Psychological Review
2003, Vol. 110, No. 4, 683–712

Copyright 2003 by the American Psychological Association, Inc.
0033-295X/03/$12.00 DOI: 10.1037/0033-295X.110.4.683

Incremental Planning in Sequence Production

Caroline Palmer and Peter Q. Pfordresher
Ohio State University

People produce long sequences such as speech and music with incremental planning: mental preparation of a subset of sequence events. The authors model in music performance the sequence events that can be retrieved and prepared during production. Events are encoded in terms of their serial order and timing relative to other events in a planning increment, a contextually determined distribution of event activations. Planning is facilitated by events' metrical similarity and serial/temporal proximity and by developmental changes in short-term memory. The model's predictions of larger planning increments as production rate decreases and as producers' age–experience increases are confirmed in serial-ordering errors produced by adults and children. Incremental planning is considered as a general retrieval constraint in serially ordered behaviors.

When people produce long, complex sequences such as speech and music, they must plan what event to produce next (the serial-order problem) and when to produce it (the timing problem). Bernstein (1967) and Lashley (1951) both pointed to music as a quintessential example of serial-ordering abilities because of its complexity, length, and temporal properties. Although musical sequences contain complex serial and temporal structure, people remember them with ease. For example, both expert and novice (child) musicians can perform sequences longer than 100 successive events *from memory* at fast production rates or tempi (4–15 events/s), with low error rates (less than 5%; Finney & Palmer, 2003; Palmer & Drake, 1997). Listeners and performers can remember musical sequences over long retention intervals (Chaffin & Imreh, 1997), even novel songs presented once (Dowling & Bartlett, 1981). Thus, musical sequences are highly memorable, despite their serial and temporal complexity.

To account for such good memory for order in sequence production, many theories assume that planning occurs: Sequence events are retrieved and organized prior to their production. Lashley (1951) proposed that sequences are planned hierarchically, in part because of the rapidity with which people produce sequences. In hierarchical plans, higher order units (such as a musical phrase) increase access to lower level units (such as the notes within a phrase). Hierarchical organization is popular in theories of memory for serial order (cf. Estes, 1972, 1997; Martin, 1972; Restle, 1972) as well as in accounts of retrieval in sequence production. However, it is unlikely that producers have simultaneous access to all elements in a long sequence, as a completely hierarchical model might suggest. Instead, theories of sequence production in domains such as speech and typing suggest that planning occurs incrementally, with producers' access constrained to some sequence elements (Bock, 1982, 1995; Dell, Burger, & Svec, 1997; Rumelhart & Norman, 1982; Sternberg, Knoll, Monsell, & Wright, 1988). Incremental approaches to sentence production propose that planning and producing co-occur in time, such that completed increments are released during production (Kempen & Hoenkamp, 1987; Levelt, 1989). This reduces demands on memory and spreads the production costs across time (Wheeldon, Meyer, & Smith, 2003). Thus, producers' access to sequence events is not evenly distributed over time, a characteristic of incrementality that we adopt here. Incremental planning is preferable to a strictly serial or strictly hierarchical view of planning because it allows more efficient use of short-term memory and other time-limited mental processes. How temporal constraints of mental processes influence the incrementality of planning is unknown.

We propose a theoretical framework in which planning arises from time-dependent incremental retrieval of sequence events in music performance. Incremental planning refers here to the sequence elements that are mentally accessible at a given time during performance. We make two claims: first, that the extent of incremental planning is constrained by production rate and by age-related memory differences and second, that each sequence element is encoded and accessed in terms of its similarity and proximity to all other elements within an increment. We refer to the span or distance between accessible sequence elements as the *range* of planning. Our approach to incremental planning is consistent with views of short-term memory as based on cues at retrieval, as opposed to rehearsal-based activation of information from long-term memory (cf. Nairne, 2002). Current elements are instantiated by their contextual ties to other sequence elements, and the strength of those ties changes across the sequence as the current context changes. Furthermore, our approach addresses memory for serial order in terms of constraints from timing. The

Caroline Palmer and Peter Q. Pfordresher, Department of Psychology, Ohio State University.

This research was supported by National Institute of Mental Health Grant R01-45764 to Caroline Palmer. We gratefully acknowledge comments on a draft of the article by Gordon D. A. Brown, Gary S. Dell, Mari Riess Jones, Edward W. Large, In Jae Myung, and Shari R. Speer; invaluable conversations with Gary S. Dell and In Jae Myung; and the assistance of Danielle Brink, who conducted one of the experiments as part of a senior thesis.

Correspondence concerning this article should be addressed to Caroline Palmer, who is now at the Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal H3A 1B1, Quebec, Canada, or to Peter Q. Pfordresher, who is now at the Department of Psychology and Institute for Music Research, University of Texas at San Antonio, 6900 North Loop, 1604 West, San Antonio, Texas 78249. E-mail: caroline.palmer@mcgill.ca or ppfordresher@utsa.edu

relative timing of sequence items during presentation influences their serial recall (Martin, 1972; Restle, 1972; Restle & Brown, 1970). In fact, the relative timing of musical events must be remembered to distinguish among different sequences with the same serial ordering. Consider the first five tones in "Mary Had a Little Lamb" and "The First Noel"; the serial order of the tones is identical, but the time between tone onsets (forming musical durations) differs across the sequences. Thus, memory for relative timing is necessary to distinguish among musical sequences.

Music poses a challenge for serial-order models, and there are virtually no models of retrieval in music performance. Musical sequences are usually composed of a relatively small set of pitches that reoccur at different serial positions; the serial ordering of a pitch event in one sequence context does not provide useful information about its serial ordering in another sequence context. In language, serial-ordering constraints on words and other linguistic items are retained (at least partially) from one sentence to another; in music, the serial ordering of pitch events is completely determined by their relation to other events within each sequence context. Nevertheless, performers easily learn and produce novel sequences that contain the same pitch events in different serial orders. We assume therefore that the representation of music is a function of contextual associations among pitch events, consistent with associative models of serial order (Lewandowsky & Murdock, 1989) and of perceptual facilitation in music (Bharucha, 1987; Tillman, Bharucha, & Bigand, 2000). We also assume that hierarchical relationships among musical events in sequences are encoded in memory for musical patterns (e.g., Collard & Povel, 1982; Deutsch & Feroe, 1981; Jones, 1974; Simon & Sumner, 1968). Our model is unique, however, in representing sequence elements in planning increments that are contextually driven distributions of event activations. A current element's activation encodes not only *its* sequence position but also that of each other element in a planning increment. Activation levels do not represent response selection so much as response preparation; they are a measure of the extent to which elements are participating in the current state.

In this article, we present a theoretical framework of retrieval in music performance that addresses the range of incremental planning. We review relevant findings from serial-recall tasks on memory for serial order and accounts of serial recall in the context of production tasks. Then, we present a formal model of incremental planning that makes two general predictions. First, similarity and proximity between the currently produced event and other events within a planning increment facilitate their simultaneous accessibility. Second, the range of planning is constrained by production rate (with faster rates narrowing the scope) and by changes in memory span with development (with younger performers exhibiting a narrowed scope). The explanatory power of the range model is contrasted with that of other models in accounting for serial-ordering errors in adults' and children's music performances. In the last section, we extend the model to other performance tasks and sequences.

Memory for Serial Order

Evidence from serial-recall tasks indicates that people often recall sequence items in the wrong serial order. Errors that reflect correct item information but incorrect serial order information, called *movement errors,* have been documented for many tasks and types of stimulus materials, including speech and music (G. D. A. Brown, Preece, & Hulme, 2000; Dell et al., 1997; Healy, 1974; Lee & Estes, 1981; Palmer & van de Sande, 1995), and are often taken as evidence of some dissociation of memory for item and memory for serial order (Bjork & Healy, 1974; G. D. A. Brown et al., 2000; Healy, 1974). Two general constraints are reflected in the study of movement errors: proximity and similarity effects. Proximity constraints are seen in the likelihood that participants recall a particular stimulus item at or near its correct serial position. The probability of an item's recall in a particular serial position tends to decrease with distance from the correct position (e.g., Healy, 1974; MacKay, 1970; Nairne, 1991). Movement gradients, or error proportions displayed by absolute distance between an item's position of recall and its correct position, indicate that memory for an item is active over a number of sequence positions proximal to the correct position. Item similarity also exhibits constraints on serial recall. Recall of lists of similar verbal items (in meaning or sound) tends to be less accurate overall than recall of dissimilar items; furthermore, recall is hurt by item similarity differentially across serial positions (G. D. A. Brown et al., 2000; Conrad, 1965; Healy, 1974). The slope of movement gradients for serial recall of similar items tends to be shallower than that for dissimilar items, reflecting an overall tendency for movement errors to arise from larger sequence distances for similar items than for dissimilar items.

Explanations for proximity and similarity constraints on serial recall focus on decay and interference. For example, Estes's (1997) perturbation model incorporates proximity-based decay to account for forgetting; G. D. A. Brown et al.'s (2000) temporal distinctiveness model incorporates similarity-based interference. These models and others combine similarity and proximity effects to explain various memory findings, including serial-position curves and retroactive and proactive interference in serial-recall tasks. However, one assumption is problematic for their application to sequence production in domains such as speech and music: The temporal properties of the items are not important during recall. We propose that the absolute time between items during their production affects proximity constraints on retrieval.

Proximity and similarity constraints on memory retrieval are also evidenced in production tasks such as music performance. A movement gradient for pitch errors in piano performances of a *Prelude* by Bach (sequence length of 400 events) is shown in Figure 1 (see also R. K. Meyer & Palmer, 2003). The chance

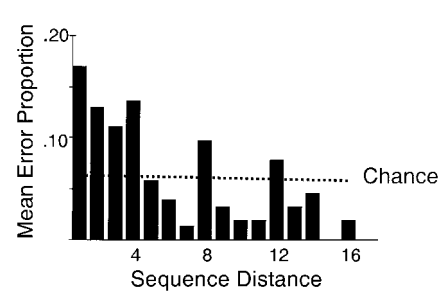

Figure 1. Movement gradient for pitch errors in piano performances of a *Prelude* by Bach (sequence length of 400 events). The dotted line indicates chance estimate.

estimate, indicated by the dashed line, reflects the proportion of all possibilities for elements to arise from each sequence distance, relative to the total length of the sequence. The movement gradient shows proximity constraints in the larger error proportions at nearby distances relative to faraway distances and similarity constraints in the peaks at certain periodic distances (4, 8, 12, and 16 events) from the correct serial position (see Palmer & van de Sande, 1995, for similar error distributions with different musical pieces). The peaks are aligned with musical events whose metrical accent is more similar than other events to the current event. The model described here uses the similarity and proximity constraints to predict differential accessibility of sequence events within a planning increment during retrieval.

This musical example raises an important distinction between serial recall of music and other sequences: the typical length of recalled sequences. The relatively short sequences (less than 15 events) of most serial-recall tasks can limit the ability to differentiate proximity constraints on recall from chance occurrences, which are sensitive to the length of the sequence. For example, errors that arise in a sequence ABCDE from a distance of four events (A replaces E's serial position or vice versa) would have a total of two opportunities by chance or 10% of all independent possibilities across distances, whereas errors from a distance of one event (A replaces B's serial position or vice versa) would have eight opportunities or 40% of all possibilities.[1] In contrast, longer sequences typical of music offer more opportunities for errors to arise from a given sequence distance and yield less variable chance estimates across different distances, as indicated by the relatively flat dotted line in Figure 1 for a 400-event sequence. Thus, an advantage of studying serial recall in the production of long musical sequences is that chance estimates for movement errors can be distinguished from typical proximity constraints.

Developmental Influences on Serial Recall

Studies of serial recall document increases in children's memory span with development (e.g., Gathercole, 1999). Memory span is measured by the length of the longest fragment in a sequence that a participant can recall in correct order. Capacity limitations on memory span have been proposed both in terms of a constant number of chunks (7 ± 2; Miller, 1956) and in terms of a constant time period of rehearsed items (2 s; Baddeley, 1986). Increases in memory span across 6–14 years of age are correlated with increases in speech (articulation) rate (Gathercole & Adams, 1994; Hulme, Thomson, Muir, & Lawrence, 1984). One explanation is that younger children's slower subvocal mental rehearsal leads to fewer items being retained and more loss of information over time (Baddeley, 1986; Towse & Hitch, 1995), which is supported by evidence of impairments in memory span that accompany certain articulatory disorders (see Gupta & MacWhinney, 1997). Accounts of developmental changes in memory span that emphasize decay of fewer items over time point to evidence that older children show increased temporal persistence of auditory sensory memory relative to younger children (Gomes et al., 1999) as well as increased short-term storage of phonological (verbal) information (Gathercole, 1999). Other explanations of age-related increases in memory span include attentional resources (Barrouillet & Camos, 2001; Cowan, 2000), reconstruction of partial traces (G. D. A. Brown & Hulme, 1995; Neath & Nairne, 1995), and increased temporal distinctiveness of items (G. D. A. Brown, Vousden, McCormack, & Hulme, 1999).

Developmental studies of memory span have focused thus far on serial-recall tasks; here, we address developmental differences in serial recall in the context of production. Incremental planning in music performance, like memory-span measures, may also change with development. Movement errors in music performance span larger sequence distances as performers increase in age and experience. Older children show a larger planning range than younger children with the same amount of musical experience (Drake & Palmer, 2000), and musically experienced children show a larger planning range than less experienced children of the same age (Palmer & Drake, 1997); adults (with the most musical experience) show the largest range. We propose that the range of incremental planning increases with age and experience, and we model differences in children's and adults' range of planning measures with changes in parameters that influence proximity- and time-based decay of event activations. In the next section, we discuss planning in the context of sequence production, which introduces time constraints on serial ordering.

Serial-Ordering Constraints in Sequence Production

Production tasks like speech and music make considerable demands on serial-ordering processes; they must be accurate and quick. In response to such demands, incremental accounts of speech production posit that mental processes work in parallel on different increments (Bock, 1995; Kempen & Hoenkamp, 1987; Levelt, 1989). Incremental planning is evidenced in grammatical encoding (Ferreira, 1996; M. Smith & Wheeldon, 1999) and phonological encoding (A. S. Meyer & Schriefers, 1991). To implement such planning, one must determine increments that are relevant to the content and form of the sequence. Errors are a common source of evidence for identifying planning units; errors tend to reflect interactions between an intended item (the target) and an unintended item (the intruder) that are similar on some structural dimension. The majority of production errors in speech and music reflect an interaction among elements of the same size; one phoneme replaces another phoneme (rather than a word), and a chord interacts with another chord (rather than a single note). Also, the contents of the interacting elements tend to share many features (semantic or phonological features for words; harmonic or diatonic features for tones). Finally, the positional contexts of the interacting elements in movement errors (the majority of errors) tend to be similar (Dell, 1984, 1986; Garrett, 1980; MacKay, 1970; Palmer & van de Sande, 1993, 1995). In summary, music and speech demonstrate constraints on the error types that arise, and these constraints often reflect similarity biases of retrieval mechanisms.

[1] Assuming independent error probabilities, we computed chance estimates as

$$(n - y)/\sum_{x=1}^{s}(n - x)$$

for a movement error arising from distance y in a span of distances s for a sequence of length n.

Similarity-based confusions arise among elements in music performance that share musical structure, including phrasal, diatonic, harmonic, and metrical relations between targets and intruders (Palmer & Drake, 1997; Palmer & van de Sande, 1993, 1995). These similarity influences often arise from retrieval or planning processes, as opposed to motoric or execution processes; structural similarity in performance errors is evidenced across a variety of motor movements that produce them (Palmer & Meyer, 2000; Palmer & van de Sande, 1995), and similarity metrics that influence motor-related errors (such as similar finger or hand movements) are different from those that influence retrieval-related errors (such as similar melodic relations; Palmer & Meyer, 2000; Palmer & van de Sande, 1993). Diatonically related errors in particular (those in which error intruders reflect a pitch appropriate to the key of the sequence) implicate nonmotoric processes. Errors arising from motor sources often result in the misaiming of actions that generate errors that are spatially proximal to the intended location (cf. Fitts & Peterson, 1964). Spatially proximal errors on a keyboard tend to be unrelated diatonically, and thus diatonic similarity is not consistent with motoric (movement-based) similarity. We show in experiments reported here that the majority of errors in music performance arises from structural similarity relations that can be accounted for by retrieval rather than execution processes.

Most informative for accounts of serial ordering in sequence production are movement errors (sequence items recalled in an incorrect order). Movement errors can be defined in terms of their directional relationship to their intended serial position (anticipatory or perseveratory) as well as their distance (how many sequence events away). Anticipatory errors reflect the influence of future events on the present; perseveratory errors reflect the influence of past events on the present. Both speech and music performance show an increasing anticipatory tendency as overall accuracy increases or practice increases (Dell et al., 1997; Drake & Palmer, 2000). One reason for the future-oriented bias indicated by anticipatory errors is future-oriented dependencies that exist among sequence events, in which the present event is determined by its relationship to some future event. Future-oriented dependencies often arise in speech ("an apple" instead of "a apple") and require a coactivation of present and future events. Although musical sequences have future-oriented dependencies (such as harmonic sequential dependencies), they tend to reflect probabilistic constraints rather than well-formedness constraints (e.g., Lerdahl & Jackendoff, 1983; Piston, 1987). Thus, language has stricter serial-ordering constraints than music. A second reason for a future-oriented bias in production is the need to turn off the past to prohibit its interference with the future. This constraint has been implemented in Dell et al.'s (1997) model that predicts a general anticipatory tendency as well as in other models of speech production (MacKay, 1987; Vousden, Brown, & Harley, 2000).

Another serial-ordering constraint in sequence production is the range over which events are simultaneously accessible. The range or sequence distance between interacting elements in movement errors can be constrained by hierarchical relationships among concurrently planned elements. For example, movement errors in speech and music performance tend to arise among elements in the same phrase rather than from different phrases, suggesting that sequences are planned in phrasal increments (Garcia-Albea, del Viso, & Igoa, 1989; Palmer & van de Sande, 1995; see also Levelt, 1983). Despite these findings, few models of sequence production have focused on range constraints (but see Vousden et al., 2000). Musicians' emphasis of different hierarchical levels influences the scope of planning; movement errors tend to span larger distances when performers emphasize larger hierarchical levels (larger scale relationships) than smaller levels in the same sequence (R. K. Meyer & Palmer, 2003). Musical elements vary on multiple hierarchical dimensions, including tonality, phrase structure, and meter, that influence movement errors in music performance (Drake & Palmer, 2000; Palmer & van de Sande, 1993, 1995). Thus, range of planning may be constrained by the producers' hierarchical representation of the sequence; the model we describe represents hierarchical weighting in its parameters. Next, we review hierarchical influences on serial ordering in production of language and music.

Serial-Ordering Constraints: Language Production

The serial ordering of words and phonemes in sentence production is often accounted for with frames (Bock, 1995; Dell, 1986; Levelt, 1989). Frames are abstract representations that specify ordering relationships among permissible item categories. The serial order for a linguistic constituent such as a noun phrase is represented in a frame as a series of slots into which items from permissible categories are positioned. Frames are hierarchical and exist at multiple levels, encoding categories such as metrical accent, syntactic categories, and consonant clusters. Frames are proposed primarily to determine the serial ordering of linguistic units; in addition, the size of the frame may regulate the scope of planning (Bock, 1995).

