Seek Well in Context: A New Model of Dynamic Melodic Expectation

Eric Nichols
Indiana University, Bloomington, IN 47401 USA
September 1, 2005

Purpose

During the past few decades, music cognition researchers have proposed several theories of melodic expectation. This paper reviews the basic features of models proposed by Narmour, Lerdahl, Larson, and Margulis. Next, it presents elements of a new theory based primarily on implementing Larson's theory of musical forces using methods developed by the Fluid Analogies Research Group (FARG). Finally, the new theory is compared and contrasted with key elements of existing theories to provide historical context.

Models of Melodic Expectation

NARMOUR'S IMPLICATION-REALIZATION (I-R) MODEL

Overview

Narmour's I-R model focuses on the implications generated by the movement from a note to another note. The model is ambitious in scope, encompassing such topics as both bottom-up and top-down perception, the effects of intra- and extra-opus style, and the multi-parametric nature of melody. A novel taxonomy was introduced (Narmour's "genetic code") for categorizing the types of implications and realizations in a melody.

Gestalt Principles and the Problem of Style

The I-R model concerns itself primarily with implications generated by subconscious, bottom-up processing. Narmour goes to great lengths to explain the important top-down influence of style. In broad terms, he states that for a listener the features of the input music result in a subconscious expectations *except* when the context of style modifies those expectations. The roles of top-down and bottom-up processing are clearly separated for Narmour, who insists on "the existence of two expectation systems. The top-down one is flexible and variable but controlled; the bottom-up one is rigid, reflexive, and automatic – a computational, syntactic input system." (Narmour 1990, 54) Although Narmour is careful to put both top-down and bottom-up processing on an even footing, the analytical symbols introduced in the theory focus on bottom-up expectations rooted in Gestalt principles. Style is not treated analytically in the way that note-to-note details of melody are, but Narmour does expose many inherent difficulties with incorporating style into a theory of expectation.

Interestingly, the I-R model states that both top-down and bottom-up processing occur at multiple levels of musical hierarchy. The "top" and "bottom" in "top-down" and "bottom-up", then, must not to be confused with "surface" and "deep" structures of musical hierarchy. This theme of application of the same principles at several hierarchical levels shows up in several models such as Lerdahl and Jackendoff's GTTM and Larson's Seek Well model, as well as in Schenkerian analysis. However, the cognitive basis for applying Gestalt principles to musical events at deeper levels (i.e. with a longer time span between events) may still need to be tested. Specifically, if low-level processing describes an innate universal mechanism for processing perceptual input, it remains to be shown how this same mechanism be applied to a more abstract level of input that has been generated with the help of top-down processing.

In the study of bottom-up expectations, Narmour invokes the Gestalt principles of *similarity*, *proximity*, and *common direction* as cognitive universals that apply to individual musical parameters such as intervallic motion, registral direction, and pitch. Additionally, he adds in two other hypotheses – *reversal* and *parametric scale* – to supplement those three principles. Invoking these hypotheses leads to five specific aspects of melody used in classification and description via Narmour's novel symbolic system. These aspects are:

1. Registral direction
2. Intervallic difference
3. Registral return
4. Proximity
5. Closure

Five Fundamental Principles

REGISTRAL DIRECTION

Registral direction refers to the change in pitch from one note to the next. The direction is either up, down, or lateral (no change). As long as the interval between two pitches is small, this principle states that a listener will subconsciously expect a *continuation* of pitch direction: upwards, downwards, or lateral motion is expected to continue. Large intervals, on the other hand, imply *reversal* of direction.

An interval is defined to be small if it is less than a tritone and large if greater than a tritone – the tritone interval itself is a boundary case. The model's prediction of either a continuation or reversal is not as simple as the interval size metric might suggest. Narmour's idea of parametric scale suggests that a very large interval generated a stronger implication of reversal than a moderately large interval; the same holds for very small intervals and implication of continuation. Additionally, even a very large interval may result in a "recessive implication" for continuation. This latent implication may be realized by retrospective processing. Narmour's symbology describes these retrospective implications by surrounding a standard implication symbol with parentheses, such as (**P**) or (**VP**).

INTERVALLIC DIFFERENCE

While registral direction refers to the direction of motion between pitches, intervallic difference relates to the size of the interval between pitches. The principle states that small intervals imply a continuation with similar-sized intervals. Large implicative intervals, however, imply relatively smaller realized intervals. Large and small intervals were already defined above in reference to registral direction, but here the concept of "similar-sized" intervals is also important. Narmour defines interval similarity as a difference of a minor third or less in interval size.

REGISTRAL RETURN

The typical motion of a melody away from a pitch and then back to the original pitch is known as registral return. Exact registral return describes the perfectly symmetrical melodic archetype **aba**, exemplified in patterns such as neighbor tones. Near registral return (**aba'**) occurs when the final tone is very similar to the original pitch (within two semitones). Patterns become less archetypal as they deviate more from the symmetrical case.

Recognition of registral return can make an otherwise surprising realization more expected. This idea is also used to explain "streaming" in the I-R model: a series of overlapping **aba'** processes allows the listener to split the music into multiple input streams in different registers (Narmour 1992, 352).

PROXIMITY

When the Gestalt notion of proximity is applied to pitches, the result in the I-R theory is that small intervals are more strongly implied than large intervals. Additionally, the implications generated by smaller-sized intervals are said to be stronger than for relatively smaller intervals.

CLOSURE

Closure describes how listeners break up melodies into separate perceived segments. According to the I-R model, closure in the pitch domain occurs in two cases:

1. The melody changes direction. That is, implicative and realized intervals are in different directions.
2. A relatively larger implicative interval is followed by a smaller realized interval.

Naturally, parameters other than pitch also can effect a feeling of closure. Narmour lists the others as:

1. Interruption of an implied pattern
2. Strong metric emphasis
3. Dissonance resolving to consonance
4. Short notes moving to long notes

I-R Symbols

The I-R model annotates melodies with symbol strings based on the implications and realizations present in terms of the parameters of registral direction and intervallic difference. The basic symbols apply to groups of three notes: the first two notes set up an implication based on the pitches and the interval between them, while the third note produces an interval with the second note; this interval and the third pitch may realize or deny aspects of the implication of the first two notes.

The most straightforward melodic structure is *duplication,* indicated by the symbol [**D**]. Duplication is the simple repetition of pitch. The I-R theory considers duplication to be an implicative structure with more importance than in, for example, Schenkerian analysis, where repeated notes are commonly reduced to a single pitch.

