Article

Musicae Scientiae
2023, Vol. 27(1) 94–116
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10298649211025491
journals.sagepub.com/home/msx

Recovery from memory failure when recalling a memorized performance: The role of musical structure and performance cues

Roger Chaffin
Department of Psychological Sciences, University of Connecticut, USA

Jane Ginsborg
Centre for Music Performance Research, Royal Northern College of Music, UK

James Dixon
Department of Psychological Sciences, University of Connecticut, USA

Alexander P. Demos
Department of Psychology, University of Illinois Chicago, USA

Abstract

To perform reliably and confidently from memory, musicians must able to recover from mistakes and memory failures. We describe how an experienced singer (the second author) recovered from mistakes and gaps in recall as she periodically recalled the score of a piece of vocal music that she had memorized for public performance, writing out the music six times over a five-year period following the performance. Five years after the performance, the singer was still able to recall two-thirds of the piece. When she made mistakes, she recovered and went on, leaving gaps in her written recall that lengthened over time. We determined where in the piece gaps started (*losses*) and ended (*gains*), and compared them with the locations of *structural* beats (starts of sections and phrases) and *performance cues* (*PCs*) that the singer reported using as mental landmarks to keep track of her progress through the piece during the sung, public performance. Gains occurred on structural beats where there was a PC; losses occurred on structural beats without a PC. As the singer's memory faded over time, she increasingly forgot phrases that did not start with a PC and recovered at the starts of phrases that did. Our study shows how PCs enable musicians to recover from memory failures.

Keywords
Memory, recall, music cognition, serial position, recovery, performance

---

Corresponding author:
Roger Chaffin, Department of Psychology, University of Connecticut, 932 Seaside St., Santa Cruz, CA 95060, USA.
Email: Roger.Chaffin@UConn.edu

Musicians performing in the tradition of Western classical music often play from memory, rapidly and accurately recalling thousands of notes in succession. Memory failure is always a possibility. When it happens, experienced performers are usually able to recover by jumping ahead and restarting the performance at a later point, filling in with contextually appropriate material so that most of the audience is unaware of the problem (Chaffin et al., 2002, p. xi). In the present study, we examine how a singer (the second author) recovered from mistakes and gaps in memory as she periodically wrote out the score of a piece from memory over a five-year period following its public performance. Our study is the first to directly examine recoveries when writing out a memorized score. The data were collected as part of a larger longitudinal case study in which the singer recorded her practice as she prepared the first Ricercar from Igor Stravinsky's *Cantata*, a movement for soprano and small instrumental ensemble, for public performance (Ginsborg et al., 2006; Ginsborg & Chaffin, 2011).

Performance memory is largely procedural and implicit (Anderson, 1983; Chaffin & Imreh, 2002; Schacter, 1987). In order to cope with the pressures of the concert stage, performances are usually practiced until they are automatic (Christensen et al., 2016; Lisboa et al., 2018). During practice, familiar sensory-motor patterns (e.g., chords and scales) are braided into new, multi-modal sequences (Diedrichsen & Kornysheva, 2015). When a performance proceeds smoothly, each action cues the next (Farrell, 2012; Logan, 2018, 2021; Pfordresher et al., 2007). Possible continuations are constrained on multiple dimensions (e.g., melody, harmony, rhythm) and time scales (e.g., note-to-note, phrase-to-phrase; Bonini et al., 2011; Halpern & Bower, 1982; Palmer & Pfordresher, 2003). These multiple constraints make performance memory surprisingly robust (Rubin, 2006) and learning largely implicit (Ettlinger et al., 2011).

When performance is disrupted, serial cuing stops. Student musicians often have to go back to the beginning and start over (Lisboa et al., 2015). Experienced musicians usually avoid this humiliation and go on. They create a safety net for when things go wrong by learning to start from a variety of different locations during practice. Like expert memorists in other domains, they use highly practiced retrieval schemes to bring the normal, slow speed of deliberate retrieval from long-term memory up to the pace required for performance (Chaffin & Imreh, 2002; Ericsson & Kintsch, 1995; Williamon & Valentine, 2002). The retrieval scheme provides content-addressable access to multiple points in the automatic action sequence (Chaffin et al., 2002, p. 216–221). When the performance is disrupted, the musician recovers by thinking of a suitable location in the music, e.g., "third repetition of the main theme," and continuing on. We refer to such thoughts as *performance cues* (*PCs*) and to the idea that PCs provide content-addressable access to the automatic sequence of a performance as *PC theory* (Chaffin & Imreh, 2002; Chaffin et al., 2002, pp. 27, 70–72, 94, 197–205; Chaffin et al., 2016; Ginsborg et al., 2012). PCs allow memories that are implicit and procedural to become explicit and declarative.

Musical structure

The organization of Western (common practice) music into movements, sections, and phrases provides performers with a convenient basis for a hierarchical retrieval scheme similar to those used to recall many other kinds of material, from stories to random digit strings (Bisesi & Windsor, 2016; Ericsson & Kintsch, 1995). In preparing for performance, experienced musicians use the musical structure to organize their practice, often starting and stopping at beginnings of phrases and sections (Chaffin et al., 2010; Lisboa et al., 2012; Williamon & Valentine, 2002). PC theory proposes that attending to features of the music while practicing in this way creates retrieval cues for which the musical structure provides a retrieval scheme (Chaffin, & Imreh, 2002).

It has often been proposed that human memory and activity are organized hierarchically (Dell, 1986; Farrell, 2012; Miller et al., 1960; Rosenbaum et al., 2001). For example, skilled typing appears to be organized into words; as each word is completed, keystrokes for the next word are retrieved from long-term memory (Logan, 2018). Similarly, music appears to be organized into phrases and retrieved from long-term memory a phrase at a time. We will refer to this as the *segmentation hypothesis.* The segmentation hypothesis is part of PC theory, which adds the further assumption that PCs provide additional retrieval cues that help with recovery after a gap in recall. Support for the segmentation hypothesis is provided by the observation that hesitations during playing frequently occur between phrases (Chaffin, 2007; Chaffin & Imreh, 2002; Ginsborg & Sloboda, 2007), and by primacy effects when musicians recall music. Primacy effects suggest that, when musicians forget, they are often able to recover at the beginning of a later segment. Recall is best in *structural* locations, that is, at the starts of sections and phrases, and declines as serial position from the structural location increases (Chaffin & Imreh, 2002; Chaffin et al., 2010; Finney & Palmer, 2003; Mishra, 2010).