Dell et al.'s (1997) frame-based model of language production makes specific predictions for anticipatory constraints on serial ordering. Structural frames are combined with a "plan," which weights past, present, and future elements in a sentence to be produced. Three parameters in the model influence anticipatory behavior. Practice increases the weighting of the present and future, but not the past; a time-step parameter (inverse of rate) indicates the number of time steps over which activation spreads to each part of the plan; and anticipatory activation weights the future. The weighting of past, present, and future events is the product of practice and the anticipatory activation. The perseveratory tendency (weighting of past events relative to current events) decreases as rate slows or as practice increases, whereas the anticipatory tendency (weighting of future events relative to current events) is defined by anticipatory activation alone.

These relationships lead to two important predictions that Dell et al. (1997) term the *anticipatory practice effect* and the *anticipatory rate effect*. The anticipatory practice effect predicts that, as practice increases, overall error probability decreases but the anticipatory proportion of errors [anticipations/(anticipations + perseverations)] increases because of reduced activation spreading to the past relative to the future. The anticipatory rate effect predicts that, as rate slows, overall error probability decreases but the anticipatory proportion of errors increases because of reduced activation spreading to the past relative to the future. These predictions are combined in a general anticipatory effect: Any influence on overall error rate should affect anticipatory proportion in the reverse way, such that anticipatory proportion is negatively correlated with the overall error rate. Experimentally elicited

speech errors resulting from tongue twisters confirmed the anticipatory practice and rate effects and the general anticipatory effect (Dell et al., 1997). The authors suggested that the anticipatory predictions of the serial-order model may apply to other production domains with complex structural sequences, such as music.

Indeed, production errors in music performance support the anticipatory practice effect (Drake & Palmer, 2000). Pianists' proportions of anticipatory errors relative to perseveratory errors increase with practice and with decreased error rates, the same relationships predicted by Dell et al. (1997). There is less evidence for an anticipatory rate effect, however; most studies of music performance do not control rate. If directional constraints on planning mechanisms such as those specified by Dell at al.'s (1997) model operate similarly across domains, then both anticipatory practice and rate effects should be seen in music performance.

One model of sentence production makes explicit predictions for the range of planning. Instead of relying on frames, Vousden et al. (2000) proposed an oscillator-based model of the sequencing of phonemes and syllables in language production that predicts the scope of movement errors. A sequence of syllables is linked to successive states of a context vector that contains a set of oscillators of different periods that encodes the learning context of a string of syllables and phonemes. This context vector encodes the serial ordering of the syllables and phonemes within the sequence. Planning a phoneme for production involves associating the phoneme features with the product of the oscillator outputs from the context vector; each phoneme is associated with a different state of the set of oscillators. During production, ordering errors can occur when successive states of the oscillators are used to activate the phonemes with which they were associated at encoding. This model predicts differential error patterns across sequence positions that reflect syllable-position constraints and phonemic similarity. The movement gradient for anticipatory errors is predicted to be steeper (predicting a smaller mean range) than the gradient for perseveratory errors because of postoutput suppression of each event after its production (which selectively reduces the likelihood of perseveratory errors from nearby distances). We return to these predictions later.

Serial-Ordering Constraints: Music Performance

Whereas serial ordering is a focus of studies of sentence production, timing has been the primary focus of study of music performance (for reviews, see Clarke, 1999; Gabrielsson, 1999; Palmer, 1997). Musical sequences contain explicit temporal structure such as rhythm and meter that specify the durations of sequence events. Rhythm and meter are hierarchical structures that reflect temporal relationships among past, present, and future events. Thus, an assumption that nonadjacent sequence events are coactivated during retrieval is consistent with the temporal structure of music. Specifically, we posit that musical meter provides a hierarchical framework that defines the serial ordering of events.

Meter is defined broadly as a regularity in the pattern of sequence events. Some metrical theories define that regularity in terms of accent strength; meter can be described in terms of strong and weak accents, usually in binary or ternary alternation at several hierarchical levels (Cooper & Meyer, 1960). An accent-based metric can assume an ordinal scale among events; the downbeat of a measure is stronger than the second beat, and so on, irrespective of the time elapsed between beats. Other approaches define that regularity in terms of time, in which metrical beats are separated by equal time units (e.g., Lerdahl & Jackendoff, 1983; Longuet-Higgins & Lee, 1984). A time-based metric can assume an interval or ratio scale among conceptualized beats, which means that one time span has twice the duration of another, and so on. In this article, we assume an ordinal-scale approach to metrical accent, which is later combined with ratio-scaled timing in the serial component of the framework. We assume metrical accent is ordinal because it provides a more parsimonious account of serial ordering. Additionally, music performance rarely exhibits equal event durations (Gabrielsson, 1987; Repp, 1992; Seashore, 1938; Sloboda, 1983; Sternberg, Knoll, & Zukovsky, 1982), despite the fact that listeners tend to perceive (ordinal) duration categories (Clarke, 1987; Desain & Honing, 1989, 2003).

Metrical accent is modeled with metrical grids in both music theory (Lerdahl & Jackendoff, 1983) and metrical phonology (Liberman & Prince, 1977). Metrical grids define a sequence of events in terms of accent strengths. Figure 2 shows an example of a metrical grid for a musical sequence with a binary meter ($\frac{4}{4}$, which refers to four quarter-note beats in a measure); the Xs below the notes in the sequence indicate an abstract representation of metrical accent strength for one measure plus one beat of the music. This grid applies in a recurrent fashion across measures of the music and is unchanged throughout most pieces in Western tonal music that share the same meter, as reflected in the time signature. The highest hierarchical level in the grid (marked Level 4) corresponds here to the downbeat of the measure (the onset of each metrical bar) and is aligned with the greatest number of accents in the grid. The number of total event locations in one cycle of the grid, *n*, is determined by the number of beat divisions within and between metrical levels. The remaining hierarchical levels define the half-note level (Level 3), the quarter-note level (Level 2), and the eighth-note level (Level 1). Although Figure 2 shows the metrical grid for only the first nine events, the grid repeats in an identical fashion throughout the music. Similar grids have been proposed to represent ternary meters, with two weak accents occurring between strong accents. We assume that the metrical grid operates during retrieval as a frame that defines the ordering of metrical accents.

Metrical frames are often reflected in the statistical characteristics of music. Frequency distributions of the number of notes at different metrical positions in musical compositions indicate more events placed at positions of strong metrical accent in samples of Western music (J. C. Brown, 1992; Huron, 1993; Palmer, 1996; Palmer & Krumhansl, 1990). We apply a similar analysis to a larger corpus of different musical styles to test the generality of metrical frames. Figure 3 shows the frequency distributions of pitch events in different musical styles and historical periods: 15th-century Chinese tone music, 15th-century European vocal music, 20th-century American swing jazz, and children's nursery tunes (circa 1600 to present). The frequency distributions depict the total number of note events that occur at each metrical position in notated music in a binary ($\frac{4}{4}$) meter (see Appendix A for complete details). Despite the different countries of origin, time periods, tonal and rhythmic characteristics, and mode of expression (ensemble or solo instrument), the distributions of note events resemble strongly the theoretical distributions of accent strength

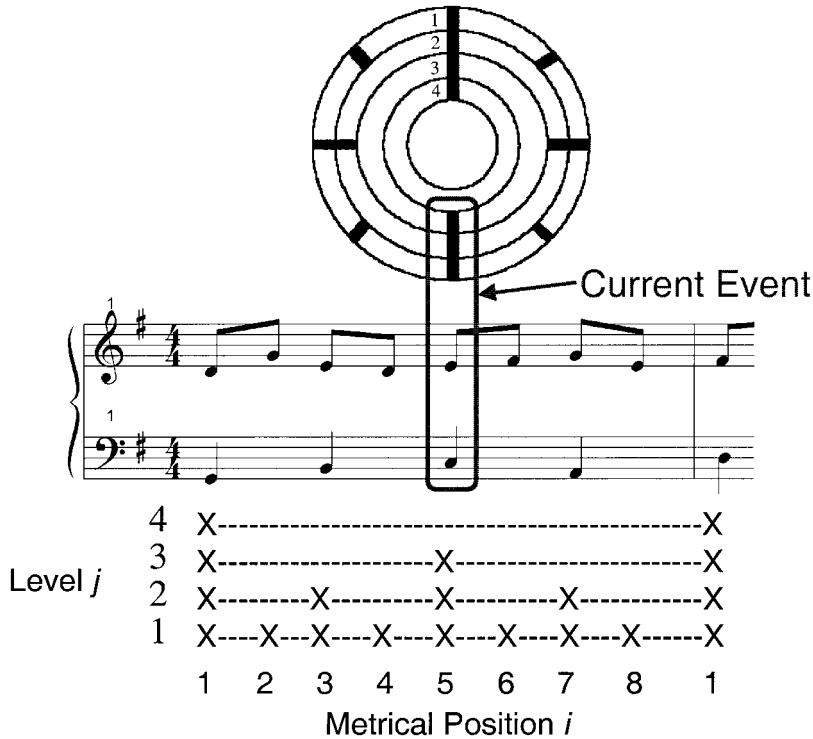

Figure 2. Depiction of a binary metrical grid (bottom) and one of the musical sequences in Experiment 2. A cyclic depiction of metrical accent, equivalent to the information in the metrical grid, is shown at the top. The Xs below the notes in the sequence indicate an abstract representation of metrical accent strength for one measure plus one beat of the music.

proposed by metrical frames. The children's tunes especially suggest that the periodic accent distribution is easily learned and is not limited to expert performers. Although the frequency distributions are susceptible to a criticism that notated music forces categorical distinctions that may or may not exist in sounded music, the evidence indicates a strong tendency for performers and transcribers (often listeners) to organize music into periodic recurrences of events that reflect metrical structure (Clarke, 1987; Desain & Honing, 2003). We test here the assumption that these same periodic patterns influence the retrieval of serial-order information during performance.

The regularity of meter offers advantages for memory of musical sequences. Metrically regular patterns are more accurately perceived and remembered (e.g., Palmer & Krumhansl, 1990; Povel, 1981; Povel & Essens, 1985), and performers can produce event sequences that match a metrical framework more accurately than sequences that do not match a metrical framework (Povel, 1981). Evidence that perception of meter may arise from a wide range of acoustic cues has inspired the view that meter arises from attention to surface-level periodicities in a sequence that generates expectancies by driving internal rhythmic oscillations (Jones & Pfordresher, 1997; Large & Jones, 1999; Large & Kolen, 1994). Other work suggests that meter is a well-learned abstract schema that guides listeners' interpretation of strong and weak beats; in the absence of surface cues, listeners tend to infer patterns of strong and weak accents (Palmer & Krumhansl, 1990; Povel & Okkerman, 1981), and musicians tend to have more well-defined metrical hierarchies than nonmusicians for the same sequences (Palmer & Krumhansl, 1990). Finally, metrical frames offer specific predictions for retrieval of serial order, based on metrical similarity between sequence elements. We elaborate on this feature in the next section.

Modeling Range of Planning

We describe here a two-component model of how musical events are retrieved from memory and ordered in music performance. In this model, meter serves as an accent-based ordinal grid that, together with serial and temporal constraints of short-term memory, predicts the activation of sequence events across the time course of a performance. Each event's metrical accent strength is modeled with metrical grids that define the number of accents that align with each sequence position; we assume that events with similar metrical accent strengths are conceived as more similar than events with different metrical accent strengths. Thus, adjacent events are often less similar than nonadjacent events, and events from farther away are often more similar to each other metrically than events close by. This assumption is consistent with findings that similarity-based interference, a common cause of serial-ordering errors in speech as well as music, often arises among nonadjacent events (Palmer & van de Sande, 1995). The top of Figure 2 shows a cyclic depiction of metrical accent, equivalent to the information in the metrical grid. Each concentric circle represents a level of metrical accent in the grid, with the outermost

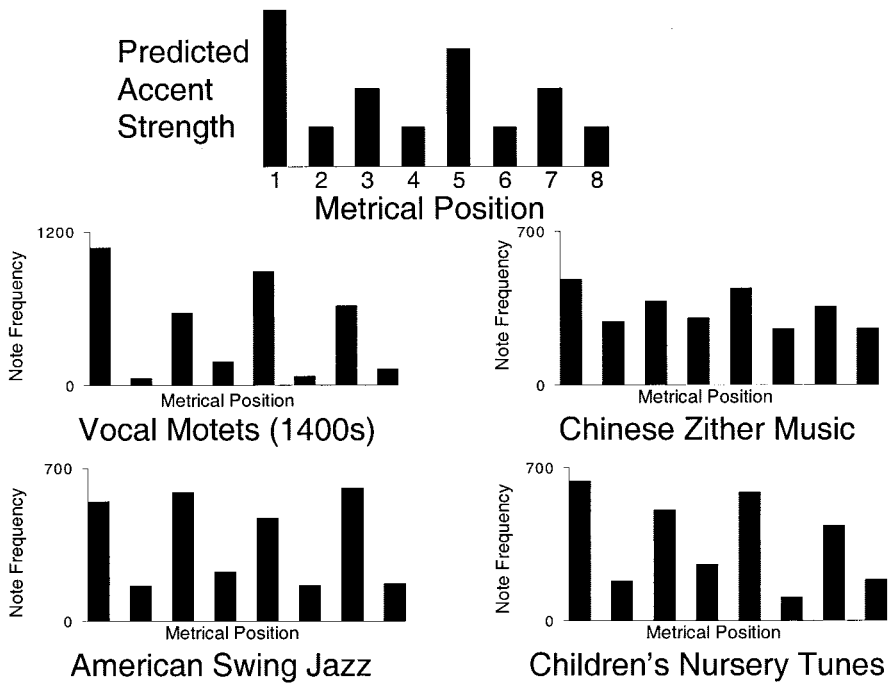

Figure 3. Frequency distributions of musical events in corpora of four musical styles. The frequency distributions depict the total number of note events that occur at each metrical position in music in a binary meter (see Appendix A for complete details).

circle (labeled 1) as the smallest metrical level and the innermost circle (labeled 4) as the largest metrical level. The period of one metrical cycle is defined by the intervals between accents at the highest metrical level in the grid. In this article, we assume a maximum of four metrical levels, on the basis of some perceptual evidence that listeners are influenced by up to four distinct levels (Palmer & Krumhansl, 1990).

The first component of the model, metrical similarity (M), defines the similarity in metrical accent strength between any two sequence events. The accent strength of an event m_i at metrical position i ($i = 1, \ldots, n$, the total number of positions in the metrical grid) is defined in terms of two variables. The first is the number of metrical accents present at position i for each metrical level j in a grid, g_{ji}. For each level j ($j = 1, \ldots, k$, the maximum number of metrical levels) within position i, g_{ji} takes the value of 1 (present) or 0 (absent). So, values of g_{13}, g_{23}, g_{33}, and g_{43} in Figure 2 = 1, 1, 0, and 0, respectively. The second variable is a weight, w_j, associated with each metrical level j, which ranges from $0 < w_j < 1$. This variable reflects differences in the salience or importance that performers and listeners often give to different hierarchical levels; an example is the tactus, or weighted metrical level that performers emphasize or at which listeners clap (cf. Drake, Jones, & Baruch, 2000; Duke, 1989; R. K. Meyer & Palmer, 2003; Parncutt, 1994). All w_js are first assumed in this article to be equal and constant: $w_j = 1/k$; then, w_js are allowed to vary, to test the model with as few parameters as possible first. The metrical accent strength of each event, m_i, is then defined in Equation 1 as the sum of the accents present at each position and level in the grid multiplied by the weight at each level:

$$m_i = \sum_{j=1}^{k} w_j \cdot g_{ji}, \quad \text{where} \sum_{j=1}^{k} w_j = 1. \tag{1}$$

For example, if $k = 4$ and all w_js = .25, then Position 2 in the metrical grid of Figure 2, m_2, has a metrical accent strength of $.25 \times 1$, and Position 3, m_3, has a metrical accent strength of $.25 \times 2$ $(.25 \times 1 + .25 \times 1)$. The metrical similarity sim between two events m_i and m_{i+x} at distance x ($x = -n, \ldots, n$; within one metrical cycle on either side of position i) is defined as $M_x(i)$, their absolute difference divided by their sum and subtracted from 1, as shown in Equation 2:

$$M_x(i) = \text{sim}(m_i, m_{i+x}) = 1 - \frac{|m_i - m_{i+x}|}{m_i + m_{i+x}} = 1 - \frac{\Delta m}{2\bar{m}}. \tag{2}$$

Thus, when w_js are equal, the similarity of events m_2 and m_3 in Figure 2 (at distance $x = 1$) is equal to $1 - (.25/.75) = .66$. According to Equation 2, $\text{sim}(m_i, m_{i+x}) = \text{sim}(m_{i+x}, m_i)$ for all values of w_j; this is an important assumption of the model because it specifies that the accessibilities of an event at distance $-x$ from the past and an event at distance x from the future are equal when averaged across current metrical positions i. The right-hand side of Equation 2 indicates that the metrical similarity function is equivalent to a simple form of Weber's law multiplied by a constant. For example, $\text{sim}(m_3, m_5) = .80$ is greater than $\text{sim}(m_2, m_3) = .66$ for events in Figure 2, even though the difference in number of metrical accents (the numerator of Equation 2) is the same in both cases, because the denominator of Equation 2 is greater for $\text{sim}(m_3, m_5)$ than for $\text{sim}(m_2, m_3)$. This is psychologically appeal-

ing because it is consistent with perceptual evidence in support of Weber's law: Perceptual sensitivity to a physical difference between two stimuli is determined relative to their absolute physical level (e.g., Yost, 1994). Similar contrast functions have been used to model the detection of luminance differences in visual perception (Walraven, Enroth-Cugell, Hood, MacLeod, & Schnapf, 1990). We assume likewise that sensitivity to a conceptual difference between two metrical accents is determined relative to their absolute metrical level.

Each current metrical position i has associated with it a metrical similarity with each other metrical position within a cycle, stored in a vector $\mathbf{M}(\boldsymbol{i})$ of up to $2n + 1$ elements (the number of events within one metrical cycle on either side of the current event, within the boundaries of the sequence). $\mathbf{M}(\boldsymbol{i}) = (\mathrm{M}_{-n}(i), \ldots, \mathrm{M}_x(i) \ldots, \mathrm{M}_n(i))$, which changes as a function of current position i as defined in Equation 2. When averaged across current positions i, $\mathrm{M}_x = \mathrm{M}_{-x}$; an event's metrical similarity with past events and with future events is equivalent, a prediction used later. The metrical similarity function reduces to Equation 3 for each distance x, averaged across i:

$$\mathrm{M}_x = \frac{1}{n}\sum_{i=1}^{n} \mathrm{M}_x(i) = 1 - \frac{1}{n}\sum_{i=1}^{n} \frac{|\mathrm{m}_i - \mathrm{m}_{i+x}|}{\mathrm{m}_i + \mathrm{m}_{i+x}}. \quad (3)$$

Thus, metrical similarity changes at various distances from the current event and tends to be larger among nonadjacent sequence elements than among adjacent sequence elements. This assumption is consistent with the caveat that although movement errors often reflect access to sequence events intended for some distance from the current location, it does not follow that a producer has equivalent access to all intervening events (Garrett, 1980). Although M_x encodes only metrical similarity in this article, it may be weighted by other psychological features that distinguish musical events, such as pitch, intensity, and so on (Bharucha, 1987, 1994; Tillman et al., 2000).