The next simple melodic structure is *process* [**P**]. Process occurs when both direction of motion and interval size are similar. Process is a nonclosural pattern that implies further continuation of the process. Narmour uses square brackets to indicate the extent of the structure: the opening bracket starts the process structure while the closing bracket indicates termination. The brackets should not be confused with parentheses: (**P**) denotes a retrospective process structure.

The opposite of process is closural reversal [**R**]. Reversal occurs when a large interval is followed by a small one and the registral direction changes (up followed by down or lateral, lateral followed by up or down, or down followed by up or lateral).

If the direction of motion to the third note in a structure changes while the intervals remain small and similar, the result is an *intervallic process* [**IP**]. Here the interval size is in-line with expectation while the direction involves reversal. The different possible combinations of interval similarity and registral directions produce the following structures indicating partial realization and partial denial of expectation (Narmour 1990, 6):

- [**IP**]: *intervallic process*: similar small intervals, different directions

- [**VP**]: *registral process*: small to large interval, same direction
- [**ID**]: *intervallic duplication*: identical small intervals, different directions
- [**IR**]: *intervallic reversal*: large to small interval, same direction
- [**VR**]: *registral reversal*: large interval to larger interval, different directions

All of these may be written with parentheses, such as (**VR**), to indicate retrospective realizations. Thus, in total there are sixteen symbols intended to represent how listeners respond to basic melodic structures. A few additional symbols are also used, such as [**M**] to represent a monad (a solitary note) and an interval size such as [4] to represent a dyad (a simple pair of notes). The symbols for registral return and near-registral return discussed earlier (**aba** and **aba**' are also used.) The I-R hypothesis that this handful of symbols can represent a listener's encoding of melody when chained together (such as in **ID-IR-VP-P-IP-VR-D-R-M**) is provocative and remains to be tested (Narmour 1990, 6-7).

Tonal Pitch Space and the I-R Model

Narmour's focus is on innate, bottom-up Gestalt laws, so it is natural to wonder how the concept of tonal scale step fits in to the theory. For example, although C-G-C# and C-G-D are both examples of near registral return, the former seems less expected if a C major context has been established. The I-R theory explicitly states that it supplements conventional notions of tonal pitch space and provides another dimension to pitch relation that should be considered. This is a case where learned top-down style interacts with bottom-up perception. Rather than defying accepted notions of pitch space, Narmour intends to contribute to a more complete account of pitch perception.

After making a point to separate the function of scale steps in harmonic vs. melodic contexts, Narmour introduces several categories of scale steps in order to introduce a parametric scale for melodic implication with respect to scale step. Degrees 1, 3, and 5 are called *goal notes* (GN), degrees 2, 4, and 6 are *nongoals* (NG), and the leading tone is a *mobile note* (MN). The I-R theory already states that small intervals imply continuation while large ones imply reversal. Incorporating scale step affects the strength of the generated implication. Within a clear tonal context, the more differentiated the two tones of an interval are with respect to the scale step categories above, the stronger the implication. Thus an interval moving from 7 to 1 (MN to GN) generates a stronger implication than 2 to 4 (NG to NG) or 1 to 3 (GN to GN). Narmour enumerates all nine possibilities: the most open (i.e. most implicative) combination is GN to MN, while MN to GN is the most closed. Within each category, the particular scale degree chosen also affects the generated implication strengths.

Narmour does not place much emphasis on tonal pitch space in the model, though, and states that because tonal style is a learned top-down schema, it deserves no more preferential treatment than other parameters affected by learning (Narmour 1992, 85). The focus of the I-R model is on bottom-up processes.

Summary

The I-R theory presents a wealth of details about low-level music cognition, especially on the note-to-note level. Although some researchers have shown that parts of the theory may be simplified without loss of predictive power (Schellenberg 1997), the philosophical foundation is especially useful. Margulis (2005) provides a useful critique of the theory that points out how the I-R symbols for basic melodic structures provide taxonomic categorization but do not clearly explain the expectations generated by a melody. Similarly, the question of how expectation relates to affect is mostly ignored by the theory. Finally, melodic hierarchy is invoked by Narmour but the focus remains on local note-to-note relations.

LERDAHL'S MODEL OF ATTRACTION

Lerhadl's model of tonal pitch space is not exclusively a model of melodic expectation, having perhaps more to do with harmony than melody. However, some elements of the model do produce quantitative predictions of melodic motion. In this model, the tones of the scale exist in a hierarchy of alphabets, after the idea of Deutsch and Feroe (Lerhadl 2001, 47). Basic tonal space, by this account, has five levels. In C major, the levels include the following notes:

1. C
2. C, G
3. C, E, G
4. C, D, E, F, G, A, B
5. C, Db, D, Eb, E, F, F#, G, Ab, A, Bb, B

That is, the levels successively add the tonic, fifth, tonic triad, diatonic octave, and chromatic scale.

The theory provides metrics for calculating "distances" between two notes based on the number of hierarchical levels and horizontal steps between the notes in this alphabet hierarchy. Next, formulas are provided that define distances between two chords in different tonal contexts. Later, the theory gives a way to calculate harmonic and melodic forms of tension.

Melodic tension (pitch instability) depends on the *anchoring strength* of the two pitches involved and the (squared reciprocal) distance between the pitches in semitones. Anchoring strength derives from the hierarchical level on the pitch in a space similar to the alphabet hierarchy above (in C major, C has anchoring strength 4; E and G both have 3; D, F, A, and B have 2, and the chromatic pitches have strength 1.) The amount of attraction is calculated by dividing the anchoring strength of a possible second note by the strength of the first note and dividing by the squared distance between notes. A pitch is thus attracted to a nearby pitches with high anchoring strength. The formula for melodic attraction provides a quantitative value for the attraction between any pitch and a possible following pitch. This formula for attraction is the analogue in this theory to the expectations generated by Narmour's I-R theory (Lerdahl 2001, 170).

Margulis provides a critique of the formula, pointing out that registral direction is ignored, semitone movement yields unreasonably high attraction values, and pitch repetition is ignored completely by the model (Margulis 2005). Margulis incorporates an extended version of Lerdahl's model of attraction in her own model. Larson's single-level Seek Well model also includes a similar parameter of attraction – magnetism – but it, too, is augmented with other model components.