Serial position effects (primacy and recency) have been reported for the recall of many types of materials, at many different time scales (Brown et al., 2007; Crowder & Greene, 2010). According to serial cuing explanations, primacy effects are due to the accumulating probability of forgetting as each event in a sequence cues the next (Ebbinghaus, 1913; Lewandowsky & Murdock, 1989).[1] According to positional explanations, both primacy and recency effects are due to the greater distinctiveness of positions at beginnings and ends of each segment, making items in these positions easier to retrieve (Brown et al., 2007). Both types of explanation require additional assumptions about segmentation to account for serial position effects in extended sequences (e.g., Burgess & Hitch, 2006; Farrell, 2012; Logan, 2018, 2021; Page & Norris, 2009).

Five studies have reported primacy effects in the recall of music memorized for performance (Chaffin & Imreh, 2002; Chaffin et al., 2010; Finney & Palmer, 2003; Mishra, 2010; Timperman & Miksza, 2019). Finney and Palmer (2003) found that pianists made fewer errors at the beginnings (primacy) and ends (recency) of short, unsegmented musical phrases (23 and 32 notes) when playing, either immediately or a short interval after learning. Two studies reported primacy effects for longer, segmented pieces, albeit without directly testing for effects of serial position. Timperman and Miksza (2019) asked student string players to learn a short étude (2 phrases, 56 notes) and then perform it twice, from memory, the first time shortly after learning and, again, 24 hours later. They report that errors increased from beginning to end of the piece in both performances. Similarly, Mishra (2010) asked students to learn a somewhat longer étude on a variety of instruments (3 sections, 9 phrases, and 72 beats) and then perform it after a retention interval of 25 minutes. The author concluded that there was a primacy effect for phrases within sections, but not for bars within phrases, and a recency effect at the end of the piece. A second experiment found that student pianists playing previously memorized pieces made more errors in locations that they identified as difficult and non-structural, suggesting that difficult structural bars were protected from being forgotten. Two case studies that reported serial position effects for PCs as well as for musical structure are described in the next section.

---

[1] This explanation is often referred to as *associative chaining* (e.g., Chaffin et al., 2017). We use the term *serial cuing* to avoid the suggestion that cuing is limited to item-to-item associations (Lindsey & Logan, 2019, in press). We assume that music performance also involves associations between non-adjacent items and between the music and the context in which it is performed (Logan, 2018, 2021).

Performance cues

Musical structure provides an abstract, bird's-eye view of an entire piece of music, revealing the relationship between its various segments. During performance, however, music unfolds sequentially, like a path revealing more of itself with each step along the way (Cook, 2013, pp. 45–49; Lisboa et al., 2018). Just as walkers use landmarks to keep track of their progress, performers track their progress through a piece by identifying musical landmarks on a mental map (e.g., Chueke & Chaffin, 2016). PC theory proposes that these musical landmarks guide the performance, serve as retrieval cues to elicit the upcoming passage from long-term memory, and provide *points of recovery* where playing can resume when the performance is disrupted (Chaffin & Imreh, 2002; Chaffin et al., 2010). Musical structure provides the map; PCs are the landmarks or points of recovery.

PC theory originated in a longitudinal case study in which an experienced piano soloist recorded her practice as she prepared J. S. Bach's Italian Concerto (Presto) for performance (Chaffin et al., 2002). The goal was to document the pianist's intuition that she prepared for performance by "somehow remap[ping her] thinking to emphasize the artistic, inspirational elements without ever losing control" (p. 27). After the performance, the pianist reported the thoughts (PCs) that she had attended to while playing, grouping them into three main types: *Expressive PCs* referred to feelings to be expressed (e.g., surprise, excitement), *interpretive PCs* to musical gestures that conveyed those feelings (e.g., staccato, forte), and *basic PCs* to critical details of technique that enabled the gestures (e.g., fingering, big leaps). (We will refer to expressive and interpretive PCs collectively as *expressive/interpretive PCs*).

When the pianist's PC reports were compared with her practice, it was clear that she had been paying attention to the musical features that she later reported as PCs throughout the 10 months of practice. PCs also affected the pianist's written recall of the piece, 27 months after the performance, when she wrote out the first page of the score from memory. There were primacy effects for three serial position predictors, coded by numbering bars from the start of each section, phrase, and expressive PC. Basic PCs had an opposite effect; there were more errors in bars containing basic PCs than in bars without. These effects were replicated in a second longitudinal case study in which a cellist recorded her learning of Bach's Suite No. 6 (Prelude) for solo cello and then wrote out the entire piece (14 sections, 44 phrases, and 1,349 notes) from memory, 10 months after last performing it (Chaffin et al., 2010). The cellist's expressive/interpretive PCs also shaped the tempo arches by which she communicated her interpretation of the musical structure to listeners (Demos et al., 2020). We refer to these studies below as the *Presto* and *Prelude* studies, respectively.

PC theory attributes primacy effects in recall to recoveries at the start of a new segment, where retrieval cues provide content-addressable access to the memorized sequence of the music (Chaffin & Imreh, 2002). When a performance proceeds smoothly, phrases are retrieved from long-term memory sequentially, one at a time (cf. Logan, 2018). When memory fails, the musician recovers by retrieving a subsequent phrase from long-term memory and restarting the sequence. After a recovery, the probability of mistakes increases again as distance from the recovery increases due to either the simple accumulation of probability, declining distinctiveness, or both. Primacy effects in the Presto and Prelude studies suggest that recoveries occur at structural locations, that is, at the starts of sections (*section starts*) and phrases (*phrase starts*), and/or at expressive/interpretive PCs. The increased error rate in bars with basic PCs may be due to the musicians writing out the music from memory, rather than playing it, rendering irrelevant the technical features to which basic PCs point (Chaffin & Imreh, 2002; cf. Godden & Baddeley, 1975). If so, then we may not find the same effect for basic PCs in the present study because, unlike instrumentalists, singers always have their instrument with them; they can sing or hum as they write.

The Presto and Prelude studies did not separate the contributions of structure and PCs. In the Presto study, structural boundaries and expressive/interpretive PCs were too closely aligned to be examined separately. In the Prelude study, they were less closely aligned.[2] A later re-analysis took advantage of this to compare phrases starting with and without PCs, using a mixed model analysis similar to that used in the present study (Lisboa et al., 2018).[3] The re-analysis showed that there was a primacy effect for phrases that started with an expressive/interpretive PC but not for phrases did not. This could mean that expressive/interpretive PCs, rather than structural boundaries, are the retrieval cues responsible for recovery. Or, expressive/interpretive PCs may protect entire phrases from being forgotten; or both. The present study provided an opportunity to explore these questions because expressive/interpretive PCs and musical structure were more independent of each other than in the earlier studies. As in the Prelude study, there were phrase starts with and without PCs; unlike the Prelude, there were PCs in non-structural, as well as in structural, locations.