The second component of the model, serial proximity (S), reflects both serial and temporal memory constraints on range of planning. Serial proximity of each event at distance x from the current event is stored in a vector $\mathbf{S} = (\mathrm{S}_{-n}, \ldots, \mathrm{S}_x, \ldots, \mathrm{S}_n)$ of up to $2n + 1$ elements (events within one metrical cycle in either direction of the current event, within the boundaries of the sequence). S varies with distance x but is identical across current metrical positions i. The serial component predicts an exponential decrease in memory for sequence events, as a function of both distance and time elapsed. Event strength decreases as event distance $|x|$ from the current event increases and as event duration t (in seconds) decreases (as production rate increases), in the exponential relationship shown in Equation 4:

$$\mathrm{S}_x = a^{(|x|/t)}. \quad (4)$$

An initial activation parameter a, a free parameter in the model, can take the positive values of 0 (no activation) $< a \leq 1$ (total activation) and represents constraints on short-term memory that change with development and task-specific memory demands. We assume a does not vary with short-term practice. Event duration t is a positive value ranging from $0.1 < t \leq 2.0$ s; limits on t stem from temporal rates observed in performance and time-perception tasks. Performers show preferred rates in this range, and music performance tends to lose temporal precision below 100–150 ms/event (MacKenzie & Van Eerd, 1990; R. K. Meyer & Palmer, 2003). The upper bound of 2 s/event, sometimes considered the upper limit of the "psychological present" (Fraisse, 1984), may be the slowest succession of elements that can be incorporated into a single stream or sequence (Drake et al., 2000) and may reflect time constraints on short-term memory capacity (Grondin, 2001). In contrast to estimates of the psychological present that assume a fixed temporal span within which listeners associate events, we show that production rate modulates the span within which producers associate events.

The exponent $|x|/t$ of the serial component affects decay rate in terms of the number of events per unit time (similar to beats per minute in musical terms) and leads to two predictions. First, the larger $|x|$ is (the farther away an event is from the current event), the weaker an event's strength or accessibility; values of S_x are symmetric around i (i.e., $\mathrm{S}_x = \mathrm{S}_{-x}$). Second, the smaller t is (as the rate gets faster), the weaker the event strength. The relationship of serial distance to time in the exponent stems in part from an analogy to auditory streaming, which refers to the perception of different events within an auditory sequence as arising from the same source or stream (Bregman, 1990). The ratio of pitch separation to temporal separation between two tones influences the perception of the tones as belonging to the same pattern (Van Noorden, 1975) and the likelihood of their serial integration (Jones, 1976); at faster rates (smaller values of t), only events with small pitch separations are perceived as arising from the same sequence. We make a related assumption here: The ratio of serial separation (x) to temporal separation (t) in the exponent influences the likelihood that two events are considered part of the same increment. At faster rates, events separated by small distances are more likely to be accessed as part of the same increment.

The serial component's two main predictions, that event strength increases with increased initial activation and with decreased production rate, are shown in Figure 4. The rate predictions are shown in the left-hand side of Figure 4 for two values of t, with a fixed (.960). For any two production rates, the strength of an event at position $i + x$ will be greater at the slower rate than at the faster rate. The initial activation predictions are shown in the right-hand side of Figure 4 for two values of a, with t fixed (.50 s). For any two initial activation values, the strength of event $i + x$ will be greater at the larger initial activation than at the smaller activation. Thus, event strength decays more quickly at smaller initial activation values and at faster production rates. Comparison of the left and right graphs in Figure 4 indicates that small changes

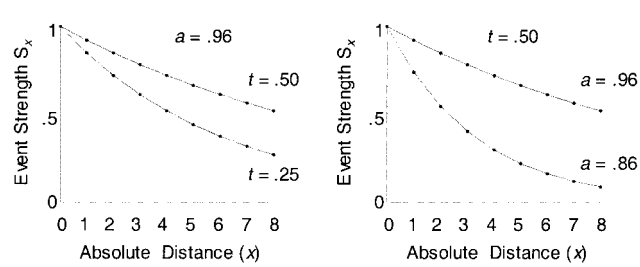

Figure 4. Predictions of the serial component for t (production rate) varying with a (initial activation) fixed at .96 (left) and for a varying with t fixed at .50 (right).

in a yield greater differences in decay than small changes in t. We test the model's predicted effects of t and a on event strength in experiments that manipulate tempo and age–experience of performers, respectively, in this article.

Finally, the metrical and serial components are multiplied to predict the strength of each event $i + x$ during the current event i, on the basis of the product $S_x \times M_x(i)$ of the serial event strength and metrical similarity vectors, as shown in Equation 5. Each sequence event i has a vector $S_x \times M_x(i)$ of up to length $2n + 1$ associated with it: a distributed representation of that event's association with other sequence events across a planning increment. The strength of event $i + x$ is its memory activation at the time that the current event i is to be produced; this function ranges from 0 (no event strength) to 1 (most event strength) and represents the activation of sequence events defined relative to the current event.

$$\text{Event strength (event } i + x \text{ when current event} = i) = S_x \cdot M_x(i). \quad (5)$$

We used the model's event strengths to generate predictions of error proportions. We computed the conditional probability that given an error at event i, an event from distance $|x|$ will be retrieved, based on its normalized event strength. Equation 6 defines the normalized event strengths at each distance $|x|$, averaged over positions i. The numerator of Equation 6 (from Equations 3 and 4) represents the strength of an event at distance x, and the denominator represents the summed strengths of all events within the metrical cycle. The movement gradient (error probabilities across sequence distances) shown in the bottom of Figure 5 reflects probabilities combined over past ($-x$) and future (x) events; event strengths are equal for past and future events when averaged over current positions i ($S_x \times M_x = S_{-x} \times M_{-x}$). The top of Figure 5 reflects the predictions of the unnormalized metrical and serial components.

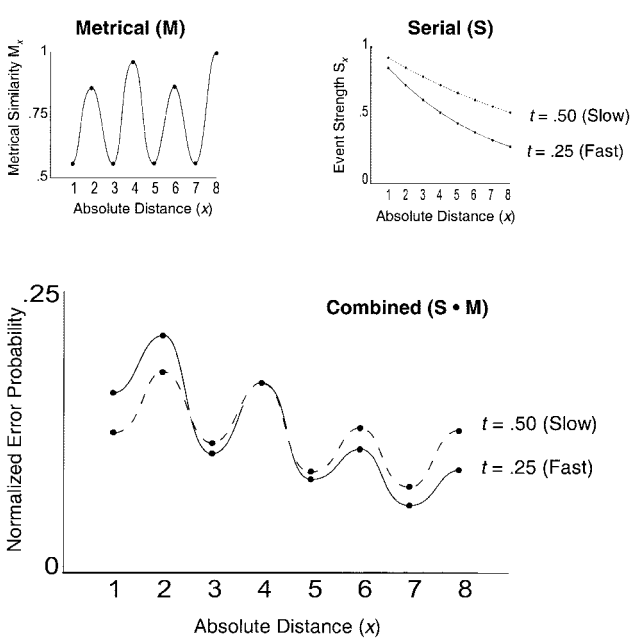

Figure 5. Predictions of the metrical and serial components (unnormalized, top) and their product (normalized, bottom) for a (initial activation) = .96.

$$\text{Probability (event at distance } |x| \text{ occurs} \mid \text{error at current event)} = \frac{S_x \cdot M_x}{\sum_{|x|=1}^{n} S_x \cdot M_x}. \quad (6)$$

The range model makes several predictions for retrieval during music performance, based on the strength of elements at each position in the sequence. First, the model predicts differential effects of metrical similarity and serial proximity across sequence distances. The periodicities in the bottom of Figure 5 reflect the metrical component's contributions to error proportions, and the higher error proportions at shorter distances characterize the serial component's contribution. Second, the model predicts differential effects of production rate on error proportions arising from different sequence distances. For any two production rates and a given initial activation value, sequence events at farther distances are more accessible (and yield more serial-ordering errors) for the slower rate than for the faster rate, and events at closer distances are more accessible for the faster rate than for the slower rate. This interaction between production rate and distance is shown in the bottom of Figure 5. The serial component decreases the activation of sequence elements at greater distances more quickly for the faster performance than for the slower performance, in essence damping the effects of metrical similarity for events from greater distances.

The model also predicts the mean absolute distance for any serial-ordering error; this measure of mean range represents the increment of planning and is computed as the normalized error probabilities from Equation 6 at each distance $|x|$, multiplied by $|x|$ and summed, as shown in Equation 7:

$$\text{mean range} = \sum_{|x|=1}^{n} \left(\frac{S_x \cdot M_x}{\sum_{|x|} S_x \cdot M_x} \cdot |x| \right). \quad (7)$$

For any two production rates or values of t, the mean range for the faster performance is smaller than the mean range for the slower performance. Figure 6 shows this prediction across t for two values of a, the initial activation parameter, with all w_js equal. The serial component decreases the activation of sequence elements at greater distances more quickly for the faster performance than for the slower performance, in essence damping the effects of metrical similarity for events from greater distances.

The final prediction of the model tested here concerns metrical position effects on item similarity. Because metrical similarity is defined relative to metrical position, the model predicts that metrical similarity and metrical position interact in event strength. For example, an event intended for a strong metrical position is more similar metrically and thus more likely to be confused with a current event on a strong metrical position than with a current event on a weak metrical position; the reverse is true for events intended for weak metrical positions. Figure 7 demonstrates these predictions by showing the model's event activation for a strong

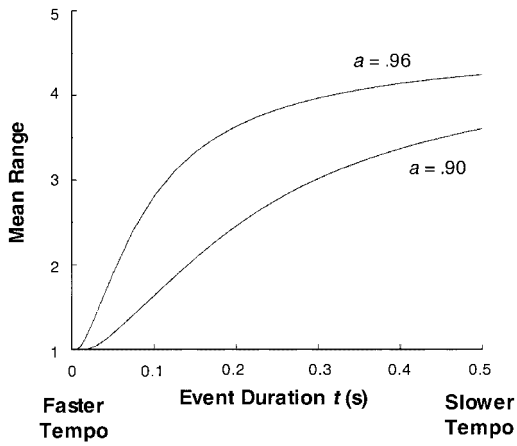

Figure 6. Mean range predictions across values of *t* (production rate) for two values of *a* (initial activation).

metrical position (marked in Figure 7 with an asterisk) when the current event (circled in Figure 7) is in a strong metrical position (Column 1) compared with the same event's activation when the current event is in a weak metrical position (Column 2). We compare the model's predicted activation patterns for each combination of metrical positions and sequence distances with obtained error frequencies. A plausible alternative is that metrical position alone, rather than its interaction with sequence distance, accounts for error distributions. We examine the relationship between metrical accent and metrical position in serial-ordering errors in this article.

Although the model's implementation here is restricted to one metrical cycle in either direction from the current event (vector of length $2n + 1$), this restriction is not a necessary condition; the model's predictions can be applied across the entire sequence. We extend the maximum values ($2n + 1$) of S, M, product vectors, and *x* (distance between current event and other events) to the length of the entire sequence minus 1 to generate the product model's activation predictions across the entire sequence, as shown in Figure 7. The metrical grid predictions repeat throughout the sequence: That is, m_is are defined for all positions *i* in the sequence as $m_{[(i-1) \bmod n]+1}$. The majority of the activation distribution falls within one metrical cycle of the current event, indicated by the solid bars in Figure 7. Model predictions generated across the entire sequence for *t* values ranging from .10 to .40 (including the range of *t* values reported in this article) in steps of .01 with the initial activation parameter held constant (at .880 and .960 values, as reported in this article) indicated that the mean proportion of the total activation across the sequence (excluding the current event, a

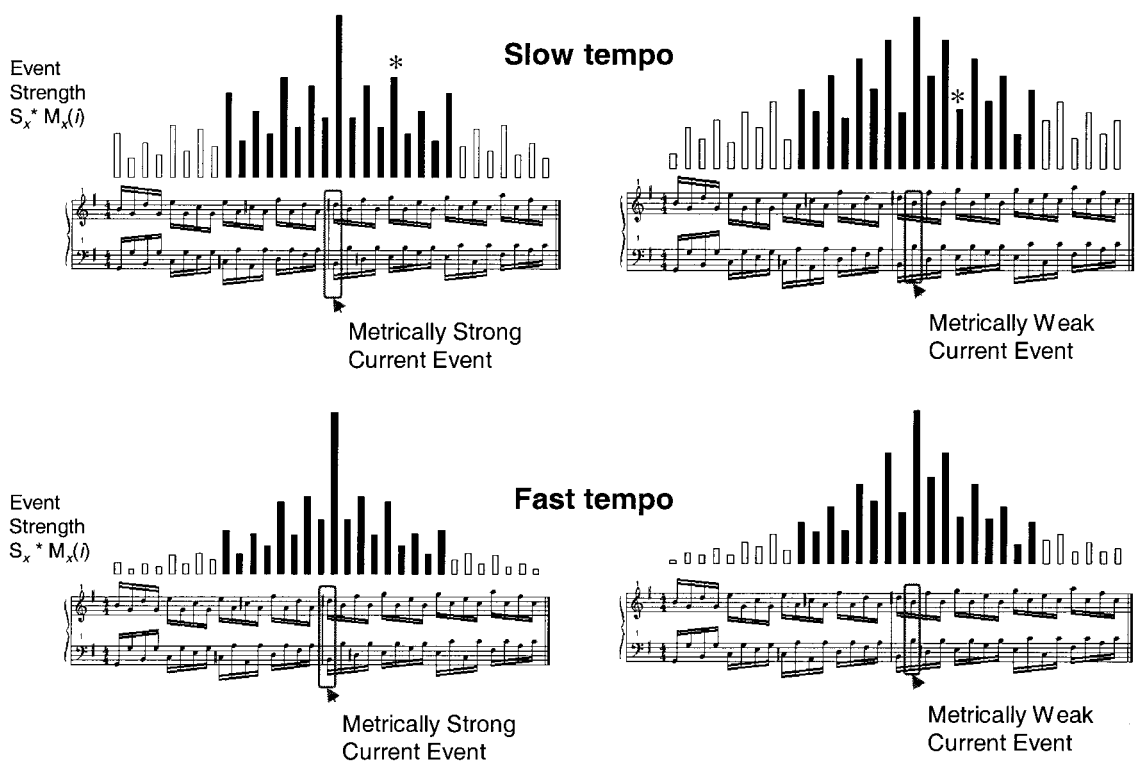

Figure 7. Unnormalized predictions of the product model ($a = .96$) for a slow tempo ($t = .50$, top row) and a faster tempo ($t = .25$, bottom row). Predictions are for a current event in a strong metrical position (left column) and in a weak metrical position (right column). Solid bars indicate event activations within one metrical cycle of the current event. Open bars indicate event activations outside one metrical cycle of the current event. Asterisks indicate predictions for the same sequence event when the current event is similar in metrical accent (left) and when the current event is dissimilar (right).

correct response) within one metrical cycle is 89%. The restriction of incremental planning measures to activations within one metrical cycle allows us to account for the chance estimates of how often randomly occurring errors would give rise to a sequence event (and be identified as a movement error) and still capture the majority of the model's predictions (and data, as we show later).

Two structural dimensions that are not incorporated in the model include rhythmic variation and phrase structure. Most musical sequences contain a variety of event durations that are distributed unequally across metrical positions. The model proposed here computes similarity on the basis of metrical accent rather than duration and can be applied to both rhythmically varying and unvarying (isochronous) sequences; we test its predictions with both types of musical sequences. Another likely constraint on retrieval processes in music performance is phrase structure. Musical phrases exist at many hierarchical levels and typically encompass more than one measure or cycle of the metrical grid. Because the majority of errors in music performance reflects interactions among elements within a single phrase (Palmer & van de Sande, 1995) and thus phrase structure alone cannot distinguish among these errors, phrase structure is unlikely to designate similarity relationships that account for memory confusions. Although the theoretical framework applies to sequences of any length, we assume that phrase structure creates boundaries that set upper limits on the range of planning; the model described here captures planning constraints for musical sequences that comprise a single phrase.

Overview of Experiments and Model Simulations

In the next section, two experiments are reported in which serial-ordering errors are elicited in music performances. The experiments are designed to test the range model's predicted effects of production rate as well as producers' age and/or experience on incremental planning. Pitch errors are identified and measured by computer, and movement gradients based on error proportions across sequence distances are compared with the range model predictions. The main assumptions of the model—four-tier metrical hierarchy of similarity, rate- and age-based decay parameters in the serial component, and a nonlinear combination of the metrical and serial components—are tested as well by contrasting models with and without these features in their ability to account for the serial-ordering errors.

In Experiment 1, skilled adult pianists performed musical exercises (musical sequences designed to elicit errors) at two production rates. The main predictions of the range model outlined earlier are tested: a smaller mean range of planning with increased rate, a characteristic movement gradient of error proportions across sequence distances that contains periodic (metrical) and sequential (serial) components, an interaction between rate and sequence distance on error proportions, and an interaction between metrical position and sequence distance on error proportions (event strength alters as a function of the current event's metrical position). Because the metrical component of the model reflects schematic knowledge that applies to many sequences, we expect no effect of short-term practice on range of planning. In Experiment 2, novice child pianists performed the same task, which permits examination of the roles of age and experience on planning constraints. The range model makes similar predictions for novices as for adults, with one important exception. The initial activation parameter, which represents short-term memory constraints, should be smaller in the model fits to the children's performances than to the adults' performances. Although age and experience are highly correlated in skills such as music and speech (there were too few adult novices and child experts to allow independent manipulation), the range of ages (8–40 years) and years of musical training (5–20 years) across participants in these studies varied sufficiently to allow separation of their relative contributions to planning measures.

A final goal of the article is to consider how range (distance) constraints on incremental planning combine with anticipatory (directional) constraints in production. One possibility is that the range and anticipatory processes constrain production independently; anticipatory tendencies are applied in speech by Dell et al. (1997) to predict the relative activation of events within a frame, which is structured by directional constraints such as future-oriented dependencies. The range model predicts the relative activation of events on the basis of metrical similarity and proximity, both of which are structured by bidirectional distance constraints across sequence positions. Another possibility is that anticipatory and range constraints are correlated. Dell et al.'s (1997) model predicts a smaller anticipatory proportion (AP) at faster production rates; the model presented here predicts a smaller range at faster production rates. Similar influences of rate on AP and range might underlie a correlation between the two. Some studies indicate increased AP and increased range measures across expertise levels of music performers (Drake & Palmer, 2000; Palmer & Drake, 1997). Production rate, however, was unconstrained, and thus the temporal characteristics of sequence planning on range and anticipatory behavior are unknown. We compare range and anticipatory constraints in music performance as well as the factors that influence each.

Experiment 1: Incremental Planning in Skilled Performance

The first experiment tested predictions of the range model in skilled adults' music performances. We used the same experimental design as in Dell et al.'s (1997, Experiment 3) study of experimentally elicited speech errors in which speakers viewed a four-word tongue twister presented visually and then produced it twice at one of two production rates indicated on a metronome, in each of eight practice blocks. In Experiment 1, 14 skilled pianists performed 32-event musical exercises designed to elicit errors (an example is shown in Figure 7) on a computer-monitored piano, first at a slow rate to ensure they had learned it correctly. Pianists then performed each sequence twice at either a fast tempo or a medium tempo, as indicated by a metronome in each of eight practice blocks, followed by an error-free performance at a slow rate to ensure that they had not relearned the sequence. Because the music stimuli were considerably longer than the tongue twisters used by Dell et al. (1997), the music notation remained in front of the performers; analyses of memorized performances are presented in a later section.