MARGULIS' DYNAMIC MODEL

Margulis developed a model of melodic expectation that included elements of both Narmour's and Lerdahl's models. Both tonal pitch space and innate bottom-up processing are given significant status in the model. The model provides solutions to issues Margulis pointed out with each in her critiques thereof. For example, it explicitly describes how expectation connects to the experience of affect and tension, how to deal with repeated notes, and how hierarchy can be used formally to include more than the two preceding notes in the generation of expectations. The model is composed of five separate components: stability, proximity, direction, mobility, and hierarchy.

Model Components

STABILITY

Stability of melodic events is calculated based on the tonal context. Rules adapted from Lerdahl (2001) select a chord and current key for each pitch event. Based on the tonal function of each pitch, a stability rating is assigned. These are numerically similar to the anchoring strengths discussed earlier. However, they are more sophisticated because they are based on the current tonal context. For example, several exceptional cases such as augmented sixths and Neapolitan chords are enumerated in the model to improve the quality of stability predictions.

PROXIMITY

Just as in the I-R model, the principle of proximity states that listeners have higher expectations for pitches nearby in frequency. This model gives a numerical proximity rating to pitches based on the distance in semitones away from the preceding pitch. Margulis selected the particular numerical values by hand based on results from various studies and intuition.

DIRECTION

The I-R model states that small intervals imply continuation of direction but large intervals imply reversal. Margulis incorporates this idea to generate particular expectations (for continuation or reversal) along with the strengths of each expectation. These expectations are based on the interval size in half steps. Instead of prediction strength being a simple linear function of interval size, Margulis uses data from Schellenberg to suggest a particular nonlinear mapping from interval size to prediction strength.

MOBILITY

Although many theories ignore the possibility of repeated notes, Margulis includes a mobility parameter that increases the expectation value for melodies that move to different pitches. Repeated notes lead to a mobility penalty whereby the stability and proximity scores are multiplied by 2/3 to encourage motion. Margulis notes that this parameter was explicitly added to reduce the strong expectations for repetition otherwise produced by the model.

HIERARCHY

Without the concept of hierarchy, the amount of expectation for a pitch to follow another pitch is given by the model as **stability × proximity × mobility + direction**. Hierarchy is incorporated in the model by generating a time-span reduction of the input and then applying the formula above to each level of the hierarchy. The final expectancy value for each pitch is given by a weighted average of the values at each level. Margulis selected the weights in this formula by hand; the surface level has a weight of 15, other levels of no more than two second duration have weights of 5, remaining levels with less than a six second duration have a weight of 2, and no levels are considered with a time span longer than six seconds. In the model, the time-span reduction is generated via an implementation of preference rules from GTTM (Lerdahl & Jackendoff, 1983), augmented with an additional preference rule.

Tension and Affect

This model defines three different types of tension based on calculated expectancy values: surprise-tension, denial-tension, and expectancy-tension. The first two types are calculated for a third event in a series based on the two prior events. For example, the notes A-B in a C major context may set up an expectancy for continued upward motion to the tonic C, but if the pattern continues A-B-F# the third event yields a high value for denial-tension. Expectancy tension, on the other hand, is calculated when looking forward to a potential expected future event. Thus, in our example, the second event, B, may have a high value for expectancy tension if the tonic C is strongly expected.

Surprise-tension describes how unexpected an event was – thus its value is inversely proportional to the expectedness of the event. An event that was not very expected at all results in high surprise. Margulis associates surprise-tension with an experience of “intensity” and “dynamism”.

Denial-tension is proportional to the difference between the expectancy value for the event and the maximum expectancy value over all possible notes. Thus denial is stronger when there is a very strongly expected event or when the actual event is quite unexpected in comparison with the maximally expected event. Denial results in feelings of “will, intention, and determinedness.”

Expectancy-tension is proportional to the expectation strength of the maximally expected event to follow. Its calculation is thus quite similar to that of denial tension except that it is forward-looking. Expectancy-tension is associated with feelings of “yearning” and “strain”.

Margulis uses the three tension formulas to generate graphs of each type of tension versus time for musical pieces. It might be informative to analyze such graphs for a large variety of music to look for possible generalized patterns in tension structure. More importantly, these measures may inspire more research into connections between expectancy and affect. If these formulas prove useful, one can imagine additional formulas based on expectancy to predict more specific types of emotional responses.

Summary

This theory has several attractive features that are good models for future theories of melodic expectancy. First, it is quite simple to describe even though it is sophisticated. This stems from the modularity of the model: it has clearly separated components that make separate calculations of expectation strengths. These individual factors are combined via simple arithmetic operators. This leads to the difficulty of weighting the separate factors – Margulis accomplished this using intuition and hand-tuning of weights. Another positive feature of the model is the generation of quantitative predictions of a field of expectations at each moment in a piece. The attempt to quantify some types of affect is also a major strength of the model.

Larson’s Seek Well Model

The theory of musical forces states that “we tend to hear music as purposeful action within a dynamical field of musical forces” (Larson 1993a), making an analogy between physical motion through space and the perceived "motion" of a melodic line. The three forces involved are musical gravity, magnetism, and inertia.

Three Musical Forces

Gravity refers to a tendency of notes heard above a stable platform to descend to that platform. For example, given the rising line C-D-E in C Major (Figure 1a), gravity would suggest a continuation back down to the stable tonic: C-D-E-D-C (Figure 1b). After the alternate beginning C-D-C (Figure 1c), however, gravity does not continue to pull the line down past C, because C has been established as a stable base.

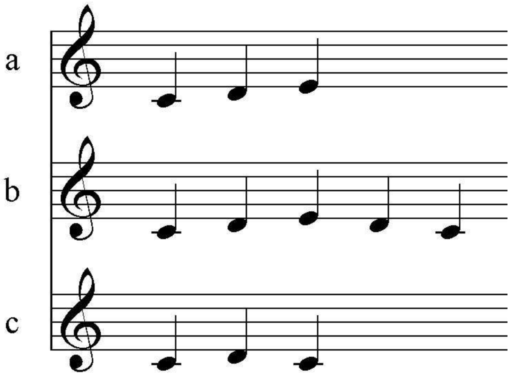

Fig. 1

Magnetism refers to the perceived attraction between notes of unequal stability. For example, in the ascending octave that stops on the leading tone, C-D-E-F-G-A-B (Figure 2a), we feel the strong magnetic force of the stable upper octave C "pulling" on the unstable B, leading to the expected completion in Figure 2b. As with physical magnetism, the strength of the force is inversely proportional to the distance between two pitches. The B would also be attracted downward to the stable G, but the magnetic pull of the upper C is stronger because it is closer as measured in half steps.