Primacy effects are not the most efficient way of identifying the location of recoveries. Recoveries occur from one beat (forgotten) to the next (remembered), whereas primacy effects encompass entire segments (phrases, sections, or piece). In the present study, we identified beats where a *recovery* occurred by scoring each beat for errors and then taking the derivative to determine where errors decreased from one beat to the next. Correspondingly, we identified beats where errors increased as cases of *forgetting*. We treated forgetting and recovery as two separate, binary variables, which we refer to as *gains* and *losses*. Losses indicated forgetting (increased error); gains indicated recovery (decreased error).

The segmentation hypothesis predicts that gains will be more frequent at the starts of segments, because this is where retrieval cues provide access to long-term memory. To test this prediction, we compared the frequency of gains on the first, *structural beat* of a segment with the frequency on later *non-structural beats*. The segmentation hypothesis also makes predictions about the effects of structure on losses but, unlike gains, the size and direction of the effect depends on whether losses are more frequent within or between segments. If forgetting is more frequent within segments, then losses will be more frequent on non-structural than on structural beats. If forgetting is more frequent between segments, then vice versa—losses will be more frequent on structural beats. The primacy effects for phrases and sections in the Presto and Prelude studies, and for sections in Mishra's (2010) study, indicate more forgetting within than between segments, suggesting that we will find more losses on non-structural than on structural beats. Conversely, if we find more losses on structural beats, then we will not expect a primacy effect at these locations.

It may seem unlikely that gains and losses would both occur on structural beats, since they are mutually exclusive—more gains imply fewer losses, and vice versa, other things being equal. However, according to the segmentation hypothesis, other things are not equal on structural beats because this is where retrieval from long-term memory occurs, or fails to occur.

---

[2] In the Presto and Prelude studies, all expressive/interpretive PCs occurred at section and phrase starts, and vice versa, for the Presto but not for the Prelude; in the Prelude study, approximately a quarter of the phrases did not start with an expressive/interpretive PC.

[3] The re-analysis of the Prelude examined phrases as a single level of musical structure, ignoring sections. In contrast, the present study examined section and phrase starts separately.

Losses result from failure to retrieve the next segment; gains from recovery when a subsequent segment is successfully retrieved. Thus, gains and losses can both occur on structural beats, just not on the same beat at the same time.

The present study

The present study was conducted concurrently with the Prelude study with the goal of expanding the study of PCs to a different composer (Stravinsky vs Bach), from a different time period (20th vs 18th century), for a different instrument (voice vs piano/cello), and musical forces (instrumental ensemble vs solo). We have previously described the singer's practice and written recall, but did not report the effects of expressive PCs, noting only that there were unresolved problems in their analysis (Ginsborg & Chaffin, 2011). Those problems are addressed here in three ways. First, we examined the data in two stages, first identifying where forgetting and recovery occurred, as described above, and then examining serial position effects with respect to those locations. Second, we reduced collinearity between predictors by restricting their number. Third, we used generalized linear mixed models (GLMM), rather than general linear models (GLM), to account more accurately for the nested nature of the data and apportion variance between the different levels of the temporal hierarchy (Demos & Chaffin, 2017; Singer & Willett, 2003).

In addition to these changes in the data analysis, we made two other important changes from the methods used in the Presto and Prelude studies. First, the singer wrote out the score from memory nine times over a five-year period, rather than just once as in the earlier studies. We examined data from the last six of these *recall events*, starting 14 months after the performance, when the singer first began to make more than trivial mistakes in recall. The multiple recall events allowed us to track the weakening of the singer's memory over time, determine when forgetting began, and examine serial position across the entire piece (as well as across sections and phrases). We accepted the limitation that we would not know the extent to which the multiple recall events affected the singer's memory. Writing out her part from memory was part of the singer's normal preparation for performance, a way of ensuring that her memory was secure. Although writing out the score repeatedly without also performing it, as in our study, was not normal, it seemed likely to enhance the kind of explicit recall that we were interested in.

Second, in most memory studies, stimulus properties affecting memory are controlled by randomization. Music, however, cannot be randomly ordered. By including the number of repetitions of each beat during practice as a predictor in our analyses, we controlled for variability in the data due to stimulus properties that affected both practice and recall, such as difficulty and familiarity.

In sum, we used GLMM to look for effects of musical boundaries and PCs on the written recall of a memorized performance, including practice as a predictor. We analyzed the data in two stages. First, we asked whether gains and losses were more frequent at section and phrase starts, and at PCs, than at other locations. Then, we looked for serial position effects at locations where there were more gains, examining three levels of musical structure, words within phrases, phrases within sections, and sections within the piece.

Table 1. Nested musical structure of Ricercar I (omitting rests).

Section #	1			2			3			4			5		6		7		8			9						
Phrase #	1	2	3	1	2	3	1	2	3	1	2	3	1	2	1	2	1	2	1	2	3	1	2	3	4	5	6	7
# of Word	3	6	7	6	4	3	3	3	3	6	4	3	6	6	6	7	6	10	6	3	4	3	1	6	3	5	4	1

Table 2. Descriptive statistics for three levels of music structure (omitting rests).

	Phrases per section	Words per phrase	Beats per word
Mean	3.22	4.54	1.78
Std	1.56	2.01	1.06
Median	3	4	2
min	2	1	1
max	7	10	8

Method

Learning the Ricercar

Jane Ginsborg, a former professional singer, performed as solo soprano in a public performance of Stravinsky's Cantata for two solo singers, women's choir and small instrumental ensemble. The Cantata includes a movement for solo soprano and ensemble, Ricercar 1, which was the subject of this study. The singer prepared the Ricercar for performance in 15 practice sessions and rehearsals, totaling just over 8 hours, over a period of one month that ended with the public performance. All practice sessions and performances were recorded and transcribed as part of a longitudinal self-study of the singer's practice (Ginsborg et al., 2006; Ginsborg & Chaffin, 2011).

The Ricercar lasts about 4 minutes, consists of 250 beats, scored in 70 bars that alternate intermittently between 3:8 and 4:8 meter, and contains 276 notes and 16 notations indicating rests, where the singer was silent. The archaic English text is divided into nine sections (according to the singer's analysis): four verses of varying lengths, separated by three refrains, followed by a recitative, and closing prayer (see Ginsborg et al., 2006, Figure 3).

Table 1 shows the nested musical structure: words within phrases, within sections. The highest level, the piece, is not labelled. Table 2 adds the lower level of musical beats to show the average number of items in each grouping at each level. Not included in either table are seven short instrumental passages at beginnings of sections in which the soloist does not sing (rests), which we excluded from our analyses. Rests varied in length from 1 to 6 beats, for a total of 26 beats. We counted 126 words, although the score actually contains 140. When a word ended on the same beat that the next word began, we counted the two as one word; for example, "in my" counted as one word.