The range model predicts different error proportions across sequence distances; faster rates should lead to higher proportions of movement errors from nearby distances and a smaller mean range overall than slower rates. The range model predicts no

effects of short-term practice on range. The serial-order model of speech (Dell et al., 1997) predicts that practice and reduced rate should increase the anticipatory proportion of errors.

Pitch errors were identified and coded by computer in terms of their distance and directional properties; Figure 8 provides examples of movement error codings (see also Palmer & Drake, 1997; Palmer & van de Sande, 1993, 1995). Because musical sequences contain pitches that reoccur, errors that arise randomly may resemble sequence elements and be mistaken for movement errors. Therefore, movement errors for which the presumed source fell within a chance estimate of how often pitches repeated on average in each stimulus (within eight events, the length of one metrical cycle) were examined for range of planning estimates. Further details of the methods and error codings are given in Appendix B.

Results

There were 3,827 pitch errors in total, 81% of which were movement errors (those with an identifiable source in the stimulus).[2] Eighty-five percent of all movement errors involved pitch sources from the sequence within one metrical cycle of the error, comparing favorably with the model's predictions (based on the same production rates) that 85% of total sequence activation falls within one metrical cycle. The mean error rate was .082. Participants' performed tempo (measured by mean event interonset interval [IOI] in the first measure) did not differ from the instructed tempo in either condition (mean medium tempo = 188 ms/IOI, instructed = 188 ms/IOI; mean fast tempo = 154 ms/IOI, instructed = 150 ms/IOI; $p > .10$). With the exception of Block 1 (which was slightly slower), the performed tempo did not change over practice in either tempo condition. Figure 9 shows the effects of practice on error rate by tempo condition in log scale. There was a significant negative linear relationship between log error rate and log practice block within each tempo condition (fast tempo: $r = -.93$, $p < .01$; medium tempo: $r = -.89$, $p < .01$). The fast tempo generated significantly higher error rates than the medium tempo condition, $F(1, 13) = 18.8$, $p < .01$. Error rates also decreased with practice, $F(7, 91) = 18.8$, $p < .01$, and there was no interaction between tempo and practice.

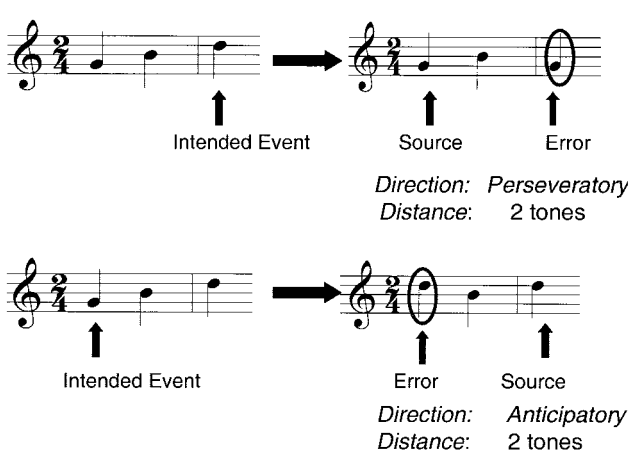

Figure 8. Coding examples for movement errors. Top: Perseveratory error. Bottom: Anticipatory error.

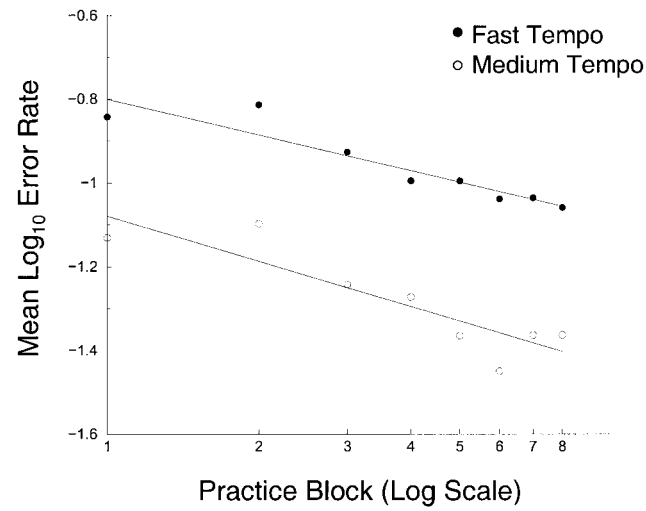

Figure 9. Log error rates by practice block and tempo condition for Experiment 1.

The range model predicts a larger span for movement errors at the slower tempo than at the faster tempo. The mean absolute error distance across performances is shown in Figure 10 by tempo condition and practice block. Three participants had no errors in over half of the medium tempo condition trials and were excluded from analyses of tempo effects. The mean distance was greater for the medium tempo condition (3.84) than for the fast condition (3.04), $F(1, 10) = 7.7$, $p < .05$, consistent with the range model predictions. There was no effect of practice block on mean range, also consistent with the model's predictions.

The range model also predicts a characteristic profile of error proportions across sequence distances and an interaction of this profile with production rate. The proportion of movement errors arising from each distance (1–8) was computed for each participant and tempo condition, combined over practice blocks; the mean error proportions are shown in Figure 11. There was a significant effect of distance, $F(7, 70) = 62.2$, $p < .01$, with more errors arising at shorter distances and at periodic distances that are integer multiples of 2. The interaction between tempo and distance approached significance, $F(7, 70) = 1.9$, $p = .08$. As shown in Figure 11, more errors tended to arise at shorter distances in the fast tempo condition and at larger distances in the medium tempo condition, consistent with the model's predictions. The error proportions did not correlate significantly with the chance estimates in either tempo condition.

[2] The single largest error type was deletions (19% of total errors). Although many studies of speech errors code some deletions as movement errors, the source of deletions in music is ambiguous. Deletions are more common in music performance, and musical events repeat more often than in speech; thus, deletions that arise by chance can appear contextual. As a conservative approach, deletions were excluded from analyses that focus only on movement errors. The same analyses repeated with deletions yielded results similar to those reported here.

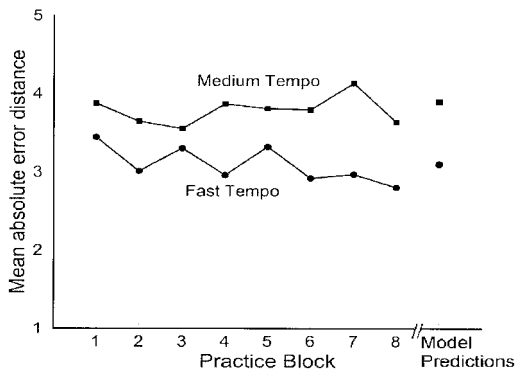

Figure 10. Mean error range and model predictions by practice block and tempo conditions for Experiment 1.

Model Fits

Model components. The range model was fit first to the mean error data from the medium and fast tempos using $t = .188$ s and .150 s (the instructed rates), respectively, in the serial component of the model and all w_js equal in the metrical component. Initial activation a, the only free parameter, was determined by choosing the estimate that minimized the sum of the squared error term for the mean data of each tempo condition across a range of values from .800 to .999 in increments of .001. The mean value of a across tempo conditions represented the optimal initial activation, from which the value of the Akaike information criterion (AIC) was computed for each model fit. AIC, which reflects a model's ability to capture the underlying process in the presence of random error (Akaike, 1973), allows comparisons among models of varying complexity; smaller numbers indicate less difference between model and data. Variance accounted for (VAF), which indicates the relative goodness of fit (unadjusted for number of parameters), is also reported for each model.[3]

Predictions of the best-fitting unweighted range model are shown in Figure 11 (top row) for the mean data of both tempo conditions. The initial activation value was .960 (AIC = −22.01), and the model accounted for a significant proportion of the variance (VAF = .57, $p < .01$). Initial activation values were then estimated for model fits to individual data in a similar fashion; estimates were first generated for each tempo condition on the basis of individual tempo (t) estimates, and the average value across tempo conditions represented the optimal initial activation for an individual participant. Fits of the range model to each participant's data indicated a values from .949 to .973 (mean AIC = −18.48), and the model accounted for a significant portion of the variance for 10 of 11 participants (mean VAF = .45, $p <$ .05). Furthermore, the initial activation parameter values did not change in model fits to the data of the first four practice blocks and the last four blocks for the mean data or for participants.

Does the range model fit the data in each tempo condition better than the data in the other tempo condition? This comparison of fits, based on t values from the correct tempo condition and the alternate tempo condition, allows a test of the role of the t parameter in the serial component independent of the a parameter (a was held constant at .960). The medium tempo data were fit better by the medium tempo predictions (AIC = −25.00, VAF = .59, $p < .01$), and the fast tempo data were fit better by the fast tempo predictions (AIC = −19.10, VAF = .55, $p < .01$). The same analyses conducted on individual data indicated that the appropriate t-parameter values provided a better fit than the alternate values for 7 of 11 participants.

Next, we tested the assumptions of the metrical component. The metrical component's assumption of four maximum hierarchical levels was tested by generating predictions for $k = 1$, 2, 3, and 4 hierarchical levels (following Equation 1) with initial activation held constant (.960). The four-level model fit the mean data better than the other models for both tempo conditions and for 10 of 11 individual fits. We also tested the assumption that metrical accent is computed in terms of relative similarity (as in Equation 2) rather than absolute similarity. Predictions of the range model were generated for absolute similarity measures by removing the denominator term in Equation 2, keeping initial activation constant (.960). The predictions based on relative similarity fit the mean data better than the predictions based on absolute similarity for both tempo conditions and for 10 of 11 individual fits. Relative similarity can provide a more sensitive measure than absolute similarity because it generates distinctive predictions for pairs of events that differ in metrical accent by the same absolute amount.

The range model was contrasted with three other models—the metrical (M) component alone, the serial (S) component alone, and an additive model (S + M, with one parameter, initial activation)—as well as with the chance estimates shown in Figure 11. Normalized values of all models were compared first with the mean data of each tempo condition. Initial activation was optimized across models (the metrical component contained no initial activation parameter);[4] the best-fitting a that minimized the AIC across models was .963. As shown in Table 1, both the medium and fast tempo data were fit better by the range model than by any of the other models or the chance estimates. Model comparisons for each participant's data, with a optimized as before across tempo conditions, confirmed that the range model fit better than the other models or the chance estimates for all 11 participants.

We tested the model's final prediction that metrical similarity reflects an interaction of metrical position with sequence distance to yield different error distributions across metrical positions. Although there were too few errors arising at each combination of metrical position and sequence distance to test individual data, the correlations between model predictions and mean error proportions across metrical positions and sequence distances were significant, both in the fast tempo condition ($r = .62$, $p < .01$) and in the medium tempo condition ($r = .58$, $p < .01$). Furthermore, the obtained mean error proportions did not correlate as strongly with either the metrical component alone (fast: $r = .27$; medium: $r =$

[3] VAF is computed as $1 - (SS_E/SS_T)$ or $1 - \Sigma(\text{data} - \text{model})^2/\Sigma(\text{data} - \text{mean})^2$, where SS_E = sum of squares—error, SS_T = sum of squares—total, and mean = mean data. The AIC is superior to the VAF, as it accounts for number of parameters as well as goodness of fit, and is well-suited for nonlinear models (Myung, 2000; Myung & Pitt, 1998).

[4] Initial activation parameters were estimated for each model that incorporated the parameter (all but the metrical component) and each tempo condition. The mean estimate across tempo condition and model combinations was taken as a measure of the optimal initial activation value to avoid weighting the estimates unduly by tempo or model.

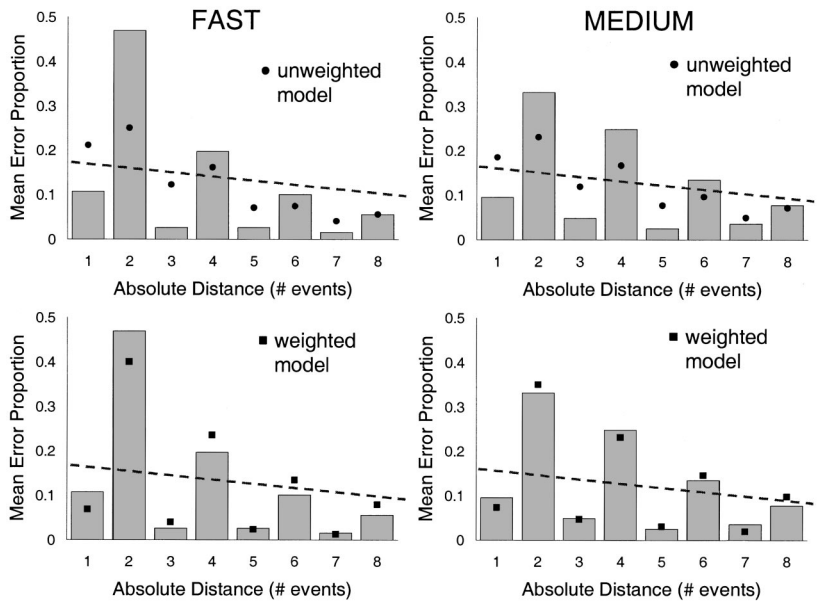

Figure 11. Mean error proportions and model fits by sequence distance and tempo conditions (left graphs: fast; right graphs: medium) for Experiment 1. Top row: Unweighted model fits (w_j = .25). Bottom row: Weighted model fits (w_2 = .76). The dashed lines indicate chance estimates.

.43) or the serial component alone (fast: r = .45; medium: r = .32), providing confirmation that the range model provides a better fit than either of the individual components.

Metrical weights. Next, we compared the unweighted range model, in which metrical levels are equally weighted (all w_js are equal), with a model in which the weight on one metrical level, l, differs and the remaining weights are set to equal values [$(1 - w_l)/(k - 1)$]. The weighted model therefore has two additional parameters: l, the metrical level being weighted (1 to 4), and w_l, its weight value, which ranges from .25 (the constant value) to .99 in increments of .01. Increasing the metrical weight at one level increases the metrical accent of events aligned with that level, relative to events aligned with other levels. This in turn increases the similarity between events aligned with that level and decreases the similarity between events aligned with that level and events aligned with other levels. (The metrical similarity between any two events not aligned with the weighted metrical level is unaffected by the weighting because the remaining metrical weights are set to equal values.)

The weighted model was first fit to the mean data with metrical weights optimized for both tempo conditions. Initial activation was held constant (a = .960) in comparisons of weighted and unweighted models. The best-fitting weighted model for the mean error proportions, indicated by squares in the bottom row of Figure 11, had an increased weight on Metrical Level 2 (w_2 = .76). Table 2 shows that the fits of the weighted model surpassed those of the unweighted model for both tempo conditions, even when two additional parameters were taken into account (the AIC values incorporate model complexity). The weighted model also surpassed the unweighted model for all individual participants (weighted model: mean VAF = .74, $p < .01$; unweighted model: mean VAF = .45, $p < .05$), and the weight was highest on Metrical Level 2 (eighth-note level) for all of the participants (mean weight = .741). To assess the role of tempo on metrical

Table 1
Model Comparisons for Adults' Mean Performance in Experiment 1

	Medium tempo condition		Fast tempo condition	
Model	AIC	VAF	AIC	VAF
Product	**−25.1**	**.60****	**−18.9**	**.54****
Serial	−18.8	.12	−14.7	.23
Metrical	−20.5	.29	−14.0	.15
Additive	−21.1	.34*	−15.0	.26
Chance estimate	−20.4	.07	−15.2	.07

Note. Boldface indicates the best-fitting model values. AIC = Akaike information criterion; VAF = variance accounted for.
* $p < .05$. ** $p < .01$.

Table 2
Comparisons of Weighted and Unweighted Models for Adults' Mean Performance in Experiment 1

	Medium tempo condition		Fast tempo condition	
Model	AIC	VAF	AIC	VAF
Weighted	**−43.9**	**.98****	**−30.4**	**.93****
Unweighted	−25.0	.59**	−19.1	.55**

Note. Boldface indicates the best-fitting model values. AIC = Akaike information criterion; VAF = variance accounted for.
** $p < .01$.

weights, we repeated the same model fits with the metrical weights allowed to vary across tempo conditions; again, weighting of Metrical Level 2 provided the best fit for each tempo condition, both for the mean data and for all participants. Thus, the range model with weighting of metrical levels provided the best fit overall and accounted for most of the variance.

Finally, the model's applicability to the remaining errors (those without an identifiable source in the sequence) is considered. Deletion errors (19% of total errors and the most common type of nonmovement error) may arise from weaker activation values for current events with lower metrical accent strength. To test this, we compared error rates for deletions of events that corresponded with strong beats (three or four levels of metrical accent) or weak beats (one or two levels of metrical accent). Deletion error rates (adjusted for number of sequence events that aligned with each metrical accent) were almost twice as high on weak beats (.025) as on strong beats (.014; $p = .05$), suggesting that the model's metrical component modulates the activation of current events as well as those of surrounding events. Second, we tested whether nonmovement errors reflected diatonic similarity-based confusions by comparing the percentage of pitch error intruders within the key with the percentage expected by chance (7 out of 12 pitch classes, or .583). Performers' proportions of diatonically related errors were higher than expected by chance ($M = .80$), $t(10) = 8.3$, $p < .01$, indicating that the majority of nonmovement errors, like the movement errors, showed contextual influences of structural similarity. Finally, the range analyses repeated for movement errors that arose across the entire sequence (without a chance estimate constraint) yielded the same effects of production rate on range as found for movement errors within one metrical cycle.

Comparisons With Directional (Anticipatory) Constraints

Here, we compared the range constraints with the anticipatory predictions of Dell et al.'s (1997) serial-order model of speech, applied to music performance. First, Dell et al.'s prediction that perseveratory errors decrease more over practice than anticipations was examined; the mean number of anticipations and perseverations decreased over practice, $F(7, 91) = 3.1$, $p < .01$, and there was greater decrease in perseverations than anticipations over practice, $F(7, 91) = 2.7$, $p < .05$. The log–log slope for number of errors by practice block was negative for perseveratory errors ($r = -.65$, $p = .08$) and near zero for anticipatory errors ($r = -.08$, $p > .10$), replicating the anticipatory practice effect seen in speech. AP, computed as the number of anticipatory errors in each performance divided by the sum of anticipatory and perseveratory errors [A/(A + P)], also increased over practice, $F(7, 91) = 4.0$, $p < .01$. The general anticipatory effect (correlation between AP and log error rate) by practice block approached significance ($r = -.69$, $p = .06$). The analyses repeated without the 3 participants with few errors indicated that the general anticipatory effect by practice block was significant ($r = -.86$, $p < .05$). As shown in Figure 12, performances early in the practice session were marked by high error rates and low AP, whereas later performances showed lower error rates and higher AP.

Next, the predictions of rate effects on anticipatory constraints were examined: Slower production rates should result in higher AP than faster rates (Dell et al., 1997). Overall, there was no difference in AP across the medium and fast tempo conditions. Figure 13 shows the relationship between AP and log error rate for each tempo condition (the 3 participants with too few errors were excluded); there was a significant negative relationship (a general anticipatory effect by practice block) for the fast tempo condition ($r = -.81$, $p < .05$) and a nonsignificant effect for the medium tempo condition ($r = -.59$, $p = .12$). Most important, the two tempo conditions do not fall along the same negative slope, as the anticipatory rate effect predicts ($r = .26$, $p > .10$). Therefore, rate had different influences on serial ordering in music performance than predicted by the speech model (Dell et al., 1997).