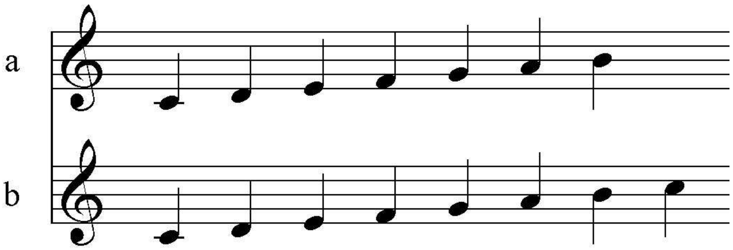

Fig. 2

Inertia is the tendency of a musical pattern to continue "in the same way". A simple example is the tendency of a moving line to continue moving in the same direction. Given the same beginning as in the gravity example, C-D-E (Figure 1a), inertia suggests that instead of falling back down to C, the melody may continue rising through F, G, A, etc. as in Figure 3. Musical inertia corresponds to the physical statement "bodies in motion tend to stay in motion". Another example of inertia is an Alberti bass, which tends to continue its characteristic pattern once set in motion.

These three forces act continuously on musical lines in a dynamically shifting musical context. A significant part of this context is provided by pairs of *reference alphabets* and *goal alphabets* (Larson 2004). We say that musical lines move through a reference alphabet between notes of a goal alphabet. Reference alphabets consist of collections of notes that make up musical lines, such as the C Major scale in the examples above. A goal alphabet is a subset of an associated reference alphabet that marks stable points of arrival, such as members of the tonic triad in these examples.

It is illustrative to reconsider the ascending octave line (Figure 2b) in the context of all three forces. The melody starts out at rest on C, the stable base. An external force starts the motion by "pushing" the melody up to D. The external force subsides and the melody begins to be tugged in various directions. Gravity exerts a constant force pulling downwards, back towards C. Magnetic forces are pulling with equal strength up to the E and back down to C. The inertia of the initial push keeps the melody going upwards to E, overcoming the force of gravity. At this point the magnetic pull of G becomes prominent as it is closer than the lower C. Magnetism accelerates the melody up through F towards G, and once reaching G it sails past due to inertia. However, the weight of gravity is still noticeable, as is the strong magnetic pull of G once the melody ascends to A. Only inertia enables the melody to rise a bit further to B. Once reaching B the magnetic force of the upper C is irresistible and overcomes both gravity and the magnetism of G, pulling the melody all the way to C.

The dynamic push and pull of the forces is significant in that it acknowledges some of the complexity of how we listen to music. The flexible interactions of the forces can also provide explanations for alternate melodic continuations. In the ascending line above, the melody might have changed direction before reaching C; perhaps the initiating "shove" was not strong enough to build up inertia to carry the melody past F, A, or even the initial D (Figure 3a,b,c below). In each of these cases, the inertial ascent "ran out of steam", overcome by gravity, magnetism, or both. Musical forces can also explain more extended motion such as that of Figure 3d: the initial ascent to A is followed by a descent via gravity and magnetism back down through G, a brief lift back up due to the magnetic pull of G, and a final giving-in to gravity to return to the stable platform of C (after briefly dipping below the base due to inertia).
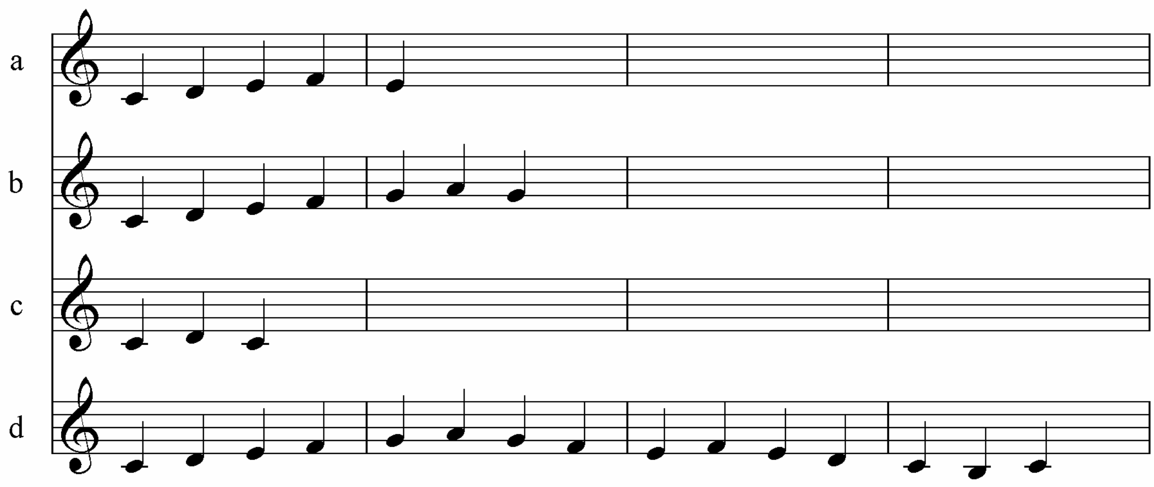
Fig. 3

Computer Models Incorporating Musical Forces

Larson's computer models of melodic expectation illustrate the theory by generating explicit predictions of melodic continuations. The *single-level* model is based on the three forces and built-in pairs of reference and goal alphabets. The *multi-level* model additionally uses information provided by an embellishment hierarchy (a simplified Schenkerian analysis).

SINGLE-LEVEL MODEL

The single-level model takes a melodic beginning as a cue and produces predicted completions of the cue based on gravity, magnetism, and inertia. For each of the three forces, the model analyzes the pitches in the cue to choose pairs of reference and goal alphabets from a predefined list. Several rules help ensure that the alphabets chosen make musical sense in the context of the force under consideration. For instance, when making a gravity or magnetism prediction, the cue must end on a pitch that is present in the reference alphabet but not in the goal alphabet (that is, the final note of the cue must be heard as unstable so that there is an impetus for continuation). Next, the model examines the final notes of the cue and produces predictions based on these notes and the pairs of alphabets under consideration. These predictions involve motion through the reference alphabet to a stable member of the goal alphabet, providing a sense of completion. The predictions are assigned probabilities according to the strength of the forces as defined by the theory. Thus, the output from the model is a set of different predictions and associated probabilities. Finally, if desired, the resulting predictions can be fed back into the system as new input so that it can predict what will follow each of those possible continuations, generating more extended predictions.