Recall

Following her usual custom, the singer wrote out the words along with their pitches and rhythms before the last rehearsal, humming and conducting to help her recall (Ginsborg,

Table 3. Number of non-structural and structural beats (section starts and phrase starts), with and without PCs, excluding beats of rest (*N*=26) and the first and last bars of the piece. Baseline frequencies are shown in boldface. Footnotes list frequencies separately for each type and combination of types of PC (exp=expressive; int=interpretive).

Musical structure	No PC	With at least 1 PC	Total
Non-structural	148	43[a]	**191**
Section start	1	7[b]	8
Phrase start	13	7[c]	20
Total	**162**	57	219

[a] exp (5), int (18), prepare (9), basic (10), int+prepare (1).
[b] exp (2), int (1), basic (1), exp+int (2), exp+basic (1).
[c] int (4), basic (1), exp+int (1), exp+basic (1).

2009). Also, as part of her preparation, she gave three uninterrupted practice performances from memory with piano accompaniment. The public performance, on 16th December 2003, was also from memory and was error-free. The initial written recall before the performance contained only two trivial errors, as did two subsequent recalls at the end of January and February 2004. These early recalls were not included in the study. Our data begin with the first time that the singer made a substantial number of errors in recall, 14 months after the performance, when she wrote out the piece in February 2005. We report data for this and five additional recall events in June 2005, August 2006, June 2007, November 2007 and November 2008 (Recall Events 1–6).[4] Intervals since the public performance were 14, 18, 32, 42, 47, and 59 months respectively.[5]

During the nearly five-year period of the study, following the performance, the singer did not practice or perform the Ricercar. She did, however, score her own recalls, transcribe her practice, and engage in other work for the self-study, including writing papers and giving talks. Each time that she resumed work on the project, she began by writing out the Ricercar from memory and scoring her written recall. The intervals of time between recall events since last consulting the score were 10, 4, 10, 6, 5, and 4 months respectively (mean = 6.5 months).

The singer scored each quaver beat (¼-note) as correct (0) or as an error (1) if the beat was omitted or recalled imperfectly. We refer to this measure as *error* or *error rate*, depending on context. Imperfectly recalled beats were scored for three types of errors (word, pitch, and rhythm/duration), ignoring misspellings of pitch and rhythm such as A# for Bb and incorrectly positioned bar lines. In combination, the two scorings provided a 5-point measure of *number of errors*, running from 0 (correct), to 1, 2, or 3 errors, to 4 (omitted), that we used to construct the gains and losses measures described below.

---

4. For Recall Events 1 and 2, after working through the piece from start to finish, the singer engaged in additional efforts to remember passages that she had forgotten, going back through the piece, reconstructing as much as she could from memory. These efforts added minimally to her accuracy and so, in Recall Events 3–6, she simply worked through the piece once from start to finish.
5. We excluded from the study two additional recall events that occurred in July 2009, when the singer attempted to sing the piece from memory twice (see Ginsborg & Chaffin, 2009).

Reports

The day after the public performance, the singer marked the musical structure and the PCs she had paid attention to during the performance on separate copies of the score, indicating each location with an arrow which she annotated to indicate the level of musical structure (section or phrase) or type of PC (expression, e.g., "yearning", n=12), interpretation (word stress or pronunciation, n=28), preparation (for an entry, n=11), and basic (e.g., "breathe", n=14). We excluded a fifth type of PC (shared), involving coordination with the conductor, because it overlapped with the other types (see Ginsborg et al., 2006).[6]

Table 3 lists the frequency of non-structural beats and two kinds of structural beat (*section starts* and *phrase starts*), with and without PCs, excluding rest beats (n=26 beats) and the first and last bars (n=8 beats), which were not included in the analyses.[7] About a third of phrases started with a PC, allowing us to separately examine phrases that started with and without a PC ($phrase_{start+PC}$ and $phrase_{start\ \text{no-PC}}$ respectively), as in the Prelude study. We could not do the same for sections with and without PCs because there was only one section start without a PC. So, we examined the effect of section starts by combining section starts with and without a PC ($section_{start}$).

The great majority of beats were non-structural (n=191; 87.21%), and without PCs (n=162; 79.97%). These frequencies are shown in Table 3 in boldface to indicate that they provided the baselines against which the effects of musical structure and PC-type, respectively, were compared, as described below. There were a substantial number of PCs of each type on non-structural beats, making it possible to examine each type of PC separately for these beats. The number of PCs of each type is listed in the footnotes to Table 3, separately for each row, with beats with multiple PCs of different types on the same beat enumerated separately.[8]

We analyzed the data using logistic mixed models, adding predictors in stages to identify the best fitting model (Singer & Willett, 2003). First, the *gain/loss analyses (G/L-Models)* examined gains and losses separately, looking at three types of structural boundary (section starts, and phrase starts with and without PCs) and at non-structural locations with each type of PC. Second, the *serial position analysis (SP-Models)* examined errors, looking at the effects of serial position at three levels of musical structure (phrase, section, and piece). Both analyses examined binary data coded for each beat (0 or 1). Both examined effects of interest while accounting for the nested nature of the temporal hierarchy (beats, words, phrases, sections, piece, recall event), as described below. Both started with a null (baseline) model (Model 0) that included, as predictors, amount of practice and recall event. Practice controlled for properties of the music, such as difficulty, and was measured as the (normalized) number of times each beat was repeated during practice (Ginsborg & Chaffin, 2011). Recall event indicated change over time; we will make this explicit by referring to *recall event (time)*. We report effect sizes as odds ratios in the tables, but show effects as predicted probabilities in the figures to facilitate comparison with other studies.[9]

---

[6] In previous reports, PCs that we refer to here as *interpretive* and *basic* were referred to more specifically as *word* and *technical* PCs, respectively (Ginsborg et al., 2006; Ginsborg & Chaffin, 2011).

[7] Section and phrase starts are counted separately because, although starts of sections are necessarily starts of phrases, we coded them as one or the other, but not both, to make them independent in the analyses.

[8] The one non-structural beat with both an interpretive and preparation PC on the same beat was classified as a preparation PC in the analysis.

[9] An odds ratio of 1 means there is no effect. For binary predictors (e.g., PC/no PC), odds ratios greater than 1 indicate the difference in the odds for the two levels of the predictor. For continuous predictors (e.g., serial position), odds ratios indicate the slope of the linear function across levels of the predictor.