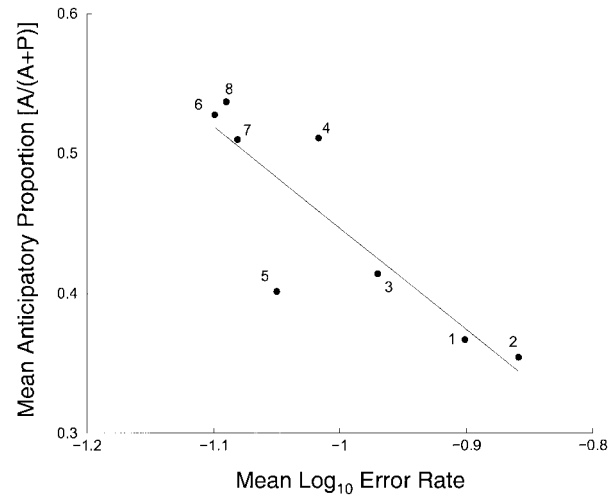

Figure 12. Mean anticipatory proportion of errors by log error rate and practice block (indicated by numbered circles) for Experiment 1. A = anticipatory errors; P = perseveratory errors.

To address how range and AP constraints combine, we first tested their relative contributions to overall error rate.[5] Mean log error rate, regressed on the mean values of range and AP for each combination of tempo and practice block (the factors that influenced range and AP, respectively), indicated a good fit overall ($R = .74$, $p < .01$). As error rates decreased, both range ($\beta = -.94$, $p < .01$) and anticipatory tendencies ($\beta = -.73$, $p < .01$) increased. AP and range predictors were negatively correlated ($r = -.63$, $p < .01$). The same regression model conducted on individual performances did not reach significance ($R = .10$) but showed the same pattern of results. To test further whether range and AP are related, we computed the mean range separately for anticipatory errors and for perseveratory errors. That difference was not significant within or across tempo conditions (anticipations: mean range = 3.34; perseverations: mean range = 3.48). A second test examined whether the movement gradients differed for anticipations and perseverations, as suggested by Vousden et al. (2000). There were no differences in the movement gradients for anticipatory and perseveratory errors within or across tempo conditions, suggesting again that directional and distance constraints do not change together.

[5] Comparisons of error distance and direction were based on the same subset of total errors (those that occurred within the chance estimate of one metrical cycle).

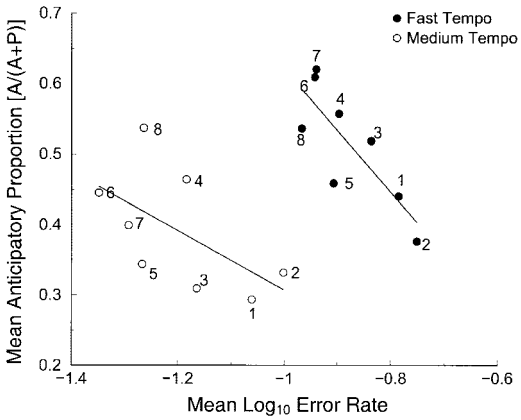

Figure 13. Mean anticipatory proportion of errors by log error rate and production rate for Experiment 1. Numbered circles indicate practice blocks. A = anticipatory errors; P = perseveratory errors.

Finally, simple correlations were conducted to identify underlying factors that contribute to range and AP. The correlation between the mean range and mean performed tempo (t) within practice and tempo condition was high ($r = .87$, $p < .01$), and the correlation between the mean AP and practice block was high ($r = .60$, $p < .05$). The same pattern of results was found for individual performances; as tempo slowed, range increased ($r = .22$, $p < .01$), and as practice increased, AP increased ($r = .16$, $p < .01$). Range values did not correlate with practice trial, and AP values did not correlate with tempo. These results, together with the findings that range of planning is influenced by tempo (not practice) and AP is influenced only by practice (not tempo), suggest that range and AP reflect distinct mental constraints.

Discussion

Serial-ordering errors in music performance were influenced by both proximity and metrical similarity of sequence events. The main predictions of the range model were supported: Error proportions showed a distinctive profile that reflected contributions of metrical similarity and serial proximity across sequence distances, faster production rates led to higher error proportions from nearby distances and a smaller mean range overall than slower rates, and error proportions changed as a function of the current event's metrical position. Production rate influenced serial-ordering errors in ways consistent with the range model's predictions but not with the anticipatory predictions of the speech model (Dell et al., 1997). The range model's nonlinear combination of proximity and similarity components described the error patterns better than either of the individual components or a simple additive model. Furthermore, each of the model's assumptions—number of metrical levels, relative versus absolute metrical similarity, and rate-dependent constraints on the serial component's decay—improved the model's ability to account for the data. When metrical weights were added, the model accounted for nearly all of the variance in the mean data (93%–98%).

The initial activation parameter values associated with the model's decay function were high for all of the individual performers. We assume that a reflects the large memory capacities of the adults in this study. Developmental studies indicate that memory span capacities are more constrained in children under 14–16 years of age. Initial activation may be influenced by music performing experience also; some performance-error studies indicate increased sensitivity to musical structure with musical experience (Drake & Palmer, 2000; Palmer & Drake, 1997). If incremental planning reflects short-term memory constraints or musical experience, then novice children's movement errors should exhibit smaller sequence distances than skilled adults', on the basis of previous measures of smaller range of planning for child than for adult pianists (Drake & Palmer, 2000; Palmer & Drake, 1997). In addition, the initial activation parameter values should be smaller in model fits to individual children's movement errors, which may correlate with age or musical experience. We test these predictions next.

Experiment 2: Incremental Planning in Novice Performance

Twelve novice child pianists performed musical exercises similar to those in Experiment 1 (an example is shown in Figure 2) on a computer-monitored piano. The design and procedure were the same as in Experiment 1, with the following adjustments that simplified and shortened the task. The children performed each stimulus only once in each block, the production rates were slower than adults' rates, and the children were reminded to try to maintain the rate indicated on the metronome. The error codings and analyses were conducted as in Experiment 1. Further details of the methods are given in Appendix C.

Results

There were 1,157 pitch errors in total, 87% of which were movement errors (those with an identifiable source in the stimulus).[6] Ninety-three percent of the movement errors involved pitches from the sequence within one metrical cycle of the error. This compares favorably with the model's predictions generated for the children's production rates that 92% of total sequence activation falls within one metrical cycle. The children's mean error rate was .139. One of the children performed significantly slower in the fast tempo condition than in the medium tempo condition; this participant's data were removed from further analyses. Both tempo conditions were performed slower than the prescribed tempo (mean fast tempo = 347 ms/IOI, instructed = 300 ms/IOI; mean medium tempo = 399 ms/IOI, instructed = 375 ms/IOI), but the fast tempo performances were significantly faster than the medium tempo performances, $F(1, 10) = 17.0$, $p < .01$. The mean tempo was faster over practice, $F(7, 70) = 4.7$, $p < .01$, and was slowest in Block 1 for both tempo conditions. There was no interaction between tempo condition and practice block.

Mean error rates were higher in the fast tempo condition (.150) than in the medium tempo condition (.128), although the difference was not significant. With the exception of Block 1 (after which the experimenter reminded participants to perform at the

[6] The largest single error type was substitutions (43% of total errors). As in Experiment 1, deletions were excluded from further analyses of movement errors. Again, results that incorporated deletions were similar to those reported here.

metronomic rate), error rates decreased significantly over practice, $F(6, 60) = 2.5$, $p < .05$. The log–log relationship between error rates and practice when Block 1 was excluded was significant ($r = -.94$, $p < .01$). The log–log fit for Blocks 2–8 was significant for the medium tempo condition ($r = -.85$, $p < .05$) and was negative but not significant for the fast tempo condition ($r = -.51$, $p = .25$).

Children's mean absolute error distance did not differ by practice block or tempo condition. Two children did not perform significantly faster at the fast tempo than at the medium tempo; the analyses repeated with their data excluded indicated the same results, and their data were excluded from remaining analyses for which production rate was the crucial predictor. Comparisons of children's error distances with those of the adults in Experiment 1 indicated a smaller mean range for children (3.18) than for adults (3.50), $F(1, 18) = 5.7$, $p < .05$, and an interaction of group and tempo, $F(1, 18) = 6.7$, $p < .05$. Children's range values in both tempo conditions resembled adults' smaller range values for the fast tempo condition.

Next, we tested the model's prediction that the movement gradients display a characteristic pattern of error proportions across sequence distances and an interaction of this pattern with production rate. The proportion of movement errors arising from each distance is shown for each tempo condition in Figure 14. Error proportions differed over sequence distances, $F(7, 56) = 17.1$, $p < .01$, with more errors arising at shorter distances and at distances that are integer multiples of 2. Furthermore, tempo interacted with distance effects, $F(7, 56) = 3.4$, $p < .01$. Figure 14 displays the model's prediction of stronger metrical effects in performances at slower rates than at faster rates. The error proportions were not correlated with the chance estimates (indicated in Figure 14 by the dashed lines) in either tempo condition.

Model Fits

Model components. The range model was first fit to the mean data of each tempo condition, on the basis of the 9 children's performed tempi of $t = .402$ s for the medium tempo and $t = .332$ s for the fast tempo condition. The model's predictions, also shown in Figure 14, were optimized as before (mean AIC $= -29.77$) and accounted for a significant portion of the variance in the data of each tempo condition (mean VAF $= .65$, $p < .01$). The best-fitting initial activation value (the only free parameter) was $a = .880$. Applied to individual data (based on individual t estimates), the model accounted for a significant portion of the variance for all 9 participants (mean VAF $= .46$, $p < .05$). Furthermore, the initial activation parameter values did not change in model fits to the first four and last four practice trials for mean data or individual data.

We tested whether the initial activation parameter values, which are assumed to capture short-term memory constraints, were lower for children than for adults. Model fits to individual data indicated significantly lower initial activation values for the children (range = .800–.940) than for the adults (range = .950–.970), $t(18) = 6.3$, $p < .01$. Because age and years of musical training were correlated among children and adults ($r = .56$, $p < .01$), a multiple regression analysis was conducted to examine relative contributions of age and experience to the individual initial activation values of children and adults. The overall fit was significant ($R = .70$, $p < .01$), and both predictors correlated positively with initial activation (age: $r = .68$; musical training: $r = .49$). Most important, partial correlations indicated that age predicted initial activation values above and beyond musical training ($\beta = .59$, $p < .01$); however, musical training did not contribute beyond age ($\beta = .16$, $p > .10$). These findings suggest that age influenced

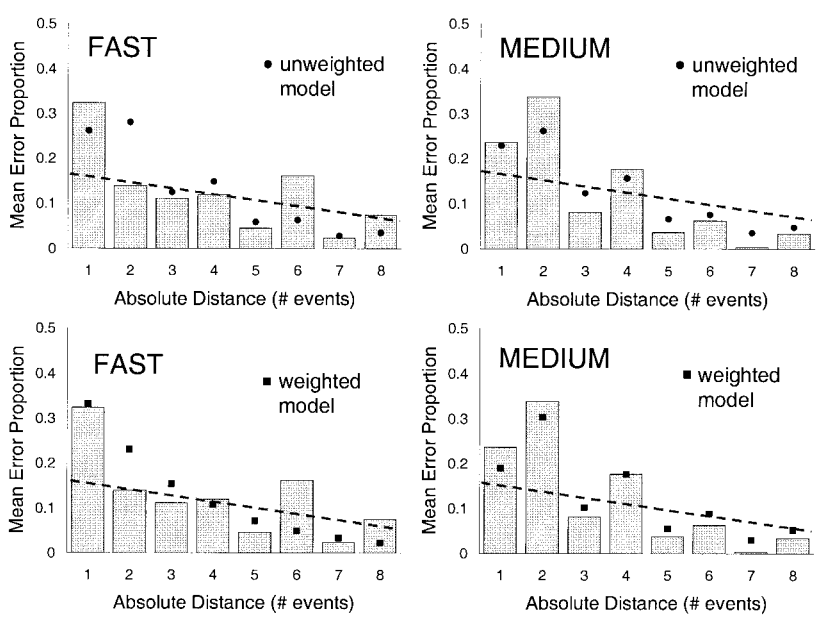

Figure 14. Mean error proportions and model fits by sequence distance and tempo conditions (left graphs: fast; right graphs: medium) for Experiment 2. Top row: Unweighted model fits ($w_j = .25$). Bottom row: Weighted model fits (fast tempo, $w_1 = .91$; medium tempo, $w_2 = .42$). The dashed lines indicate chance estimates.

initial activation values beyond musical training, such that younger participants had stronger serial and temporal decay.

The range model was contrasted with three other models—the metrical (M) component alone, the serial (S) component alone, and an additive model (S + M, with one parameter, initial activation)—as well as with chance estimates. The initial activation parameter (the only free parameter) was optimized simultaneously across the normalized model predictions, as before. As shown in Table 3, the mean data for the medium tempo condition were fit better by the range model than by any of the other models or the chance estimates. The mean fast tempo data were fit better by the serial component alone than by any of the other models or the chance estimates. The range model fit the individual data for the medium tempo condition better than the chance estimates for all participants and better than the other three models for 6 of 9 participants. No one model best fit the individual data for the fast tempo condition.

The range model's assumption of a four-tier metrical component was tested in comparisons with one-, two-, and three-tier models. The four-tier model surpassed the other models in the medium tempo condition; individual fits confirmed the better fit of the four-tier model for all participants. Although a one-tier (serial) model provided a better fit for the mean data of the fast tempo condition, no one model provided a best fit to the individual data. The range model's assumption of relative metrical similarity was compared with predictions of absolute (unadjusted) similarity (by removing the denominator from Equation 2). The relative similarity predictions surpassed the absolute similarity predictions in fits to the mean data of the medium tempo condition, but the serial model provided a better fit for the mean data of the fast tempo condition than either the relative or absolute similarity metrics. The relative similarity metric surpassed the absolute similarity metric for 8 of 9 participants.

The range model's predictions that metrical similarity changes with metrical position and sequence distance were correlated with the mean error distributions in the children's performances. The correlations across metrical positions and sequence distances were weaker than those of the adults overall but were significant both in the fast tempo condition ($r = .37$, $p < .01$) and in the medium tempo condition ($r = .59$, $p < .01$). The correlation of the model's serial component predictions alone with the mean data of the fast tempo condition was also significant ($r = .39$, $p < .01$).

Table 3
Model Comparisons for Children's Mean Performance in Experiment 2

	Medium tempo condition		Fast tempo condition	
Model	AIC	VAF	AIC	VAF
Product	**−34.4**	**.89****	−24.6	.41*
Serial	−25.5	.67**	**−27.5**	**.59****
Metrical	−17.0	.03	−19.1	−.18
Additive	−20.6	.39*	−23.8	.34*
Chance estimate	−21.7	.32	−25.5	.32

Note. Boldface indicates the best-fitting model values. AIC = Akaike information criterion; VAF = variance accounted for.
* $p < .05$. ** $p < .01$.

Metrical weights. The unweighted range model was compared next with a model in which the weight on one metrical level differs and the remaining weights are set to equal proportions $[(1 - w_l)/(k - 1)]$. The weighted model had two additional parameters: l, the metrical level being weighted (1 to 4), and w_l, its weight value. The weighted model was fit to the data of each tempo condition, with initial activation values held constant from the optimized fit of the unweighted model ($a = .880$). The weighted model surpassed the unweighted model for the medium tempo condition (AIC = −36.71, VAF = .95, $p < .01$); the best-fitting weighted model, shown in the bottom right panel of Figure 14, had an increased weight on the second metrical level (quarter-note level; $w_2 = .42$). Fits to individual data indicated likewise that the weighted model surpassed the unweighted model for 7 of 9 participants in the medium tempo condition.

The best-fitting weighted model for the mean data of the fast tempo condition emphasized Level 1 (eighth-note level; $w_1 = .91$, AIC = −23.12, VAF = .57, $p < .05$). The product model predictions are identical to the serial model when Level 1 is maximally weighted ($w_1 = 1$) because all values in the Vector **M** equal 1 when other metrical weights, w_j, equal 0 (this follows from Equation 2 when all $m_i = 1$) and the product model S × M = S when M = 1. Not surprisingly, the predictions of the serial component and best-fitting weighted model were highly similar ($r = .999$). Because of the two additional parameters, the weighted model yielded a higher AIC (an inferior fit) than that of the serial model (serial AIC = −27.12), although it accounted for the same amount of variance. Likewise, the AIC values for the serial model's fits to individual data surpassed those of the weighted model, although the weighted model accounted for more of the individual variance on average (weighted model: mean VAF = .40; serial model: mean VAF = .30). The weighted product model generalizes to a broader set of experimental findings; the weighted model accounts for all of the main experimental conditions (both adult tempo conditions and the child medium tempo data), and its ability to surpass the serial model in other data sets is demonstrated later. Therefore, the weighted model's greater complexity than the serial model (which causes the inferior AIC scores) is warranted by its ability to generalize.

Finally, the model's applicability to other error types is considered. Deletion errors, a subset of nonmovement errors (22% of total errors), were examined for effects of metrical accent strength by comparing deletion rates on strong beats (three or four levels of metrical accent) and on weak beats (one or two levels of metrical accent). Deletion error rates did not differ by metrical accent but were slightly higher on weak beats (.029) than on strong beats (.027). Next, nonmovement error intruders were examined for evidence of diatonic similarity-based confusions. The percentage of error intruders that were diatonically related (within the key of the sequence) was higher ($M = .82$), $t(8) = 3.5$, $p < .01$, than expected by chance (7 out of 12, or .583), indicating that the majority of nonmovement errors, like the movement errors, showed contextual influences of structural similarity. Finally, the range analyses conducted on all movement errors (without the chance estimate constraint) yielded the same effects of production rate on range measures.

Comparisons With Directional (Anticipatory) Constraints

We tested how the range constraints combine with the anticipatory predictions of Dell et al. (1997) in children's performance. The children's movement errors did not become more anticipatory over practice; the log–log slope of number of errors by practice block was negative for anticipatory errors ($r = -.71$, $p < .05$) and not significant for perseveratory errors, the opposite of the predicted anticipatory practice effect (Dell et al., 1997). The AP did not change reliably over practice blocks or tempo conditions, although the AP for the medium condition was slightly higher (.51) than for the fast tempo condition (.44). The children's performances also failed to show the anticipatory rate effect. The two tempo conditions did not fall along the same regression line of AP and log error rate, as predicted by the general anticipatory effect, and only the medium tempo condition showed a negative (nonsignificant) relationship across practice blocks ($r = -.40$, $p > .10$).[7]

Next, we tested contributions of range and AP measures to overall performance by predicting log error rates from an additive combination of range and AP, on the basis of the children's individual performances. The regression model provided a modest but significant fit ($R = .24$, $p < .01$), and both AP ($\beta = -.18$) and range ($\beta = -.24$) contributed significantly; as error rates decreased, both anticipatory tendencies and range of planning increased ($p < .01$). AP and range predictors were negatively correlated ($r = -.38$, $p < .01$). The same regression analysis on averaged performances across block and tempo was not significant but showed the same pattern of results. Simple correlations conducted to test factors that contributed to range and AP predictors did not reach significance for the children's data, probably because tempo and practice effects on overall error rates were small. In general, range and AP increased as overall error rates decreased, indicating that scope and directional constraints on planning influenced novice children's ordering errors in similar but weaker ways to those of adults.