MULTI-LEVEL MODEL

The theory of musical forces applies not only to the surface-level notes in a melody, but also to deeper levels of an embellishment hierarchy (a Schenkerian analysis). In Figure 4 there are two levels in the hierarchy. Staff A represents the deepest (background) level, while staff B depicts the foreground level, made up of the notes actually heard by the listener. While there are only two levels of hierarchy in this example, there could be several middleground levels in more extended examples. Any discussion in this paper of relations between the background level and the surface applies equally well when the background or surface is replaced by a middle level.

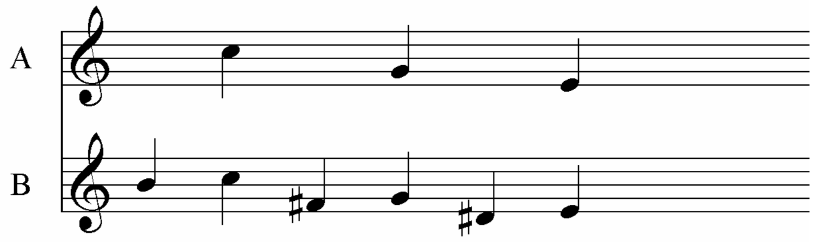

Fig. 4

The current implementation of the multi-level model requires the user to provide not only the surface notes of the melodic beginning and the key of the melody, but also a specific type of Schenkerian analysis. In this case, both staves A and B would be supplied to the computer, along with a description of the embellishment function of those notes not present in the background level. For instance, the user might describe the initial note B as a "prefix chromatic neighbor" to the following C. This type of analysis differs from a typical Schenkerian analysis in its specificity: the precise embellishment function is explicitly described in the analysis. At the same time, the theory also claims that the analysis must be flexible in describing the embellishment function. For instance, in some musical contexts a description such as "suffix diatonic lower neighbor" might be appropriate, while in other cases the description might be generalized to "suffix diatonic neighbor".

The multi-level model considers each level as an individual melody line and generates a completion of each line. In this example, the *multi*-level model would begin a prediction by applying the *single*-level model to staff A. The single-level model would generate middle C as a likely final note, due to the influence of both inertia and gravity. (The other possibility, discussed below, would be a return to G based on the magnetism between E and G.) Here the reference alphabet in use is the C Major triad, with a goal alphabet made up of only the tonic and dominant. Figure 5 demonstrates the partially completed prediction after this step.

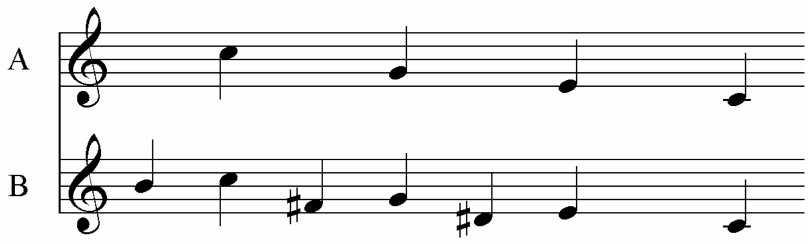

Fig. 5

Having generated a completion for the background level, the multi-level model proceeds to complete the surface level. This step depends on the particular analysis supplied to the system. In the current example, if all the embellishment notes are described as "chromatic lower-neighbor prefixes", the force of inertia indicates that to continue "in the same way" the final C should also have this type of prefix embellishment. The word "chromatic" indicates that the alphabet to be used for the prefix note is the chromatic scale (as opposed to the triadic alphabet used in the background level). Hence, the predicted embellishment note is the note B as shown in the analysis in Figure 6. In this diagram, the slurs connect embellishment notes (shown without stems) with notes present at a deeper level in the analysis.

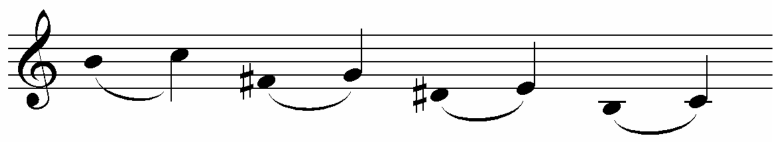

Fig. 6

A Dynamic Seek Well Model

This paper presents another model of melodic expectation, which derives from Larson's multi-level Seek Well model. Ideas from each of the other models are incorporated as well, with a goal of modeling melodic expectation at a deeper subcognitive level. The architecture to provide this level is detail is quite different from the models discussed previously and requires some explanation before the specific model components are presented.

FARG ARCHITECTURE

Douglas Hofstadter and FARG developed the idea of studying cognitive mechanisms via a particular method involving *microdomains* (Hofstadter 1995). Microdomains are restricted domains typically involving a much smaller problem space than typical problem domains. Larson (1993b) uses the term *creative microdomain* to further require that the process of devising solutions must arguably involve creativity (as opposed to brute-force search, for instance.) A particular creative microdomain developed by FARG called *Copycat* (Hofstadter, Mitchell 1994) involves letter-string analogy problems such as:

"If ***abc*** goes to ***abd***, what does ***xyz*** go to?"

This problem is purposefully restricted such that the solution involving "wrapping around" from '***z***' to '***a***' is disallowed. Creativity is implicit in interesting responses to this question. Indeed, many answers are possible besides the forbidden ***xya***, such as ***xyd, xyy,*** and the surprising ***wyz***. Creative microdomains generally do not admit a single correct response, as the process of coming up with an answer is the primary object of study. Other FARG microdomains include *Seek Whence, Tabletop, Jumbo,* and *Letter Spirit*, with subjects ranging from number sequences to font design (Hofstadter 1995).

FARG has devised architectures to model creative solutions to problems in these microdomains. Central to these architectures is the notion of simulating parallel subcognitive processes via a community of specialized agents. Recalling ideas from *Society of Mind* (Minsky 1988), these agents perform a *parallel terraced scan* in which solutions are formed, rejected, modified, and eventually settled upon in a stochastic, flexible, simulated community working environment (Hofstadter 1995).