Analyses used the glmmTMB (1.0.0) package in R (3.6.2) for generalized linear mixed modeling with a logit linking function (Brooks et al., 2017), graphed results using the effects (4.1-4) package (Fox, 2003) and the ggplot2 (3.3.0) package (Wickham, 2016), and adapted tables from the sjPlot (2.8.3) package (Lüdecke, 2020).

Gain/loss analysis

Dependent variable. Losses were beats on which forgetting increased compared to the previous beat; gains were beats on which there was recovery from forgetting, either partial or complete, compared to the previous beat. We calculated gains and losses by differentiating the 5-point measure of number of errors described above. After differentiation, we separated gains and losses into two separate dependent variables which we converted to binary scales. Gains were negative changes (fewer errors), coded as 1. Losses were positive changes (more errors), also coded as 1.[10]

Analysis method. As shown in Table 3, there were substantial numbers of non-structural beats with and without PCs, a much smaller number of phrase starts with and without PCs, and only one section start without a PC. Accordingly, we treated PCs differently for non-structural beats, section starts, and phrase starts, but did so in a single analysis.

Fixed effects. We examined three types of structural beat (section starts, phrase starts without a PC, and phrase starts with a PC) treating PCs as a single type, and four types of non-structural beat (beats with expressive, interpretive, preparation, and basic PCs), examining each PC-type separately.[11] We compared each type of structural beat to a baseline of non-structural beats (n=191), and compared non-structural beats with PCs to a slightly different baseline of beats without PCs (n=162), as indicated in Table 3 in boldface. Non-structural beats without PCs (n=148) were common to both baselines.

Random effects. We used words (made up of beats) as the random intercept, and allowed a random slope of recall event (time) as a function of words. In all of the models, the random slopes were not allowed to correlate with the random intercepts to encourage the models to converge.

Serial position analysis

Dependent variable. We measured error rate by coding each beat as an error (1) or not (0), as described above.[12]

---

[10] For both measures (gains and losses), presence was coded 1, and absence 0.

[11] To make these comparisons we created two predictors, Structure and PC-type, by dummy-coding each beat based on the singer's reports. Structure examined structural beats and treated PCs as a single type: 0 = non-structural (PC = 0 or 1); 1 = section start with/without PC; 2 = phrase start with no PC; 3 = phrase start with PC. PC-type examined non-structural beats and treated PCs as four different types: 0 = no PC (Structure = 0 or 1); 1 = basic PC (Structure = 0); 2 = preparation PC (Structure = 0); 3 = interpretive PC (Structure = 0); 4 = expressive PC (Structure = 0).

[12] We used the binary measure for the error rate because it provided a more stable and replicable analysis than the count of number of errors (0–5) which we examined in preliminary modeling using zero-inflated count GLMM models.

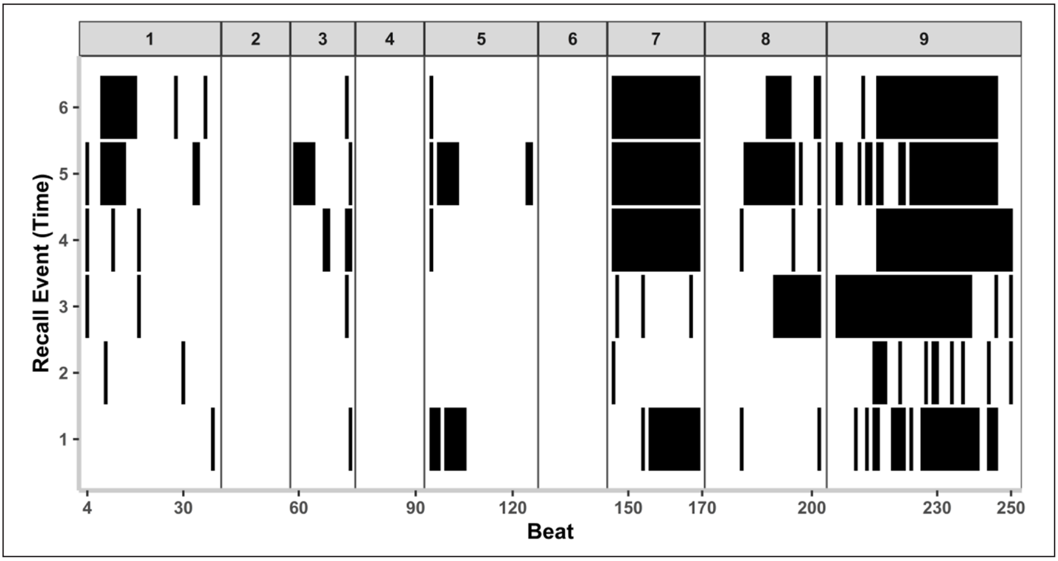

Figure 1. Heat plot of errors (error=black; correct=white) for each beat (excluding rests), with each of the nine sections in separate panels, as a function of time (represented by successive recall events).

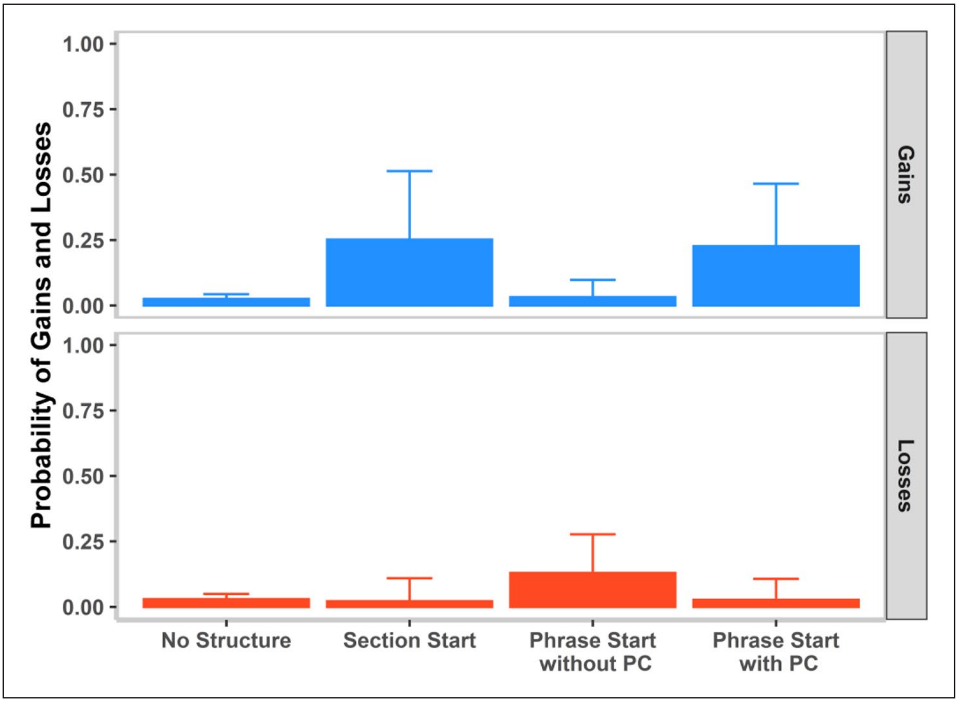

Figure 2. Probability of gains and losses for non-structural beats (no structure) and three types of structural beat (starts of sections and starts of phrases with and without a PC), estimated by G/L-Model 2. Error bars indicate 95% confidence intervals.