Discussion

Novice children's performance errors indicated proximity and metrical similarity influences on serial ordering, consistent with the range model predictions. Fits of the range model indicated that novice children's initial activation parameters were smaller and did not overlap with those of skilled adults in Experiment 1. Although initial activation values increased with both age and musical experience, individual differences in initial activation were best accounted for by age; musical experience did not contribute beyond age. Overall, these findings are consistent with the assumption that the initial activation parameter reflects slow-changing memory constraints rather than factors such as short-term practice.

Similarity-based interference, based on metrical accent strength, also influenced children's serial-ordering errors. The metrical similarity effects are further evidence that performers' sensitivity to metrical accent strength does not require many years of exposure (Drake & Palmer, 2000; Palmer & Drake, 1997). Although the children's performances at the fast tempo were accounted for by both the weighted range model and a serial model, only the range model accounted for performances at the slow tempo. Thus, the range model is preferable to a serial model because it is the only one that generalizes across production rates and skill levels (children's and adults' data). The fast tempo used in these experiments may have reached the limits of the novice children's performance rates, as evidenced in less variance accounted for by either model. Finally, children weighted lower metrical levels at fast production rates than did adults, consistent with other evidence that children tend to focus attention on smaller time spans in auditory patterns than do adults (Drake et al., 2000).

General Discussion

We have described a theoretical framework for retrieval in music performance that specifies the activation of each sequence event during production. Incremental planning arises from a combination of hierarchical, serial, and temporal constraints on retrieval during production. Performers' memory for serial order is viewed in this framework in terms of response preparation; each event's activation encodes contextual information about the serial order of other sequence events. Metrical (hierarchical) similarity constraints combine with serial and temporal proximity constraints to determine the planning increment: the subset of sequence events that are accessible at any point during production. The metrical similarity component boosts the activation of metrically similar events, and the serial proximity component dampens the activation of events at progressively greater sequence distances and faster rates. Finally, the similarity and proximity components combine nonlinearly to modulate the range of planning differentially at each sequence distance and production rate. Thus, our approach defines the planning of serial order in terms of timing; the theory distinguishes among memory for musical sequences that have the same serially ordered events but different rates.

The serial proximity component of the model predicts that memory for sequence events decays exponentially, on the basis of an initial activation parameter. Initial activation reflects short-term memory constraints that change developmentally; the model predicts increased planning scope with age and experience, consistent with developmental accounts of increased memory span. Measures of increased memory in skilled tasks are consistent with proposals that tie working memory abilities to the hierarchical organization of retrieval structures acquired with skill (Ericsson & Kintsch, 1995). However, our findings suggest that expansions to planning scope (as evidenced by the initial activation parameter) in the context of performance result more from age than experience. Furthermore, the initial activation parameter's exponent, the ratio of serial separation to temporal separation, is related to two general assumptions about conditions under which listeners associate successive events with the same auditory stream or sequence. First, the absolute temporal separation between events (0.1–2.0 s) constrains which events can be incorporated into the same sequence (Drake et al., 2000; Fraisse, 1984; Grondin, 2001). Second, the ratio of pitch separation to temporal separation between tones (analogous to the serial/temporal separation) influences their likelihood of being perceived as part of the same pattern (Jones, 1976;

[7] The same results were found for children's anticipatory practice and rate effects when Practice Block 1 (which was performed slower) and 2 participants whose fast tempo was not faster than their medium tempo were removed from the analyses.

Van Noorden, 1975). Although several retrieval models predict highest activation for proximal events in time or in number of events (Anderson & Matessa, 1997; G. D. A. Brown et al., 2000; Estes, 1997), the relationship of serial and temporal separation posited here makes unique predictions for rate-dependent effects on retrieval.

The metrical similarity component defines event similarity in terms of a hierarchical frame of periodically recurring accents, consistent with metrical grid theory. Metrical similarity is defined in relative, rather than absolute, terms in the range model; two events with different metrical accent strengths are more similar if their absolute metrical accents are large than if they are small. These predictions are consistent with perceptual identification findings that the relative (not absolute) separation of items along a perceptual dimension affects their discriminability (Bower, 1971; Garner, 1962; Murdock, 1960; Neath, 1993). Metrical frames are advantageous because they are ubiquitous across a variety of musical styles, as indicated by the statistical characteristics of our corpus analysis; they have a wide range of acoustic correlates in production; and their regularity offers perceptual and memory advantages for pattern recognition. The similarity-based retrieval predictions of the model can be extended to other musical behaviors. For example, the range model can be applied to address overall similarity between melodies that differ in rate or meter; forgetting of particular events within melodies; and segmentation of melodies into smaller units, on the basis of activation differences between successive sequence events.

In the next section, we consider the model's applicability to other performance tasks and sequences. Then, we compare the assumptions of the theoretical framework with those of serial-recall theories and with accounts of retrieval in the context of sequence production. We consider how domain specific the framework is through comparisons with language production. Finally, we discuss how different planning constraints (including scope and anticipatory tendencies) combine in music performance.

Model Extensions

In this section, we apply the range model to other music performance contexts. Although the model reflects only one dimension of similarity (that of metrical accent), many musical dimensions influence similarity judgments and confusion errors in comprehension and performance, including melodic contour, tonality, harmony, rhythm, and timbre (Crowder & Pitt, 1992; Cuddy, Cohen, & Miller, 1979; Dowling & Fujitani, 1971; Krumhansl, 1990; Palmer & van de Sande, 1993, 1995). First, we address the model's application to performance from memory (with musical notation absent during performance). Second, we address how the model applies to other meters, particularly ternary meters. Third, we test further the assumption that initial activation reflects age-related memory differences in extensions to child and adult performances of the same musical sequence. Finally, we test the range model's weighting of metrical levels by comparison with performances in which performers intentionally weighted a particular metrical level. Although the weighted model provided a better fit to the current experiments than the model without weights, the choices of metrical levels and weights were ad hoc. Each of these extensions incorporates more complex musical sequences than the musical exercises and thus provides a test of generalization.

The first extension tests the model's applicability to performance in the absence of visual stimulus presentation. The experimental findings reported above relied on stimulus presentation through music notation, which may have reduced memory demands in the production task. Other factors may contribute to performance from notation; for example, experience and age influence visual search (Clancy & Hoyer, 1994), and novice–expert differences in chess players' visual span have been attributed to perceptual coding factors, on the basis of eye-fixation differences (Reingold, Charness, Pomplun, & Stampe, 2001). Although a larger fixation range can in principle yield an increased serial span, eye fixations could not account for metrical similarity effects, which reflect events at alternating distances in the notation.

We compared the model's predictions with error patterns from adult musicians' memorized performances (Palmer & van de Sande, 1995) of excerpts from two-part *Inventions* by Bach. The musical sequences feature rhythmically varying (nonisochronous) lines of scalar melodies in a $\frac{4}{4}$ meter. Thus, the performances in Palmer and van de Sande (1995) provide a test of whether the range model extends to performance from memory as well as to more complex musical forms (both rhythmically and harmonically) than the musical exercises used in the experiments. Because the pianists performed the music at a consistent tempo (mean t = .177 s/IOI) that did not change much across performances, their data provide a good test for the range model. The unweighted range model was fit to the mean error proportions shown in the top of Figure 15, and the model was compared with three other models: serial, metrical, and additive models as well as chance estimates. Initial activation was optimized as before across the models. The range model provided the best fit to the memorized performances (AIC = −37.38, VAF = .95, $p < .01$), and initial activation was .900. This initial activation value is lower than that obtained in Experiment 1 (a = .960) with pianists of comparable skill levels, which may reflect the greater memory demands of performance in the absence of a visual stimulus. These findings also confirm that the model extends to performance of rhythmically and harmonically complex music.

We also tested the range model's applicability to other meters, particularly those that contain ternary patterns of alternating metrical accents. An example of a metrical grid for the ternary meter of $\frac{3}{8}$ is shown in the bottom of Figure 15. Ternary meters in Western tonal music usually contain a ternary alternation of strong and weak beats between one of the hierarchical levels (Level 3 in Figure 15) and the next lowest level (Level 2). Palmer and van de Sande (1995) reported movement errors from 18 adults' memorized performances of two *Inventions* in $\frac{3}{8}$ meter by Bach. The range model was fit to the mean data of the ternary meter performances in that study. The mean tempo was .190 s/IOI. The error proportions and the range model fits are shown at the bottom of Figure 15. Again, the range model provided a better fit (AIC = −30.16, VAF = .95, $p < .01$) than the serial component, the metrical component, an additive model, or the chance estimates shown in Figure 15, and the initial activation was a = .880, again a lower value in memorized performance than in performance with visual presentation. These findings confirm that the model extends to ternary meters as well as to memorized performance.

Next, we tested whether the model's initial activation parameter captures performer-specific memory constraints. The higher initial activation values obtained for adult than for child performers in the

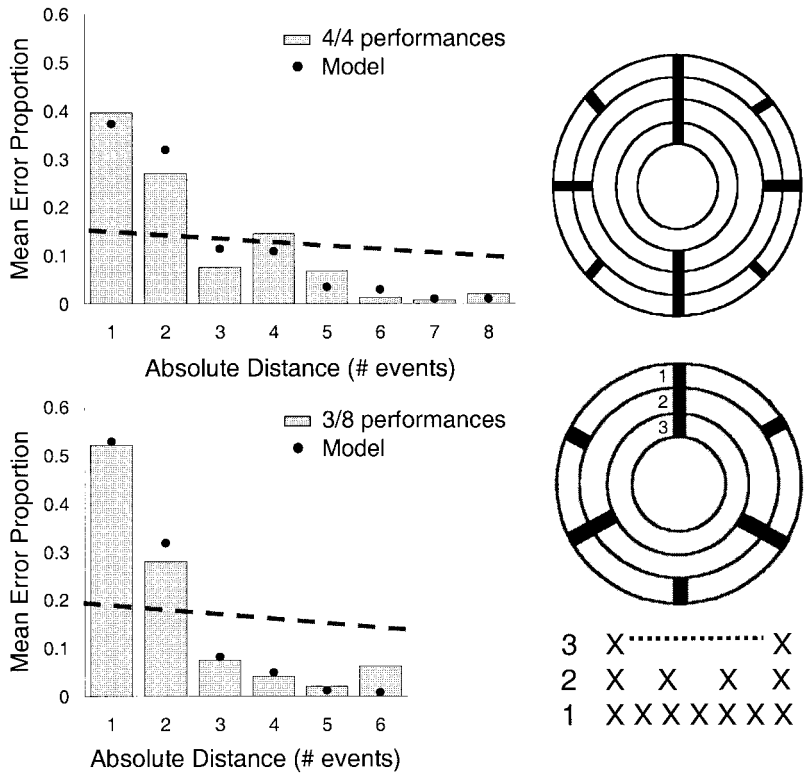

Figure 15. Left: Unweighted model fits of memorized performances from Palmer and van de Sande (1995). The dashed lines indicate chance estimates. Right: Cyclic depictions of metrical accent, in which Xs indicate an abstract representation of metrical accent strength for one measure plus one beat of the music. Top: Binary meter. Bottom: Ternary meter.

experiments may have reflected stimulus or task differences; the adults performed different musical sequences than the children to elicit equivalent error rates. Here, we applied the model to children's and adults' performances of the same musical sequence. Drake and Palmer (2000) compared child and adult pianists' performances of unfamiliar music; pianists at higher skill levels had a significantly larger range of planning, on the basis of error distances, than pianists at lower skill levels. We first compared the performances of 12 advanced children (aged 12–16 years, with 8.4 years mean piano instruction) and 12 adults (aged 20–30 years, with 15.3 years mean piano instruction) of the same musical sequence in ternary meter; the performance tempo did not change much within groups, and thus their data provide a good test for the range model. The fit of the range model was optimized on Drake and Palmer's (2000) mean range values across participants, on the basis of the mean performance tempo (adults: $t = .221$ s/IOI; children: $t = .296$ s/IOI). Fits were based on mean range values rather than movement error gradients because the error gradients were difficult to distinguish from chance estimates for the nonuniform distribution of events in the musical sequence. Initial activation of the best-fitting model was higher (although not significantly so) for adults (.957) than for children (.930) who performed the same musical sequence.

We also compared two groups of less advanced child performers from the same study (Drake & Palmer, 2000): 12 younger children (aged 8–11 years, with a mean of 4.7 years of piano instruction) and 12 older children (aged 12–15 years, with a mean of 6.6 years of piano instruction) who performed an easier musical sequence than the advanced performers. These two groups performed the same musical sequence in binary meter at the same rate (younger children: $t = .222$ s/IOI; older children: $t = .207$ s/IOI; $p > .10$), allowing a test of the initial activation parameter when tempo does not vary. The model fits were optimized on the mean range values as before. Initial activation estimates were significantly higher for the older children ($a = .890$) than for the younger children ($a = .864$; $p < .05$, 95% confidence interval). Furthermore, both groups had significantly lower initial activation values than the more advanced children and adults in the same study ($p < .05$, 95% confidence intervals). Thus, the initial activation parameter values increased monotonically with age and experience across the four comparison groups from Drake and Palmer (2000), suggesting that the model parameters remained stable across performers and stimuli. Moreover, these comparisons suggest that the initial activation differences obtained in the experiments reported here are more likely due to performer-specific factors than to stimulus-specific factors.

Finally, we tested further the validity of the model's metrical weight parameters. The metrical level chosen to be weighted was unconstrained in the model fits reported earlier. We compared a constrained weighted model with data from an experiment in which 18 skilled pianists performed the same musical piece (*Prelude* by Bach, 400 events in length) while emphasizing (weighting)

specific metrical levels (eighth-note, quarter-note, or half-note) as the primary metrical level (R. K. Meyer & Palmer, 2003, Experiment 2). Error proportions changed across metrical conditions, such that more errors arose from distances equivalent to the emphasized metrical level. Because pianists were familiar with the music, error rates were low (less than 2%) and only the quarter-note and half-note conditions (which pianists performed at a faster tempo) contained sufficient numbers of errors for modeling. Figure 16 shows the error distribution from the quarter-note (t = .132 s/IOI) and half-note (t = .116 s/IOI) metrical conditions. The range model was first fit to the mean error proportions with no metrical weights, by fitting initial activation across the metrical conditions as before; the initial activation was .962 (AIC = −21.40, VAF = .37, $p < .05$). Then, the constrained-weighted model was fit to the data with the same initial activation value, optimizing across weight values corresponding to the metrical level, l, that performers emphasized; weights at other levels were set equal to $(1 - w_l)/(k - 1)$, as in Equation 1. Thus, the weighted model has one more free parameter (for weight value) than the unweighted model.

The circles in Figure 16 indicate where the metrical weights increase the metrical similarity of events at distances of every four events (in the quarter-note condition) or eight events (in the half-note condition). Increasing the weight at a given metrical level in the model increases the metrical accent of events aligned with the weighted level, which increases the similarity between those events and decreases the similarity between events at the weighted level and other events. The fit of the model with metrical weights on the quarter-note level surpassed the unweighted model for the quarter-note condition (weighted model: AIC = −20.50, VAF = .44, $p < .05$; unweighted model: AIC = −20.00, VAF = .30, $p < .10$). The fits to the half-note condition with metrical weights added, also shown in Figure 16, improved but not significantly; both weighted and unweighted models provided a good fit. This is not surprising because the model's predictions for the half-note metrical level primarily affect error distances of 8, which are most dampened by the serial component and especially so at the fast production rates in this condition. Notably, models with weights on the wrong metrical levels (ones not intended by performers) provided a worse fit than the unweighted model for both half-note and quarter-note conditions, indicating that the improvement was not due simply to the inclusion of an increased weight on any level. Although performers' error rates were too low (less than 2%) to allow individual fits, this extension of the model provides some evidence for the validity of the metrical weights.

In summary, these extensions verify that the range model applies to a variety of musical sequences, to memorized and unmemorized performances, to different meters, to well-practiced and unpracticed (unfamiliar) music, and to different weighted metrical levels. Furthermore, the parameter estimates were fairly stable across these applications, supporting an interpretation of the model's parameters as representing psychological processes rather than stimulus or task demands. The extensions of metrical weights in particular are important because they enlarge the model's ability to account for performances of different musical sequences that are based on the same meter as well as different performances of the same musical sequence. The range model's metrical weights have analogous effects to the amplitude of an oscillator in a dynamical systems model. Dynamical systems models, based on internal oscillations that entrain to periodicities in external events, have

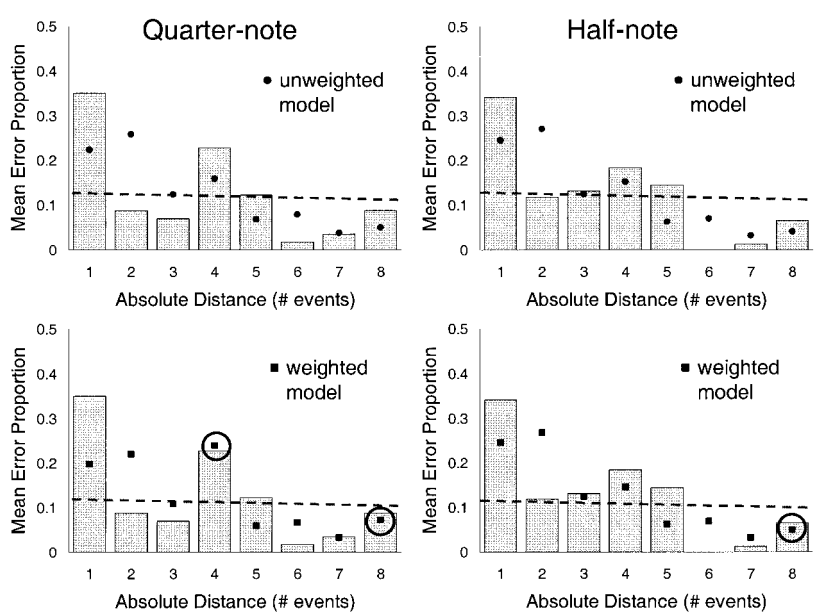

Figure 16. Model fits to performances in which quarter-note or half-note metrical levels are emphasized (from R. K. Meyer & Palmer, 2003, Experiment 2). Top row: Fits (circles) of the unweighted model. Bottom row: Fits (squares) with one metrical level weighted (w_3 in the quarter-note condition; w_4 in the half-note condition). The circled squares indicate where the metrical weights increase the metrical similarity of events at distances of every four events (in the quarter-note condition) or eight events (in the half-note condition). The dashed lines indicate chance estimates.

been applied to auditory sequences to predict expectancies for when events will occur (Large & Jones, 1999; Large & Kolen, 1994). Different oscillators with adjustable period and phase components respond to periodicities equivalent to each level in a metrical grid and are capable of adapting to the rate changes that characterize music performances (Large & Palmer, 2002). Thus, metrical frameworks may be instantiated in a process in which internal predispositions toward a range of periodicities entrain to structural characteristics of auditory sequences (e.g., Jones, 1976; Large & Jones, 1999).