THE CREATIVE MICRODOMAIN *SEEK WELL*

Steve Larson and FARG developed the creative microdomain *Seek Well* to study melodic expectation (Larson 1993b). To reduce the potentially overwhelming complexity of the domain while retaining key features of interest, *Seek Well* involves melodies in the classical tradition (i.e. Western tonal music) conforming to the following list of restrictions:

- Only one note sounds at once; the sound represented is monophonic
- All notes have the same duration
- All notes sound the same except for pitch (dynamics, articulation, etc. are not involved)
- Rests are not allowed (notes come immediately one after another until the melody ends)

Figure 1a is a typical melodic beginning in the *Seek Well* domain. Larson studied typical responses to this cue by asking listeners to sing a completion of any length (Larson 1997). Two common continuations are shown in Figures 7b and 7c. Note how the response 1b implies the listener heard the cue in C Major, while response 7c implies F Major.

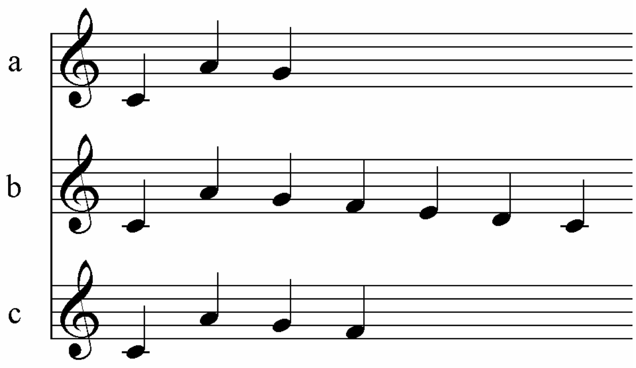

Fig. 7

KEY MODEL FEATURES

The desired extended version of the multi-level model is conceived as a recasting of the model in terms of a dynamic simulation. In particular, the influence of the three musical forces at different levels is transformed into a set of behaviors of simulated subcognitive agents (**Workers**). Features of additional models are also naturally implemented as unique worker types. The proposed system draws inspiration from other FARG models that function via a parallel terraced scan. The simulation involves a collection of workers working together to 1) build up a shared representation of a musical cue input to the system and 2) induce a natural expected continuation of the cue. The workers act in parallel to explore various representations and continuations until the system judges the result to be satisfactory.

Amongst the myriad features this architecture could attempt to provide, the focus is on several particular features motivated by the strengths and weaknesses of previous models of melodic expectation discussed previously. These include the concepts of musical forces, tonal pitch space, meter, bottom-up processing, Gestalt principles, top-down processing, tension, and musical memory.

Musical Forces

Larson's three musical forces are incorporated in the system by providing workers that specialize in each force.

Gravity workers are the most straightforward: they respond to and predict downward movement. Note that in some cases these workers will be called on to analyze existing structures to notice the pattern of downward motion. In other cases the workers may predict later events by simply predicting motion downward towards a stable base.

Magnetism workers can make multiple predictions based on strength of magnetism between a given note and a predicted successive note. Predictions are generated randomly but with a distribution correlation with magnetic strength. Magnetism depends on the harmonic analysis performed by other workers in the system.

Inertia is a more complicated force to implement than the other two. Inertia due to continued motion is a simple case that works much like gravity; this type of inertia is similar to Narmour's principle of registral direction. Generalized inertia, however, refers to the continuation of a pattern, where pattern can be defined in many different ways. For example, expecting the motion of an Alberti bass to continue is an instance of inertia. Likewise, inertia also predicts that a melodic sequence will continue. These types of inertia require the cooperation of workers assigned to detect grouping structure so that the extent of an inertia pattern can be determined. Melodic sequence behavior also makes use of musical hierarchy. For example, in a melodic fragment that sequences downward by step, the linear stepwise descent is predicted to continue by simple inertial prediction at a higher level, while the particular pattern of the melodic fragment is predicted to continue by inertia at the surface level. Thus, inertia workers must be operating on at least two hierarchical levels to make correct sequencing predictions.

Tonal Pitch Space and Meter

Incorporating the force of musical magnetism requires a notion of tonality implemented via an alphabet hierarchy. Determining which alphabet should be in use at any given time at each hierarchy is a difficult problem. Larson's original Seek Well model gives simple rules of thumb for choosing alphabets; similarly, Margulis uses preference rules to pick the current tonal context. This model attempts to determine the current context via the same architecture used for the rest of the system: workers are assigned to the analysis task.

We hypothesize that harmonic analysis will be more successful when it works in tandem with grouping and metric structure determination. Thus, in contrast to other models, the key and meter will not be given to the system as input. Instead, all of these will be determined in parallel with successful exploration in one area influencing the results in another area. Note that there will be a top-down mechanism in place to favor deciding on a key and entraining to a meter quickly (within the first several notes of a piece).

Bottom-Up Processing (Parallel Terraced Scan/Gestalt Principles)

The workers representing musical forces certainly represent bottom-up processing. Additional workers will function as Gestalt agents, responding to and making predictions based on the musical Gestalt laws given by Narmour. Note that Narmour's detailed classification system does not necessarily need to be built into the model because the I-R symbols stand for combinations of Gestalt continuations or denials in separate musical parameters. These components will be part of the model but the particular I-R symbols are not necessary (although they might be generated to produce system output in a manner more easily interpretable by those versed in those symbols.)

Top-Down Processing

As in other FARG models, there will be a Metacontroller component to provide top-down processing. That is, when the system has a set goal (such as to determine the key signature and meter), workers will be directed towards achieving the goal. Top-down processing will also dictate the activity level of the workers by monitoring system *temperature*.

Narmour's notion of top-down processing also involves the influence of learned style. In our model, tonal pitch relationships are dealt with using magnetism and other tonal pitch space ideas discussed earlier. However, the general influence of style is not present here.

Tension

Margulis' formulas for calculation of tension will be adapted for use in this model. Workers will at a minimum notice the degree of tension and influence the activation of tension in long-term memory. In the future, tension might be used to more directly affect predictions.

Musical Memory and Tempo

A key goal of this work is to model temporal effects in music cognition. The different types of memory used in human cognition (i.e. working memory vs. long-term memory) are associated with different time spans (Snyder 2000). In this model, the workers represent the action taking place in working memory. Long-term memory is stored separately. Short-term echoic memory also has a place in the model: very recent musical events are allowed to influence the workers more strongly than events in the past. This part of the model should be based on the tempo of performance. Even though the microdomain requires all notes to have the same duration, if this basic tempo changes it has the potential to affect cognitive processes.

