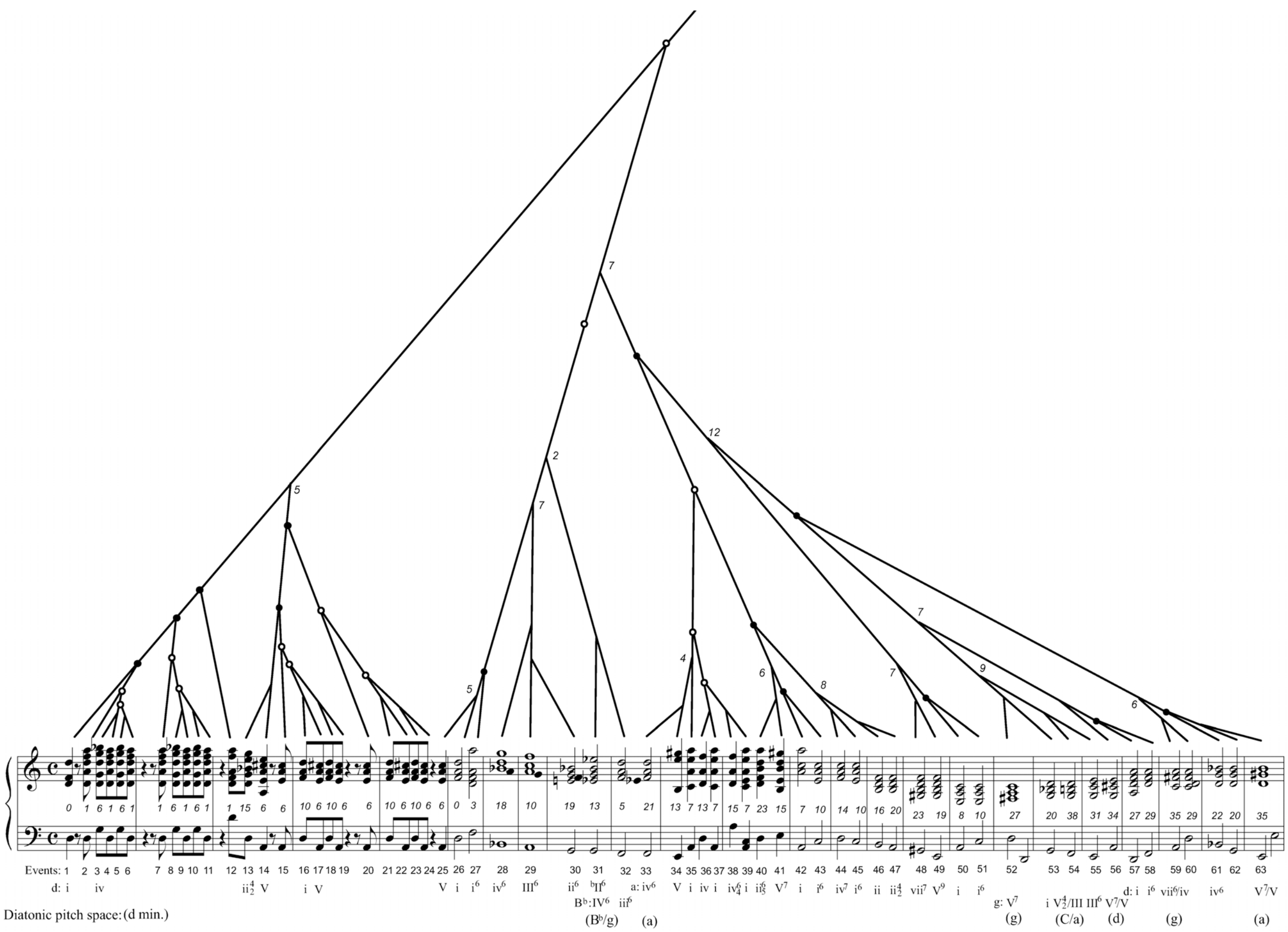

Figure 2: Prolongational reduction and hierarchical tension values for the Bach-Vivaldi excerpt.