Table 4. Summary of comparisons between mixed effect logistic models for gains and losses, showing whether the change in predictors improved the fit of the model compared to the previous model.

Model comparision	Change in DF	Gains			Losses		
		Deviance	χ^2	p	Deviance	χ^2	p
G/L-Model 1 [Vs 0]	+3	607.67	34.87	<.0001	585.62	11.51	.009
G/L-Model 2 [Vs 1]	+4	600.13	7.55	.110	584.51	1.11	.89

Note: G/L-Model 0=Practice + Recall Event (Time) + Wordstart.
G/L-Model 1=G/L-Model 0 + Structure.
G/L-Model 2=G/L-Model 1 + PC-Type.

Analysis method. We tested the effects of serial position at three levels of musical structure: words within phrases, phrases within sections, and sections within piece. For brevity, we sometimes refer to serial position for *phrases*, *sections*, and *piece* respectively.

Fixed effects. The effects of serial position were tested simultaneously at all three levels (fixed and random slopes) in the same model. We coded serial position separately for each level by numbering from the beginning of each grouping, expressing each serial position as a percentage (0–100), and coding each beat in the same grouping with the same percentage. For example, we numbered sections from the beginning of the piece (1–9), converted these serial positions to percentages of the piece [$SP_{section}$ = (Section # -1/total number of sections -1)*100], and coded each beat in the same section with the same value.

Random effects. We used three sets of random terms for this model. First, we allowed a random slope for the serial position of sections within piece as a function of recall event (random intercept). Second, we allowed a random slope of serial position of phrases within sections as a function of section (random intercept, nested within each recall event). Finally, we allowed a random slope of serial position of words within phrases as a function of the phrase (random intercept, nested within sections). (Words were made up of beats, the lowest level of the temporal hierarchy.) In all of the models, the random slopes were not allowed to correlate with the random intercepts to ensure each model converged.

Results

Figure 1 shows errors (black) and correct (white) recall for each beat, with sections of the piece in separate, vertical panels (left-to-right) and successive recall events in separate horizontal panels, reflecting the passage of time (bottom-to-top). The figure shows the singer's written recall from the first time that she made a substantial number of errors (error rate = 0.25), 14 months after the performance, until the sixth recall event, almost four years later, by which time the error rate had increased (error rate = 0.36). The separation of sections into panels in Figure 1 reveals that forgetting (loss) and recovery (gain) often occurred at the starts of sections. Figure 1 does not show the lower level of musical organization into phrases, but, as described next, losses and gains also often occurred at starts of phrases. Forgetting started with a failure to recall the next segment and continued until the successful retrieval of a subsequent segment allowed written recall to restart (recover). The resulting *gaps* in recall, between forgetting and subsequent recovery, appear in black. The increase in errors over time, noted above, was a result of gaps growing longer.

Table 5a. Summary of logistic mixed models for gains showing odds ratio, standard error, and *p* value for practice, recall event (time), structure (section$_{\text{start}}$, phrase$_{\text{start no-PC}}$, phrase$_{\text{start+PC}}$), and PC-type (basic, preparation, interpretive, expressive) on non-structural beats.

Predictors	G-Model 0			G-Model 1			G-Model 2		
	Odds ratio	*Std. error*	*p*	*Odds ratio*	*Std. error*	*p*	*Odds ratio*	*Std. error*	*p*
(Intercept)	0.03	0.32	**<.001**	0.03	0.32	**<.001**	0.03	0.33	**<.001**
Practice	1.79	0.18	**.001**	1.73	0.17	**.002**	1.68	0.18	**.003**
Recall event (time)	0.98	0.14	.905	0.99	0.15	.919	0.98	0.15	.887
Word$_{\text{start}}$	1.80	0.26	**.024**	1.08	0.30	.791	1.22	0.30	.521
Structure									
Section$_{\text{start}}$				14.40	0.61	**<.001**	12.97	0.61	**<.001**
Phrase$_{\text{start no-PC}}$				1.42	0.56	.534	1.27	0.57	.669
Phrase$_{\text{start+PC}}$				12.10	0.60	**<.001**	11.29	0.60	**<.001**
PC-type									
PC$_{\text{basic}}$							0.87	0.70	.840
PC$_{\text{preparation}}$							1.94	0.52	.202
PC$_{\text{interpretive}}$							0.21	1.08	.143
PC$_{\text{expressive}}$							0.18	1.23	.168
Random effects (variance)									
(Intercept) | Word	1.38			1.13			1.13		
Recall | | Word	0.45			0.57			0.60		
Model fit									
AIC	654.541			625.673			626.126		
log-Likelihood	–321.270			–303.837			–300.063		

Gains and losses

The effects of musical structure and PCs on structural beats are summarized in Figure 2, which was generated by the best fitting model, G/L-Model 2. (The effect of PC-type on non-structural beats was not significant and is not shown). Gains and losses are shown in separate panels. There were very few gains or losses on baseline beats. Most gains and losses occurred on the relatively small number of structural beats, as predicted by the segmentation hypothesis. Losses occurred at phrase starts without PCs; gains occurred at phrase starts with PCs and at section starts. Since there were PCs at most section starts, this means that most gains occurred on structural beats with PCs, as predicted by PC theory. Thus, the gaps in recall noted in Figure 1 were mostly due to the forgetting of phrases without PCs, followed by recovery at phrase and section starts with PCs.

Model fit. Table 4 compares the models, separately for gains and losses, showing which models provided significantly better fits to the data. The null model included practice, recall event (time), and starts of words. Adding the three types of structural beat to the null model in Model 1 significantly improved the fit to the data for both gains and losses. Adding the four types of PC on non-structural beats in Model 2 did not significantly further improve the fit for either measure.

Table 5b. Summary of logistic mixed models for losses showing odds ratio, standard error, and p value for practice, recall event (time), structure (section$_{start}$, phrase$_{start\ no\text{-}PC}$, phrase$_{start+PC}$), and PC-type (basic, preparation, interpretive, expressive) on non-structural beats.