Long-term memory in this model refers to two related ideas. First, there are built-in concepts in the Conceptual Memory (such as tension, forte, or major third) that have relatively fixed definitions and can be used by workers as necessary. However, conceptual memory can also be altered by workers to simulate changes in long-term memory. A common change in long-term memory is to elevate a characteristic musical fragment from the input to the status of a motive or theme. Motives (such as the distinctive opening four notes of Beethoven's 5th Symphony) must be stored in long-term memory, even during the initial hearing of a piece, because listeners notice repetitions of the motive even after a long time has passed from the beginning of the piece. It seems unlikely that motives persist in working memory throughout a listening session; hence they must be moved to long-term memory. This model will provide for the movement of perceptually significant musical structures from working memory to long-term memory. This also provides the interesting opportunity of having a model that becomes acquainted with a particular style over time as long as the long-term memory is preserved between program runs.

SYSTEM ARCHITECTURE

Although some names and ideas of the main components below are borrowed from other projects (Hofstadter 1995), each of those projects uses the components in a domain-specific manner. Likewise, in this case certain details are unique to the *Seek Well* microdomain. The main components of the architecture are the **Metacontroller**, **Workspace**, **Worker Collection**, and the **Conceptual Memory**.

Metacontroller

The Metacontroller component coordinates the various parts of the system to provide overall control. It maintains a record of the system *temperature*, which indicates how satisfied the system is with its current predicted completion. When the simulation starts running, the Metacontroller directs the other parts of the system to build up a representation of the given cue. This representation is built and stored in the Workspace. Then the Metacontroller directs the system to generate possible completions of the cue. It stops the process when the temperature reaches a low enough value.

Workspace

The Workspace represents the simulated cognitive space where an internal representation of a musical cue is generated, stored, and modified. When the system starts running, the representation includes only the given cue, but as the system runs it builds up an embellishment analysis of the cue and eventually the expected continuation. The Workspace is intended to evoke an image of a bustling construction area: subcognitive agents are busy working in the Workspace to construct an embellishment hierarchy in the same manner that a group of construction workers might work together to build a multi-story house. However, the analogy must be extended to allow for flexibility: instead of working from blueprints, imagine the construction workers building the house quickly in an improvisatory manner. These workers have no qualms about tearing down a wall if they decide it was badly planned. The musical working agents build up structures such as groups of notes, notes at deeper hierarchy levels, or analytical links between notes. They would also tear apart these structures when necessary to try different ideas (i.e. major changes would be made if the system temperature remained too high).

Worker Collection

The Worker Collection stores a large collection of workers that each have individual tasks to perform. Depending on the demands of the musical context, the Metacontroller will select certain Workers from the collection to join the construction process in the Workspace. Some workers are intended to focus on bottom-up tasks such as analyzing individual notes, while others deal with top-down ideas such as considering high levels of the embellishment hierarchy. Certain workers will analyze groups of notes to determine the key or meter of a cue. Others will determine whether certain notes are more structural or more embellishing in function. Finally, some workers will be involved with constructing actual predicted continuations: at each level of the hierarchy, there may be different workers building continuations based on the forces of gravity, inertia, magnetism or Gestalt continuation principles.

Conceptual Memory

The Conceptual Memory is a fixed representation of the musical knowledge of the simulation and acts as a source of long-term memory. Examples of the knowledge in the memory might include the concept of interval size, stepwise vs. disjunct motion, the tendency for melodies to build to a climax, understanding of sequence, known pitch alphabets such as the major and minor scales, simple harmonic schemes such as **V→I,** and motives such as the Beethoven's Fifth opening mentioned above. Workers will use knowledge in the Conceptual Memory to perform their tasks. Thus, changes in the Conceptual Memory will influence worker behavior. Conversely, workers may modify knowledge stored in the Conceptual Memory. For instance, workers noticing interesting features of the musical cue may cause concepts in memory related to those features to become more active. In general, the Conceptual Memory provides an important top-down influence on the processes in the workspace: active concepts are rather persistent, in contrast to the ephemeral nature of the individual workers.

Model Comparison

To put the dynamic Seek Well model into context with the other models mentioned above, we present a brief discussion of key similarities and differences between the models. First, however, we point out some of the unique aspects of Seek Well compared with the others.

UNIQUE ASPECTS

1. Seek Well operates in a restricted microdomain of isochronous monophonic tonal melodies. The other models dealt with more complex music.
2. This model is stochastic. Expectations are generated in an idosyncratic way dependant on the conceptual memory of the modeled listener. The model depends upon the content of the conceptual memory, which might change after a program run. The rest of the models would give a certain, fixed output each time the same musical cue was presented, but here multiple hearings may produce different dynamic expectations in the modeled listener.
3. A natural mechanism is provided to infer the key and meter of the cue. This is done using the same mechanisms that are used to build other expectations. In other models, the key and meter are either presented to the computer or simple rules of thumb determine the key.
4. Musical memory is incorporated into the model.

NARMOUR AND SEEK WELL

Similarities

1. Both models are primarily focused on bottom-up processing, the subconscious perception of musical parameters.
2. Narmour's concept of registral direction states that small intervals imply continuations in the same direction whereas large intervals imply a change in direction. The first part is like inertia in Seek Well, and reversal after large intervals will be implemented separately by workers and a reversal concept.
3. Both models enjoy the flexibility of sliding parametric scales. For example, although Narmour discusses "large" and "small" intervals, membership in these categories is by degree. The stochastic nature of the Seek Well model provides this, in a sense, because the same music can be perceived differently on different runs of the model. Categories like large and small depend on the state of the system.
4. Both models allow multiple expectations at different hierarchic levels to conflict in particular expectations at particular metric points.

Differences

1. Intra- and extra-opus style are not considered explicitly in Seek Well. Learned style structures do have a place, however, as long-term memories. The difference between intra- and extra-opus does not appear in Seek Well because both occur at a time scale

that required long-term memory to be involved, and there is not a concept in the model of some intermediate-term memory that only refers to music within the current piece.

2. Even though Gestalt principles are used, the direct implications A+A->A or A+B->C at the note-to-note level do not form the foundation of the Seek Well model as they do in the I-R model. Two notes are not seen to have direct fixed implications for a successive note; instead, certain patterns may be more or less expected based on the active concepts and the types of workers involved in the analysis. Although the notes C-D may indeed imply E due to the force of inertia, a successive C might also be expected due to gravity.
3. The emphasis in Seek Well is in modeling subcognitive processes, whereas in the I-R model there is a large focus on categorizing the types of implication and realization present in a melody.
4. Tonality plays a secondary role in the I-R model, while it forms an important part of the Seek Well model, especially as used in the magnetism force.