Memory for Serial Order

In this section, we compare the range model with three serial-order models that are sensitive to temporal features. Although most serial-order models predict recall of a list after its presentation (outside of a production context), we can draw comparisons between the models in terms of their assumptions about the periodic nature of activation, temporal decay, and item similarity. The range model's incorporation of periodic activation that strengthens associations among elements is similar to aspects of the perturbation model (Estes, 1972, 1997) and the temporal distinctiveness model (G. D. A. Brown et al., 2000). The range model is also similar to the perturbation model in that the positional information of each item within a sequence is assumed to be stored and used at retrieval. Temporal decay of information in the perturbation model is a function of retention interval (time elapsed since the beginning of the list). In the temporal distinctiveness model (G. D. A. Brown et al., 2000), the probability of correctly recalling an item's serial order is reduced when its immediate temporal neighborhood is crowded, similar to the range model's temporal proximity predictions. The primary difference from both of these models is that decay is governed by the ratio of serial separation to temporal separation between events, a feature unique to the range model.

One model based on the adaptive character of thought—rational (ACT–R; Anderson, 1993), applied to serial memory, combines (in a different way) time-based decay with similarity-based interference (Anderson & Matessa, 1997). Serial lists are represented as hierarchical structures, and declarative knowledge units encode the position of items; production rules use this information to organize the serial recall. One important difference is that the ACT–R model computes item similarity in absolute terms, whereas the range model computes metrical similarity in relative terms. In addition, time in the ACT–R model is a scaling (constant) factor that modulates the activation, which is defined independently of the production rate. Because decay occurs in the range model relative to the number of items elapsed per time period, only the range model predicts the interactions of rate and sequence distance reported here.

Planning Constraints in Language and Music

How specific are the range of planning constraints described here to music performance? Of the range model's parameters, production rate is established by the experimental conditions; initial activation, which modifies exponential decay, is posited to increase with age and experience. The parameters associated with metrical structure appear more specific to music; however, the periodicities represented in the metrical component resemble in fundamental ways those proposed for language (Hayes, 1984; Liberman & Prince, 1977). Linguistic rhythm is defined in metrical phonology as arising from the relative prominence of linguistic units (e.g., syllables) aligned with the hierarchical periodicities in a metrical grid (Liberman & Prince, 1977). Metrical grids capture patterns of relative prominence among syllables: the regular recurrence of strong and weak beats, defined by the periodicity at each level in the metrical grid (Prince, 1989). Many theoretical comparisons have been drawn between metrical structure in music and language (Cooper & Meyer, 1960; Hayes, 1989; Kiparsky, 1975; Lerdahl & Jackendoff, 1983; Prince, 1983; Selkirk, 1984). Metrical rules of prominence have been viewed as "a bridge between the rhythm of music and speech" (Kiparsky, 1975, p. 583).

Theoretical applications of metrical grid theory in the domains of music and speech indicate strikingly similar assumptions. First, prominence is defined in relative (not absolute) terms, following from the relative placement of syllables within the meter (Hayes, 1988; Liberman & Prince, 1977), similar to definitions of metrical similarity in the range model and in music–theoretic applications (Cooper & Meyer, 1960; Lerdahl & Jackendoff, 1983). Second, the linguistic grids are defined on an ordinal scale: They are not tied to any particular (interval- or ratio-scale) realization of isochronous intervals (Hayes & MacEachern, 1998; Prince, 1989), as implemented in the range model. Third, a preference for binary alternation is acknowledged in both domains (Hammond, 1995; Liberman, 1975; Liberman & Prince, 1977; Palmer & Kelly, 1992; Prince, 1989; K. C. Smith & Cuddy, 1989), although metrical grids accommodate ternary alternation as well as mixtures of binary and ternary alternation, as seen in Finnish mixed meter (Hanson & Kiparsky, 1996) and in complex musical meters (London, 1995), like Paul Desmond's song "Take 5" (Desmond, 1960). Finally, metrical prominence rules in language are similar to the rules proposed to guide perception of musical meter. For example, applications of optimality theory to metrical prominence posit a prioritized hierarchy of metrical constraints such as avoid empty (unfilled) strong metrical positions (Hayes & MacEachern, 1998); this same rule underlies perceptual models of musical meter (cf. Longuet-Higgins & Lee, 1984; Povel & Essens, 1985). Thus, the periodic hierarchies represented in metrical grids may reflect general cognitive constraints of rhythmic competence (Hayes, 1984, 1988) that include preferred periodic relationships found in music, speech, and memory for serial order.

The notion that general periodic functions underlie serial-order effects in music and speech is also supported by Vousden et al.'s (2000) model of phonological encoding. Oscillators, modeled as sinusoidal signals, operate at different periodicities to encode the learning context of a string of syllables and phonemes. At retrieval, the reinstatement of the original states of the oscillators activates the segmental information (phonemes) with which the phonological context was associated at encoding. A novel metrical structure is assembled for each sequence, which involves linking the phoneme sequence to successive states of the oscillators. The model's oscillator components capture metrical structure in a way similar to the periodic metrical accents assumed in the range model. Furthermore, both approaches assume that contextual information about positional relationships between the current event and other sequence events is available at retrieval.

Two predictions of Vousden et al.'s (2000) model differ from those of the range model. The first is a shallower movement gradient (larger range) predicted for perseveratory errors than for anticipatory errors because of postoutput suppression of the previously produced event in the model; Vousden et al. (2000) found some supporting evidence in speech errors. In contrast, the experiments reported here indicated no differences among range measures for anticipatory and perseveratory errors, consistent with the range model's predictions. Postoutput suppression may be less useful in production of musical sequences, which contain many repeating events, because the suppression of a previously produced event may impede the activation of the same event at an upcoming position. The second difference is a shallower movement gradient (larger mean range) predicted by Vousden et al. (2000) for children's than for adults' errors, attributed to less temporal distinctiveness in children's representations; McCormack, Brown, Vousden, and Henson (2000) reported supporting evidence from a written-recall task. Movement gradients for music performance errors, in contrast, tend to show a larger scope for adults than for children (Drake & Palmer, 2000; Palmer & Drake, 1997), consistent with the initial activation predictions of the range model. Production rate may contribute to the disparity; speech error corpora (Vousden et al., 2000) and written-recall tasks (McCormack et al., 2000) often reflect a variety of production rates, with slower rates for children. Rate and age-related effects, which interact in the range model, may also covary in verbal tasks and require further evaluation.

Combining Range and Directional Constraints on Planning

Here, we consider how range constraints combine with directional (anticipatory) constraints in music and other sequence-production domains. Performers' range of planning and anticipatory tendencies increased as overall error rate decreased in skilled performance. Furthermore, the proportion of anticipatory errors increased with practice in the music performances. These findings support Dell et al.'s (1997) model in which practice affects the weighting of the future relative to the past and the notion that anticipatory planning is a domain-general constraint; the findings have been replicated in other music performance tasks as well (Drake & Palmer, 2000). However, the current experiments failed to replicate the anticipatory rate effect predicted by the same model (Dell et al., 1997): The anticipatory proportion of errors was no higher at a medium production rate than at a fast rate. One explanation is that anticipatory activation, which influences the anticipatory rate effect, does not remain constant across production rates as assumed by Dell et al. (1997). Anticipatory activation can reflect structural factors such as the size of phrasal units or the location of heads in phrases. Faster rates in music performance can coincide with an increase in the size of units such as phrases (Clarke, 1982; R. K. Meyer & Palmer, 2003); if unit size varies with production rate and anticipatory activation is sensitive to unit size, then anticipatory activation may also vary with rate. Thus, a failure to replicate the anticipatory rate effect in music may result because performers adapt to faster production rates by anticipating more. Musical interpretation may allow more flexibility than speech in the size of structural units, as indicated in experimental studies that compare different performances of the same music (Clarke, 1982; Palmer, 1989; Palmer & van de Sande, 1993, 1995).

Our findings suggest that range and anticipatory constraints are distinct planning processes in music that arise from different underlying variables. Practice influenced anticipatory behavior and rate influenced planning range in music performance. Our theoretical framework assumes that past and future events are equally active on average, and thus the range model does not predict an overall anticipatory tendency. Furthermore, the range model's predictions are insensitive to short-term practice; the metrical frames' binary and ternary periodic components do not change much across frames or across sequences (the metrical weights exaggerate but do not distort the underlying hierarchy). In contrast, anticipatory tendencies arise in Dell et al.'s (1997) model from directional weighting of serially ordered items within linguistic frames that are not periodic and vary with the complexity and form of each sequence. Anticipatory practice effects in music may likewise reflect sensitivity to harmonic and tonal relationships that are not as periodic as meter. A wide variety of nonperiodic frames underlies speakers' and musicians' competence; short-term learning (including practice effects) of irregular changes in sequence structure may be more important for directional (anticipatory) than for range constraints on production. As to production rate effects, we speculate that rate produces comparable effects on range measures in other serial behaviors with temporal structure.

How anticipatory and range constraints combine depends in part on the length of produced sequences, according to the range model's predictions. Periodic functions that modulate retrieval, like the metrical similarity function in the range model, are susceptible to sequence length; larger periodicities can have an impact only in longer sequences. Directional biases can also be affected by sequence length; if errors or other measures of anticipatory tendencies are not uniformly distributed across sequence positions, then their directionality can be exaggerated in shorter sequences. For example, the range model's predictions on a small scale of the sequence (consider four events on either side of the current event) reveal activation biases toward immediate future events that differ across production rates (see Figure 7, right-hand column). Thus, distance and directional effects can be correlated for short sequences, such as those used in tongue twister studies or serial-recall tasks (e.g., G. D. A. Brown et al., 1999; Dell et al., 1997; Hulme et al., 1984). The longer the sequence, the more locations across which the range model's activations are averaged; eventually, the anticipatory proportion approaches the expected value (50%), and directional and distance effects become uncorrelated in longer sequence segments. In summary, longer sequences can increase the likelihood of obtaining uniform distributions of errors across sequence positions, which influences evaluations of directional and scope constraints in sequence production.

Finally, we consider the limitations of the range model. First, the formal model is not implemented here as a process model. We discussed one possibility, oscillator mechanisms that generate the periodic functions of the metrical component in the model. Another implementation feature is raised by the second limitation, that the model does not specify error rates on an absolute scale. The evidence reported here suggests that differential event activations at each sequence position are modulated by metrical accent strength. Combined with a threshold criterion that regulates response selection, the activation values can specify absolute error

rates. These predictions, which incorporate production rate, are currently being extended to assess speed–accuracy trade-off relationships. Finally, other musical dimensions such as pitch may influence similarity among sequence events; these dimensions might be incorporated as features in the model's item vector representations, similar to the representation of metrical similarity. The model presented here constitutes the simplest possible implementation of the core assumptions of our theoretical framework. This raises the possibility that the model's limited parameter space cannot account for a wide set of behaviors. We offered reasons why similar metrical constraints on retrieval may hold in domains like language and in other temporally patterned behaviors. We also reported applications of the model to a range of studies with different musical sequences, participants, and procedures. The fact that the parameter estimates were relatively stable across these applications suggests that the model fits a variety of data for reasons inherent in its theoretical assumptions, as opposed to experimental artifacts. Thus, the simple model (with only three free parameters in its most elaborate form) is a first step in uncharted territory of specifying the extent of incremental planning in complex production tasks.

Study of retrieval in music performance offers three practical implications for theories of memory. The first is the role that repeating items play in theories of serial recall; we began with the claim that music offers a difficult challenge because it contains many repeating items that occupy different serial orders within the same sequence. To differentiate the serial ordering of each item in the different contexts in which it appears, we proposed a contextually determined retrieval mechanism. Thus, the presence or absence of item repetition may influence accounts of memory processes that underlie sequence production. A second implication is the role that sequence length plays in measures of serial recall. Although short sequences allow more experimental control, they do not always permit the separation of different planning constraints such as anticipatory and range biases or the measurement of large periodic influences on recall. A final implication is the role that auditory nonverbal information plays in studies of memory. Studies of serial recall tend to rely on verbal material for which acoustic confusions are commonly found, lending support to phonological loop or articulatory buffer accounts of memory. Memory for music also involves acoustic confusions (such as diatonic or harmonic confusions) that are not phonological. Nonetheless, memory confusions in music show similar patterns to those found in verbal domains, such as the anticipatory practice effect or age effects on memory constraints described here. These implications advocate a general view that music and other natural complex domains of memory deserve more critical examination to promote theories that apply to a broad range of serial behaviors.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrox & F. Caski (Eds.), *Second International Symposium on Information Theory* (pp. 267–281). Budapest, Hungary: Akademiai Kiado.

Anderson, J. R. (1993). *Rules of the mind.* Hillsdale, NJ: Erlbaum.

Anderson, J. R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review, 104,* 728–748.

Baddeley, A. D. (1986). *Working memory*. London: Oxford University Press.

Barrouillet, P., & Camos, V. (2001). Developmental increase in working memory span: Resource sharing or temporal decay? *Journal of Memory and Language, 45,* 1–20.

Bernstein, N. (1967). *The coordination and regulation of movements*. London: Pergamon.

Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception, 5,* 1–30.

Bharucha, J. J. (1994). Tonality and expectation. In R. Aiello (Ed.), *Musical perceptions* (pp. 213–239). New York: Oxford University Press.

Bjork, E. L., & Healy, A. F. (1974). Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior, 13,* 80–97.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89,* 1–47.

Bock, K. (1995). Sentence production: From mind to mouth. In J. L. Miller & P. D. Eimas (Eds.), *Handbook of perception and cognition: Speech, language, and communication* (Vol. 11, pp. 181–216). Orlando, FL: Academic Press.

Bower, G. H. (1971). Adaptation-level coding of stimuli and serial position effects. In M. H. Appley (Ed.), *Adaptation-level theory* (pp. 175–201). New York: Academic Press.

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Brown, J. C. (1992). Determination of musical meter using the method of autocorrelation. *Journal of the Acoustical Society of America, 91,* 2374–2375.

Brown, G. D. A., & Hulme, C. (1995). Modeling item length effects in memory span: No rehearsal needed? *Journal of Memory and Language, 34,* 594–621.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review, 107,* 127–181.

Brown, G. D. A., Vousden, J. I., McCormack, T., & Hulme, C. (1999). The development of memory for serial order: A temporal-contextual distinctiveness model. *International Journal of Psychology, 34,* 389–402.

Chaffin, R., & Imreh, G. (1997). Pulling teeth and torture: Musical memory and problem solving. *Thinking and Reasoning, 3,* 315–336.

Clancy, S. M., & Hoyer, W. J. (1994). Age and skill in visual search. *Developmental Psychology, 30,* 545–552.

Clarke, E. F. (1982). Timing in the performance of Erik Satie's "Vexations." *Acta Psychologia, 30,* 1–19.

Clarke, E. F. (1987). Categorical rhythm perception: An ecological perspective. In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 19–34). Stockholm: Royal Swedish Academy of Music.

Clarke, E. F. (1999). Rhythm and timing in music. In D. Deutsch (Ed.), *The psychology of music* (Vol. 2, pp. 473–500). San Diego, CA: Academic Press.

Collard, R., & Povel, D.-J. (1982). Theory of serial pattern production: Tree traversals. *Psychological Review, 89,* 693–707.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior, 4,* 161–169.

Cooper, G., & Meyer, L. B. (1960). *The rhythmic structure of music*. Chicago: University of Chicago Press.

Cowan, N. (2000). The magical number 4 in short-term memory: Reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24,* 87–185.

Crowder, R. G., & Pitt, M. (1992). Research on memory/imagery for musical timbre. In D. Reisberg (Ed.), *Auditory imagery* (pp. 29–44). Hillsdale, NJ: Erlbaum.

Cuddy, L. L., Cohen, A. J., & Miller, J. (1979). Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology, 33,* 148–157.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from

the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 222–233.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language production and serial order: A functional analysis and a model. *Psychological Review, 104,* 123–147.

Desain, P., & Honing, H. (1989). Quantization of musical time: A connectionist approach. *Computer Music Journal, 13,* 56–66.

Desain, P., & Honing, H. (2003). The formation of rhythmic categories and metric priming. *Perception, 32,* 341–365.

Desmond, P. (1960). Take 5 [Recorded by the Dave Brubeck Quartet]. On *Time out* [CD]. New York: Columbia Records.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88,* 503–522.

Dowling, W. J., & Bartlett, J. C. (1981). The importance of interval information in long-term memory for melodies. *Psychomusicology, 1,* 30–49.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49,* 524–531.

Drake, C., Jones, M. R., & Baruch, C. (2000). The development of rhythmic attending in auditory sequences: Attunement, referent period, and focal attending. *Cognition, 77,* 251–288.

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition, 74,* 1–32.

Duke, R. A. (1989). Musicians' perception of beat in monotonic stimuli. *Journal of Research in Music Education, 37,* 61–71.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

Estes, W. K. (1972). An associative basis for coding and organization in memory. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 161–190). Washington, DC: Winston.

Estes, W. K. (1997). Processes of memory loss, recovery, and distortion. *Psychological Review, 104,* 148–169.

Ferreira, V. S. (1996). Is it better to give than to donate? Syntactic flexibility in language production. *Journal of Memory and Language, 35,* 724–755.

Finney, S. A., & Palmer, C. (2003). Auditory feedback and memory for music performance: Sound evidence for an encoding effect. *Memory & Cognition, 30,* 51–64.

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology, 47,* 381–391.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology, 35,* 1–36.

Gabrielsson, A. (1987). Once again: The theme from Mozart's piano Sonata in A Major (k. 331). In A. Gabrielsson (Ed.), *Action and perception in rhythm and music* (pp. 81–104). Stockholm: Royal Swedish Academy of Music.

Gabrielsson, A. (1999). Music performance. In D. Deutsch (Ed.), *The psychology of music* (Vol. 2, pp. 501–602). New York: Academic Press.

Garcia-Albea, J. E., del Viso, S., & Igoa, J. M. (1989). Movement errors and levels of processing in sentence production. *Journal of Psycholinguistic Research, 18,* 145–161.

Garner, W. G. (1962). *Uncertainty and structure as psychological concepts.* London: Wiley.

Garrett, M. F. (1980). Levels of processing in sentence production. In B. Butterworth (Ed.), *Language production: Speech and talk* (pp. 177–220). London: Academic Press.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3,* 410–418.

Gathercole, S. E., & Adams, A. M. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language, 33,* 672–688.

Gomes, H., Sussman, E., Ritter, W., Kurtzberg, D., Cowan, N., & Vaughan, H. G. (1999). Electrophysiological evidence of developmental changes in the duration of auditory sensory memory. *Developmental Psychology, 35,* 294–302.

Grondin, S. (2001). From physical time to the first and second moments of psychological time. *Psychological Bulletin, 127,* 22–44.

Gupta, P., & MacWhinney, B. (1997). Vocabulary acquisition and verbal short-term memory: Computational and neural bases. *Brain and Language, 59,* 267–333.

Hammond, M. (1995). Metrical phonology. *Annual Review of Anthropology, 24,* 313–342.

Hanson, K., & Kiparsky, P. (1996). A parametric theory of poetic meter. *Language, 72,* 287–335.

Hayes, B. (1984). The phonology of rhythm in English. *Linguistic Inquiry, 15,* 33–74.

Hayes, B. (1988). Metrics and phonological theory. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge Survey* (Vol. 2, pp. 220–249). Cambridge, England: Cambridge University Press.