Predictors	L-Model 0			L-Model 1			L-Model 2		
	Odds ratio	*Std. error*	*P*	*Odds ratio*	*Std. error*	*p*	*Odds ratio*	*Std. error*	*p*
(Intercept)	0.03	0.32	**<.001**	0.03	0.32	**<0.001**	0.03	0.32	**<.001**
Practice	1.99	0.17	**<.001**	1.90	0.17	**<0.001**	1.89	0.17	**<.001**
Recall event (time)	0.95	0.15	.749	0.95	0.15	0.717	0.95	0.15	.718
Word$_{start}$	1.07	0.25	.792	0.80	0.29	0.439	0.82	0.29	.498
Structure									
Section$_{start}$				0.78	0.86	0.773	0.74	0.87	.724
Phrase$_{start\ no\text{-}PC}$				5.21	0.50	**0.001**	4.92	0.51	**.002**
Phrase$_{start+PC}$				0.95	0.73	0.943	0.90	0.73	.884
PC-type									
PC$_{basic}$							0.59	0.67	.430
PC$_{preparation}$							0.93	0.54	.896
PC$_{interpretation}$							0.60	0.82	.537
PC$_{expressive}$							1.14	1.16	.912
Random effects (variance)									
(Intercept) | Word	1.02			0.97			0.95		
Recall | | Word	0.48			0.53			0.53		
Model fit									
AIC	609.129			603.619			610.508		
log-Likelihood	–298.564			–292.810			–292.254		

Model description. Tables 5a and 5b summarize the models for gains and losses respectively (*G-Models* and *L-Models*). For each predictor, the tables show the odds ratio, indicating effect size; standard error, indicating fit to the data; and p value, based on Wald-z. We describe the effects in the order listed.

Gains and losses both increased with amount of practice, as indicated by the significant positive effect of practice on both gains and losses in all models. Additional analyses (see Supplementary Materials) showed that removing practice from the final models did not change the significance of the effects of musical structure and PCs described below, although it did slightly decrease the size of those effects and significantly decreased model fit. The effect of recall event (time) was not significant for gains or losses, indicating no change in the number of gains or losses over time.

Gains occurred on structural beats with PCs. Losses occurred on structural beats without PCs. There were minimal gains or losses on non-structural beats, with or without PCs. Gains on structural beats with PCs are indicated by the significant effects of section starts and phrase starts with PCs, showing that gains were 12.97 times more likely at sections starts and 11.29 times more likely at phrase starts with PCs than on baseline beats. The difference between phrase starts with and without PCs was significant, $OR = 8.86$, $p = .006$, while the difference between section starts and phrase starts with PCs was not, $OR = 0.871$, $p = .86$. Since all but one section started with a PC, as described above, this means that gains occurred at PCs located on structural beats with PCs, not on structural beats without PCs, or on non-structural beats.

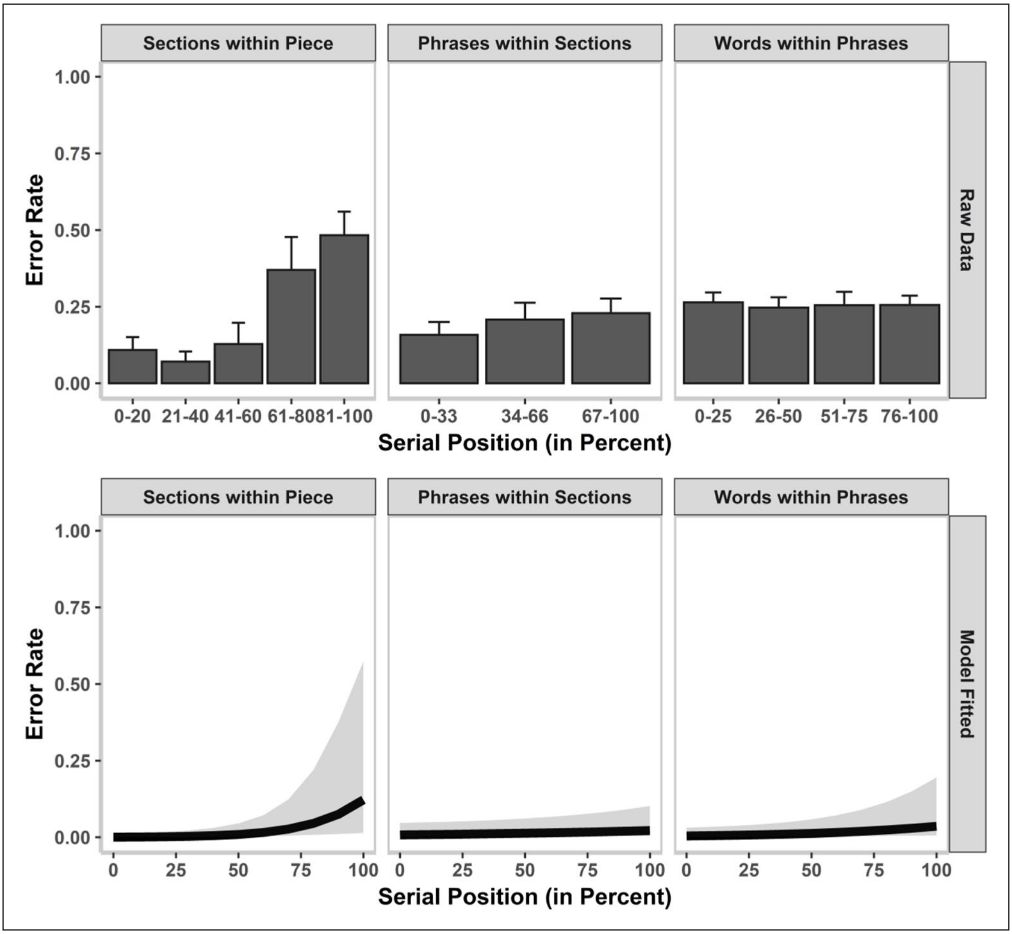

Figure 3. Error rate as a function of serial position at three levels of musical structure. The top row shows the data averaged across levels with error bars indicting the standard error. The bottom row shows probability of generating an error estimated by the best fit serial position model (SP-Model 2), with shading indicating 95% confidence intervals.

Losses on structural beats without PCs are indicated by the significant effect for phrase starts without PCs, indicating that losses were 4.92 times more likely at phrase starts without PCs than on baseline beats (Table 5b, L-Model 2). Thus, PCs at the start of a phrase prevented forgetting as well as enhancing recovery.

PCs on non-structural beats had no effect on gains or losses, as indicated by non-significant effects for all four PC types in Tables 5a and 5b. Odds ratios for most PC types were close to 1, indicating that gains and losses were equally minimal with or without a PC. Thus, PCs were not effective retrieval cues on non-structural beats, only on structural beats.