LERDAHL AND SEEK WELL

Similarities

1. Both use idea of the magnetism of stable pitches. Lerdahl, however, explicitly rejects the separate force of gravity, stating that it is unnecessary and does not provide extra predictive power.
2. Both use pitch alphabets to determine pitch stability.

Differences

1. Lerdahl's intent was to study tonal pitch space, while the Seek Well model is clearly intended for melodic expectation.
2. The Seek Well model involves many more musical features than Lerdhal's expectation model. Lerdahl's model contains a wealth of knowledge about harmony and tonality that is not available to the Seek Well model.

MARGULIS AND SEEK WELL

Similarities

1. Just as in the case of the I-R model, both models allow for different, conflicting expectations at multiple hierarchical levels. Margulis provides a particular weighting system to take the multiple levels into account, but she notes that these weights are a

ripe area for further research. Seek Well indirectly uses a similar approach, where the ”weight” of each level is influenced by the concepts active in memory and the mix of workers involved in expectation generation.

2. Margulis incorporates memory and tempo in her model by basing the weights for the weighted sum for hierarchical predictions on the time-span length of each hierarchy. This model goes into more detail by explicitly modeling echoic memory, the phonological loop in working memory, and long-term memory.
3. Margulis generates a dynamic field of expectations at each point in the music. In Seek Well, a dynamic field of musical forces is hypothesized at each point.

Differences

1. Margulis describes different types of tension and calculates the amount of each type of each tension as each point. Tension is not a central feature of the Seek Well model, but Margulis’ tension values may be calculated and then used as a concept that can be activated if necessary to influence worker behavior or provide top-down guidance.
2. Margulis’ model has a small set of parameters that were tweaked by hand until model results matched her intuitions. The componentized nature of the model likely facilitated this assigning of parameter values. In contrast, Seek Well’s cognitive architecture makes parameter values more difficult to cope with because some “parameters” are not simple weights as in Margulis’ model. Rather, they are virtual parameters defined by the manner in which components of the architecture interact. Similarly, each individual component such as a worker will have its own parameters that affect its behavior. The many degrees of freedom in this model make it complex.
3. The architecture of Seek Well is more amenable to acknowledging the complexity of the listening experience. Margulis’ model, for instance, allows for the influence of multiple layers of hierarchy in expectation generation, but combining these multiple expectations in one weighted sum masks the cognitive complexity of the influence of these layers. When these weights are selected by hand or even fit statistically to experimental results, the particular weight values are prone to inherent fragility. Care must be taken, for instance, to make sure that introducing additional features to a model does not invalidate the specific weights already selected.
4. Whereas in Margulis’ model expectation values are directly calculated based on previous pitches present at several layers in the hierarchy, Seek Well allows expectations to emerge from the lower-level modeling of subcognitive agents. This may reduce the importance of individual weight values. While at the most basic level there are parameters involved that may be “tweaked” to produce optimum results, the hope is

that the global architecture is more robust to changes in parameter values. Seek Well does not eliminate parameters but does push the parameters down to a lower subcognitive level.

Summary and Future Work

The models of melodic expectation developed by Narmour, Lerdahl, Larson, and Margulis each provide useful ideas that should be carefully considered in any other work on melodic expectation. The dynamic Seek Well model has a significantly different architecture than those models, but as the modeling goal is similar for each of these projects there are philosophical principles as well as details of implementation that should be useful in implementation of the present model.

The new model will have its own strengths as well as weaknesses. Assuming a successful implementation of the model, key strengths will hopefully include its stochastic nature which facilitates modeling multiple hearings of the same music, the ability to infer key, meter, and hierarchical structure, incorporation of bottom-up Gestalt principles, and the use of subcognitive agents and memory that is rooted in intuition about musical cognition. Weaknesses include the restricted microdomain of the model, complexity due to multiple model components such as workers, and the difficulty of understanding interactions between model components in a stochastic environment.

The next step in this research is to actually formalize and implement the model as a computer simulation. A working version of this model would ideally function not only to make melodic expectation predictions but also to serve as a test bed for hypotheses in music cognition. For example, it would be interesting to implement Narmour's fundamental principles, Margulis' expectation formulas, and Larson's musical forces as separate groups of workers in this model. Then each group could be activated for separate runs to compare the utility of the principles embodies by each. Similarly, individual components of one model could be disabled; for example, Larson's gravity force could be disabled here to investigate its effect on predictions.

Finally, more research is necessary into representation of musical memory. The temporal aspect of music requires special care to respect the listening experience. Snyder (2000) summarizes relevant work in the field. Of particular interest is the possibility of modeling echoic memory and working memory in addition to long-term memory.

References

Hofstadter, D. R. & Fluid Analogies Research Group. (1995). *Fluid concepts & creative analogies: computer models of the fundamental mechanisms of thought*. New York, Basic Books.

Hofstadter, D. R. & Mitchell, M. (1994). The Copycat Project: A Model of Mental Fluidity and Analogy-Making. In Holyoak, K. and Barnden, J. (eds.) 1994. *Advances in Connectionist and Neural Computation Theory Volume 2: Analogical Connections* pp 31-112. Norwood NJ: Ablex Publishing Corporation.

Larson, S. (1993a). Computer Models of Melodic Implication and Key Determination in Tonal Music. Presented at the meeting of the Society for Music Perception and Cognition, Philadelphia, Available (from Center for Research on Concepts and Cognition; 510 North Fess; Bloomington, IN 47408) as CRCC Technical Report #77.

Larson, S. (1993b). Seek Well: A Creative Microdomain for Studying Expressive Meaning in Music, Available (from Center for Research on Concepts and Cognition; 510 North Fess; Bloomington, IN 47408).

Larson, S. (1997). Continuations as Completions: Studying Melodic Expectation in the Creative Microdomain Seek Well, Available (from Center for Research on Concepts and Cognition; 510 North Fess; Bloomington, IN 47408) as CRCC Technical Report #117.

Larson, S. (2004). Musical Forces and Melodic Expectations: Comparing Computer Models and Experimental Results. *Music Perception* **21***(4): 457-498.*

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

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

Margulis, E. (2005). A Model of Melodic Expectation. *Music Perception* **22***(4)*: 663-714.

Minsky, M. L. (1988). *The society of mind*. New York, Simon & Schuster.

Narmour, E. (1990). *The Analysis and Cognition of Basic Melodic Structures*. Chicago: University of Chicago Press.

Narmour, E. (1992). *The Analysis and Cognition of Melodic Complexity*. Chicago: University of Chicago Press.

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