Hayes, B. (1989). The prosodic hierarchy in meter. In P. Kiparsky & G. Youmans (Eds.), *Phonetics and phonology: Rhythm and meter* (pp. 201–260). San Diego, CA: Academic Press.

Hayes, B., & MacEachern, M. (1998). Quatrain form in English folk verse. *Language, 74,* 473–507.

Healy, A. F. (1974). Separating item from order information in short-term memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 644–655.

Hulme, C., Thomson, N., Muir, C., & Lawrence, A. (1984). Speech rate and the development of short-term memory span. *Journal of Experimental Child Psychology, 38,* 241–253.

Huron, D. (1993). Note-onset asynchrony in J. S. Bach's two-part *Inventions. Music Perception, 10,* 435–444.

Jones, M. R. (1974). Cognitive representations of serial patterns. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 187–229). Hillsdale, NJ: Erlbaum.

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

Jones, M. R., & Pfordresher, P. Q. (1997). Tracking musical patterns using joint accent structure. *Canadian Journal of Experimental Psychology, 51,* 271–290.

Kempen, G., & Hoenkamp, E. (1987). An incremental procedural grammar for sentence formulation. *Cognitive Science, 11,* 201–258.

Kiparsky, P. (1975). Stress, syntax, and meter. *Language, 3,* 576–615.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch.* New York: Oxford University Press.

Large, E. W. (1993). Dynamic programming for the analysis of serial behaviors. *Behavior Research Methods, Instruments, & Computers, 25,* 238–241.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106,* 119–159.

Large, E. W., & Kolen, J. F. (1994). Resonance and the perception of musical meter. *Connection Science, 6,* 177–208.

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science, 26,* 1–37.

Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior* (pp. 112–136). New York: Wiley.

Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory, 7,* 149–169.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Levelt, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition, 14,* 41–104.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review, 96,* 25–57.

Liberman, M. (1975). *The intonational system of English*. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry, 8,* 249–336.

London, J. (1995). Some examples of complex meters and their implications for models of metric perception. *Music Perception, 13,* 59–72.

Longuet-Higgins, H. C., & Lee, C. S. (1984). The rhythmic interpretation of monophonic music. *Music Perception, 1,* 424–441.

MacKay, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia, 8,* 323–350.

MacKay, D. G. (1987). *The organization of perception and action: A theory for language and other cognitive skills*. New York: Springer-Verlag.

MacKenzie, C. L., & Van Eerd, D. L. (1990). Rhythmic precision in the performance of piano scales: Motor psychophysics and motor programming. In M. Jeannerod (Ed.), *Attention and performance XIII* (pp. 375–408). Hillsdale, NJ: Erlbaum.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review, 79,* 487–509.

McCormack, T., Brown, G. D. A., Vousden, J. I., & Henson, R. N. A. (2000). Children's serial recall errors: Implications for theories of short-term memory development. *Journal of Experimental Child Psychology, 76,* 222–252.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture–word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1146–1160.

Meyer, R. K., & Palmer, C. (2003). *Rate and tactus effects on timing and serial order in music performance*. Manuscript submitted for publication.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63,* 81–97.

Murdock, B. B. (1960). The distinctiveness of stimuli. *Psychological Review, 67,* 16–31.

Myung, I. J. (2000). The importance of complexity in model selection. *Journal of Mathematical Psychology, 44,* 190–204.

Myung, I. J., & Pitt, M. A. (1998). Issues in selecting mathematical models of cognition. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 327–355). Mahwah, NJ: Erlbaum.

Nairne, J. S. (1991). Positional uncertainty in long-term memory. *Memory & Cognition, 19,* 332–340.

Nairne, J. S. (2002). Remembering over the short-term: The case against the standard model. *Annual Review of Psychology, 53,* 53–81.

Neath, I. (1993). Contextual and distinctiveness processes and the serial position function. *Journal of Memory and Language, 32,* 820–840.

Neath, I., & Nairne, J. S. (1995). Word length effects in immediate memory: Overwriting trace decay theory. *Psychonomic Bulletin & Review, 2,* 429–441.

Palmer, C. (1989). Mapping musical thought to musical performance. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 331–346.

Palmer, C. (1996). Anatomy of a performance: Sources of musical expression. *Music Perception, 13,* 433–454.

Palmer, C. (1997). Music performance. *Annual Review of Psychology, 48,* 115–138.

Palmer, C., & Drake, C. (1997). Monitoring and planning capacities in the acquisition of music performance skills. *Canadian Journal of Experimental Psychology, 51,* 369–384.

Palmer, C., & Kelly, M. H. (1992). Linguistic prosody and musical meter in song. *Journal of Memory and Language, 31,* 525–542.

Palmer, C., & Krumhansl, C. L. (1990). Mental representations for musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 728–741.

Palmer, C., & Meyer, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science, 11,* 63–68.

Palmer, C., & van de Sande, C. (1993). Units of knowledge in music performance. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 457–470.

Palmer, C., & van de Sande, C. (1995). Range of planning in skilled music performance. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 947–962.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception, 11,* 409–464.

Piston, W. (1987). *Harmony* (5th ed.). New York: Norton.

Povel, D. (1981). Internal representation of simple temporal patterns. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 3–18.

Povel, D., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2,* 411–440.

Povel, D., & Okkerman, H. (1981). Accents in equitone sequences. *Perception & Psychophysics, 30,* 565–572.

Prince, A. (1983). Relating to the metrical grid. *Linguistic Inquiry, 14,* 19–100.

Prince, A. (1989). Metrical forms. In P. Kiparsky & G. Youmans (Eds.), *Phonetics and phonology: Rhythm and meter* (Vol. 1, pp. 45–80). San Diego, CA: Academic Press.

Reingold, E. M., Charness, N., Pomplun, M., & Stampe, D. M. (2001). Visual span in expert chess players: Evidence from eye movements. *Psychological Science, 12,* 48–55.

Repp, B. H. (1992). Probing the cognitive representation of musical time: Structural constraints on the perception of timing perturbations. *Cognition, 44,* 241–281.

Restle, F. (1972). Serial patterns: The role of phrasing. *Journal of Experimental Psychology, 92,* 385–390.

Restle, F., & Brown, E. (1970). Organization of serial pattern learning. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 4, pp. 249–331). New York: Academic Press.

Rumelhart, D. E., & Norman, D. A. (1982). Simulating a skilled typist: A study of skilled cognitive-motor performance. *Cognitive Science, 6,* 1–36.

Seashore, C. E. (1938). *Psychology of music*. New York: McGraw-Hill.

Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.

Simon, H. A., & Sumner, R. K. (1968). Pattern in music. In B. Kleinmuntz (Ed.), *Formal representation of human judgment* (pp. 219–250). New York: Wiley.

Sloboda, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology, 35,* 377–396.

Smith, K. C., & Cuddy, L. L. (1989). Effects of metric and harmonic rhythm on the detection of pitch alterations in melodic sequences. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 457–471.

Smith, M., & Wheeldon, L. (1999). High level processing scope in spoken sentence production. *Cognition, 73,* 205–246.

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica, 45,* 175–197.

Sternberg, S., Knoll, R. L., & Zukovsky, P. (1982). Timing by skilled

musicians. In D. Deutsch (Ed.), *Psychology of music* (Vol. 1, pp. 181–239). New York: Academic Press.

Tillman, B., Bharucha, J. J., & Bigand, E. (2000). Implicit learning of tonality: A self-organizing approach. *Psychological Review, 107,* 885–913.

Towse, J. N., & Hitch, G. J. (1995). Is there a relationship between task demand and storage space in tests of working memory capacity? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 48*(A), 108–124.

Van Noorden, L. P. A. S. (1975). *Temporal coherence in the perception of tone sequences*. Unpublished doctoral dissertation, Eindhoven University of Technology, Eindhoven, the Netherlands.

Vousden, J. I., Brown, G. D. A., & Harley, T. A. (2000). Serial control of phonology in speech production: A hierarchical model. *Cognitive Psychology, 41,* 101–175.

Walraven, J., Enroth-Cugell, C., Hood, D. C., MacLeod, D. I. A., & Schnapf, J. L. (1990). The control of visual sensitivity: Receptoral and postreceptoral processes. In L. Spillman & J. S. Werner (Eds.), *Visual perception: The neurophysiological foundations* (pp. 53–101). San Diego, CA: Academic Press.

Wheeldon, L. R., Meyer, A. S., & Smith, M. (2003). Incremental language production. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (Vol. 4, pp. 760–764). London: Nature Publishing Group.

Yost, W. A. (1994). *Fundamentals of hearing: An introduction* (3rd ed.). San Diego, CA: Academic Press.

Appendix A

Music Corpus for Frequency Distribution Analysis

European Vocal Motets (1400s, for 4–6 Voices)

Osthoff, H. (Ed.). (1960). *Drei Motetten* (Vol. 57). Wolfenbuttel, Germany: Moseler.

"O Virgo Virginum" (6 voices, entire motet); "De Profundis Clamavi" (5 parts, Measures [mm.] 1–37).

Osthoff, H. (Ed.). (1957). *Zwei Psalmen* (Vol. 64).Wolfenbuttel, Germany: Moseler.

"Domine Dominus Noster" (5 voices, entire motet); "Domine, Exaudi Orationem Meam" (4 voices; first part, mm. 1–36); "Secunda Pars" (mm. 100–132); "Tertia Pars" (mm. 258–294).

Chinese Tone Music (1400s, for Eight-Stringed Zither)

Yung, B. (Ed.). (1997). *Celestial airs of antiquity: Music of the seven-string zither of China*. Madison, WI: A–R Editions.

"Huaxuyin (Song of Huaxu)," "Yangchun (Sunny Spring)" (mm. 1–20, 51–69). "Liushui (Flowing Waters)" (mm. 64–82); "Xiaohujia (Little Barbarian Horn)" (mm. 69–101, 217–242); "Guangingsan (Niezheng Assassinates the King)" (mm. 30–69, 170–210, 335–372).

Liang, T. (1971). *Music of Cheng*. Taipei, Taiwan: Chinese Classical Music Association.

"Buddhist Prayer Song," "Hundred Birds Courting the Phoenix," "Wild Geese Alighting on the Sandy Shore," "Relieving My Heart," "Flowers on the Brocade," and "Winter Ravens Sporting Over the Water."

American Swing Jazz (1930s, Transcriptions of Count Basie's Ensemble)

Hathaway, C. (Ed.). (1940). *Count Basie's piano styles*. New York: Bregman, Vocco, and Conn. Transcriptions from 1930s recordings: "Jumpin' at the Woodside" (mm. 1–24).

"Every Tub" (mm. 1–24); "Sent for You Yesterday" (mm. 1–24); "Swingin' the Blues" (mm. 1–24); "Blue and Sentimental" (mm. 1–24); "Good Morning Blues" (mm. 1–24); "Panassie Stomp" (mm. 1–24); "Shorty George" (mm. 1–24); "Out the Window" (mm. 1–24).

Children's Nursery Songs (All Pieces in Binary Meter)

Palmer, W. B. (Ed.). (1931). *American songs for children*. New York: Macmillan.

Bacon, D. (Arr.). (1993). *Twenty-four favorite nursery and folk songs.* Wellesley, MA: Kodaly Center of America.

Chroman, E. (Ed.). (1970). *Songs that children sing.* London: Music Sales Limited.

Mitchell, D. (Ed.). (1968). *Every child's book of nursery songs*. New York: Crown Publishers.

50 nursery songs and games. (circa 1940). Philadelphia: Theodore Presser.

Sharon, Lois, and Bram's Mother Goose. (1985). Boston: Atlantic Monthly Press.

Winn, M., & Miller, A. (Ed.). (1974). *The fireside book of fun and game songs.* New York: Simon & Schuster.

Winn, M. (Ed.), & Miller, A. (Arr.). (1966). *The fireside book of children's songs.* New York: Simon & Schuster.

Appendix B

Method and Error Coding for Experiment 1

Participants

Fourteen adult pianists from the Columbus, Ohio, music community (mean age = 24.3 years) participated. Participants had an average of 15.3 years of private instruction (range = 8–22 years).

Materials

The stimuli were four short musical phrases from *Czerny, Opus 299* (technical exercises for piano), adapted to be of equal length (2 measures) and time signature ($\frac{4}{4}$), and all contained 32 sixteenth-note events. The stimuli were repetitive in melodic content and were designed to be technically difficult. Similar to phonemic repetitions in tongue twisters, the repeating melodic sequences established pitch/duration patterns that were violated in the composition. The goal was to create phrases that were novel but in a familiar musical style and short enough to be learned easily. Two of the stimuli had sixteenth notes in both left and right hands and two had sixteenth notes in the right hand and quarter notes in the left hand. There were no successive repeating pitches in the stimuli.

Apparatus

Participants performed on a computer-monitored Yamaha Disklavier acoustic upright piano, recorded with 2-ms resolution. Keypresses on the acoustic piano were recorded with optical sensors that do not affect the touch or sound of the instrument. Software (Large, 1993) detected pitch errors by comparing each performance with the contents of a musical score.

Design and Procedure

The design and procedure were based on Experiment 3 by Dell et al. (1997). There were eight trial blocks, during each of which participants performed each of the four stimuli twice. Half of the phrases in each block were produced at a medium rate (187 ms/IOI), and half were produced at a faster rate (150 ms/IOI). The participants were assigned randomly to one of two groups; one group played the first and second stimuli at a medium tempo and the third and fourth stimuli at a fast tempo, and the second group had the reverse assignment. Thus, production rate was a within-subject factor; each trial block involved practice at a fast rate for some sequences and a slow rate for others. Stimulus presentation order was manipulated in a Latin square design across trial blocks, and production rate was manipulated within subjects. This arrangement formed a repeated measures design with 2 (performances) × 8 (blocks) × 2 (tempi), with two stimuli nested within each tempo. Assignment of stimuli to tempo condition was the only between-subjects factor. Thus, all participants received the same stimulus order, with tempo counterbalanced across stimuli.

At the beginning of an experimental session, each participant completed a musical background questionnaire. Then, the participant was presented with a notated musical sequence and was first asked to perform at a slow tempo of sixteenth notes = 535 ms/IOI as presented on a metronome, until an error-free performance was obtained. The performance at the slow rate was designed to ensure that there were no reading or perceptual errors. Then, the metronome was altered for the fast and medium tempo conditions, and participants were instructed to perform the same stimulus twice at this faster rate, with a pause between repetitions. This procedure was repeated for all four stimuli to complete the first block. The second through eighth blocks were the same as the first, with the exception that participants were not asked to perform the stimuli at the slower rate. In the ninth block, participants performed each stimulus once more at the slow tempo (535 ms/IOI) to ensure that the stimuli were not relearned incorrectly during the experiment. Participants were instructed to stop and start over if they made an error in the slow tempo performances but were instructed to not correct errors while performing at the medium or fast tempi. The notation remained in front of the participants throughout the experiment, and the participants were instructed to maintain the metronomic tempo, which continued during each trial.

Error Coding

Most errors involved an intruder (unintended event) that interacted with the target (intended event, as determined by the musical notation). (The assumption that notated events were intended is based on the participants' error-free performances at the slower tempo.) The intruder can reflect a source that is contextual (from the musical stimulus) or noncontextual. Contextual errors, also called movement errors, were coded for direction and distance from the source: The direction of the error was either anticipatory (an event produced before it was intended to occur) or perseveratory (an event produced after it was intended to occur). Distance was measured in terms of the number of intervening events (sixteenth notes) between the error and its source (nearest same pitch) in the musical score. Statistical analyses were conducted on log error rates (log of number of errors relative to number of error opportunities) combined across the two repetitions of the same stimulus for all errors with the exception of corrections (interrupted errors, in which an event was performed incorrectly and then corrected), which were excluded because of ambiguities in their coding. Error rates were calculated separately for single notes and chords (which occurred with different frequencies in the musical pieces) and then combined for the overall error rate per performance (see Drake & Palmer, 2000, for further error coding information). Error types were similar across single note and chord errors.

Because musical sequences often contain pitches that reoccur, errors that occur by chance (i.e., nonmovement errors) may resemble sequence elements and be mistaken for movement errors. A chance estimate was computed of how often pitches repeated in each stimulus. The absolute distance between each pair of repeating pitches within each stimulus provided a mean chance estimate of eight events (one cycle of the metrical grid); only movement errors whose presumed source was within this chance estimate were included in error distance analyses.

(*Appendixes continue*)

Appendix C

Method and Error Coding for Experiment 2

Participants

Twelve child pianists from the Columbus, Ohio, music community (mean age = 12.7 years, range = 9–16 years) recruited through local music teachers' associations participated. All participants had at least 3 years of private piano lessons (mean years of instruction = 6.4, range = 3–13 years).

Materials and Equipment

The tunes from Experiment 1 were simplified for novice pianists. The sixteenth-note stimuli from Experiment 1 were reduced to eighth-note versions, the tunes were transposed to key signatures that contained few sharps or flats (C Major, F Major, and G Major), and initial finger positions were indicated in the notation to simplify the task. The four stimuli were all of equal length (2 measures) and time signature ($\frac{4}{4}$), and all contained 16 eighth-note events. Participants performed on the same computer-monitored piano as in Experiment 1.

Design and Procedure

The design and procedure were the same as in Experiment 1, with the following adjustments. To reduce the total duration of the experiment, the children performed each stimulus only once in each block. To make the task easier, we slowed the production rates; the fast tempo was 300 ms/eighth-note IOI, and the medium tempo was 375 ms/IOI. The slower rate used in the first and ninth practice blocks was 900 ms/IOI, and the stimulus was performed repeatedly at this rate until an error-free performance was obtained. The participants were reminded after the first performance at the faster rate to try to maintain the rate indicated on the metronome; it was necessary to add this instruction for the children, who had less experience playing with a metronome than adults. At the end of the experiment, the child and the parent completed a questionnaire about the child's musical background. The children received a small gift for their participation, and the parents were paid a small amount to defray travel expenses. The entire experiment lasted about an hour.

Error Coding

Errors were coded as in Experiment 1. The mean chance estimate of distance between any pitch and its nearest reoccurrence in the notated score was eight events (equivalent to Experiment 1), and only movement errors for which a source could be identified within this chance estimate were included in error distance analyses.

Received August 15, 2001
Revision received June 24, 2002
Accepted August 9, 2002 ■

Call for Nominations: *Rehabilitation Psychology*

The APA Publications and Communications (P&C) Board has opened nominations for the editorship of ***Rehabilitation Psychology*** for the years 2006–2011. Bruce Caplan, PhD, is the incumbent editor.

Candidates should be members of APA and should be available to start receiving manuscripts in early 2005 to prepare for issues published in 2006. Please note that the P&C Board encourages participation by members of underrepresented groups in the publication process and would particularly welcome such nominees. Self-nominations are also encouraged.

Rehabilitation Psychology will transition from a division publication to an "all APA" journal in 2006, and the successful candidate will be involved in making suggestions to the P&C Board and APA Journals staff about the transition process.

Gary R. VandenBos, PhD, and Mark Appelbaum, PhD, have been appointed as cochairs for this search.

To nominate candidates, prepare a statement of one page or less in support of each candidate. Address all nominations to

Rehabilitation Psychology Search Committee
Karen Sellman, Search Liaison
Room 2004
American Psychological Association
750 First Street, NE
Washington, DC 20002-4242

The first review of nominations will begin December 8, 2003. The deadline for accepting nominations is **December 15, 2003**.