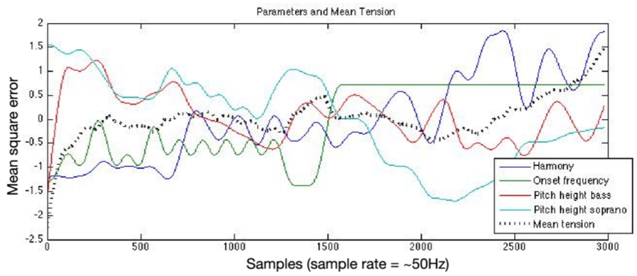

Figure 3: Graphs showing some input parameters to the regression analysis (harmonic tension, onset frequency, and pitch height of soprano and bass lines) and the mean of the subject data (perceived tension).

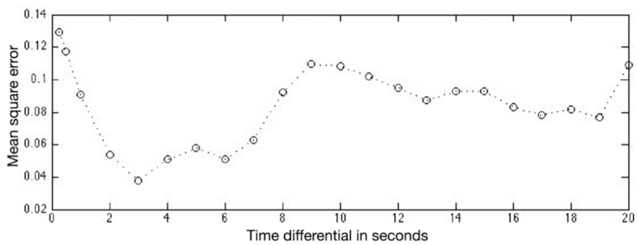

Figure 4: Graph of error values resulting from regression analysis with all parameters. Note local minimum around 3s.

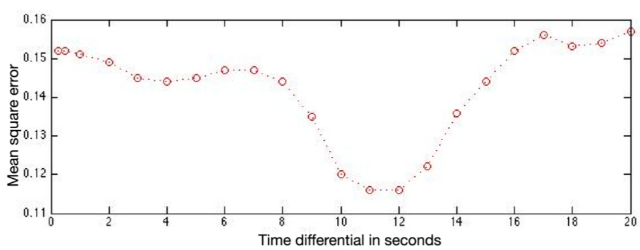

Figure 5: Graph of error values resulting from regression analysis with harmony alone. Note local minimum around 11s.

When listeners interpret musical structures, working memory keeps track of the phrase-level structures (short-term memory) and brings relevant schematic and veridical factors into play (long-term memory). The perceived key of the previous phrase might be maintained longer in working memory, but ultimately stored in a relatively abstract way—or not at all—in long-term memory. On the other hand, the immediacy of motivic elements (repeated elements with particular melodic or rhythmic features) might quickly fade from short-term memory when supplanted by new motivic elements in a following musical phrase. However, these fragments are more easily stored for later retrieval in long-term memory and recalled when similar patterns trigger the retrieval, perhaps due to their more concrete nature. Studies on real-time key perception and familiarity of musical material will hopefully provide further insight into these questions.

4. REFERENCES

Bigand, E. and Parncutt, R. (1999). Perceiving musical tension in long chord sequences. *Psychological Research*, *62*, 237-254.

Bigand, E., Parncutt, R., & Lerdahl, F. (1996). Perception of musical tension in short chord sequences: The influence of harmonic function, sensory dissonance, horizontal motion, and musical training. *Perception & Psychophysics*, *58*, 124-141.

Boltz, M. G. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception & Psychophysics*, *53*, 585-600.

Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception*, *5*, 197-205.

Farbood, M. (2008). A Global Model of Musical Tension. In *Proceedings of the 10th International Coference on Music Cognition and Perception*, Sapporo, Japan.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge: MIT Press.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, 83, 323-355.

Lerdahl, F. (2001). *Tonal Pitch Space*. New York: Oxford University Press.

Lerdahl, F. and Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. Cambridge: MIT Press.

Lerdahl F. & Krumhansl, C. L. (2007). Modeling Tonal Tension. *Music Perception*, *24*, 329–366.

Meyer, L. (1973). *Explaining Music: Essays and Explorations*. Berkeley: University of California Press.

Snyder, B. (2000). *Music and Memory.* Cambridge: MIT Press.

Thompson, W. F., & Parncutt, R. (1988). Using a memory-fade model to track the movement of musical keys. Contributed paper at the International Congress of Psychology, Sydney.

Toiviainen, P. and Krumhansl, C. L. (2003). Measuring and modeling real-time responses to music: The dynamics of tonality induction. *Perception*, *32*, 741–766.