Serial position

We examined the effect of serial position on error rate with respect to starts of phrases, sections, and the piece. Figure 3 shows the effect of serial position at the three levels of musical structure.

Table 6. Summary of comparisons between mixed effect models for error rate, showing whether the additional predictors changed the fit of the model to the data compared to the previous model.

Serial position models		Accuracy		
Model comparision	Change in DF	Deviance	χ^2	*p*
SP-Model 1 [Vs 0]	+3	763.94	15.30	.002
SP-Model 2 [Vs 1]	+1	758.10	5.83	.015

Note: SP-Model 0=Practice + Recall.
SP-Model 1=SP-Model 0 + SP Sections + SP Phrases + SP Words.
SP-Model 2=SP-Model 1 + Words per phrase.

Table 7. Summary of mixed effects models for error rate, showing effect size, standard error, and *p* value for practice, recall event (time), serial position (SP) at three levels of musical structure (piece, section, and phrase), and number of words per phrase.

Predictors	SP-Model 1			SP-Model 2		
	Odds Ratio	*Std. Error*	*p*	*Odds Ratio*	*Std. Error*	*p*
(Intercept)	0.01	0.8	**<.001**	0.01	0.8	**<.001**
Practice	4.45	0.48	**.002**	4.48	0.48	**.002**
Recall event (time)	3.81	0.61	**.027**	4.29	0.63	**.021**
SP sections in piece	6.97	0.71	**.006**	6.68	0.72	**.008**
SP phrases w/ sections	2.44	0.47	.057	2.47	0.48	.061
SP words w/ phrases	1.53	0.26	.108	1.5	0.27	.131
Words per phrase				2.92	0.46	**.021**
Random effects (variance)						
(Intercept) | Piece	1.31×10^{-11}			1.15×10^{-9}		
SP Sections | Piece	3.12×10^{-8}			3.72×10^{-8}		
(Intercept) | Piece:Sections	14.30			14.34		
SP Phrases | Piece:Sections	6.02			6.67		
(Intercept) | Piece:Sections:Phrases	6.12			5.20		
SP Words | Piece:Sections:Phrases	3.39			3.93		
Model fit						
AIC	787.935			784.103		
log-Likelihood	–381.968			–379.051		

The figure displays the same data twice, in two different ways. In the top row, error rates were calculated by averaging across beats and other levels of the temporal hierarchy, as is typical when data are analyzed using GLM. In the bottom row, error rates (i.e., the model predicted probability of making an error) were estimated by the best fitting mixed model, as is typical when using GLMM (Table 4, Model 2). The raw data (shown in Figure 1) were the same in each case. We display the data in both ways to demonstrate that the two methods of computation produce similar results; we format them differently as a reminder of the difference in computation.

Both rows in Figure 3 show a primacy effect at the top level of musical structure (left panel), and not at the two lower levels (middle and right panels). For sections within the piece (left

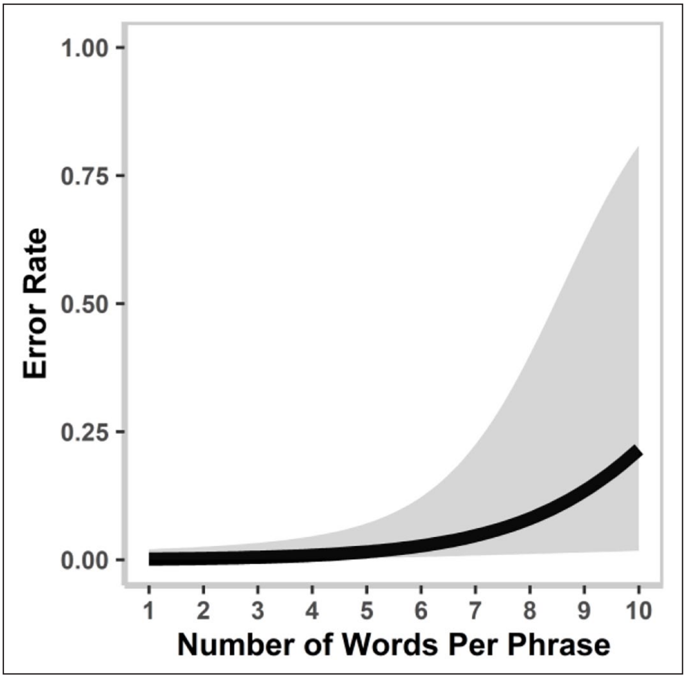

Figure 4. Error rate as a function of number of words per phrase estimated by SP-Model 2, with shading indicting the 95% confidence intervals.

panel), errors increased with serial position with most of the increase at later serial positions. In contrast, the serial position functions for phrases within sections and words within phrases (middle and right panels respectively) were almost flat, indicating no effect of serial position. To see if the sharp increase at later serial positions for sections within the piece (left panels) was due to increased errors for longer phrases, rather than to serial position, we included the number of words in a phrase as a predictor in Model 2.

Model fit summary. Table 6 compares the models, showing which provided significantly better fits to the data. The null model included practice and recall event (time). Adding the musical structure to the null model in Model 1 significantly improved the fit to the data. Adding the number of words per phrase in Model 2 significantly improved the fit still further.

Model description. Table 7 summarizes the serial position models. We describe the effects in the order listed. There were more errors on beats that were practiced more, as indicated by the significant effect of practice in both models. As with the gain/loss analyses, additional analyses (see Supplementary Materials) showed that removing practice from the models did not change the other effects, described below, while slightly decreasing their size and significantly decreasing model fit.

There was a significant effect of recall event in both models, indicating that errors increased over time, as seen in Figure 1. Coupled with the finding, reported above, that losses and gains did not increase over time (see Tables 5a & 5b), the increase in errors means that gaps between forgetting and recovery grew longer, while the number of gaps remained roughly the same.

When the singer forgot, she had to go further forward in the piece before reaching a point of recovery. Thus, points of recovery became fewer and further between.

The primacy effect for sections in the piece, seen in Figure 3 (left panels), is reflected in Table 7 in a significant positive effect of the serial position of sections in the piece in both models. The odds of an error increased by 6.68 with each successive section (SP-Model 2). In contrast, the flat serial position functions for phrases within sections, and words within phrases, seen in in the center and righthand panels of Figure 3, were reflected in odds ratios that were close to 1 and not significant.

Figure 4 shows the error rates as a function of number of words in a phrase. Errors were more likely in longer phrases, as indicated by the significant main effect for words per phrase in SP-Model 2 (see Table 7). Errors were minimal for short phrases and increased sharply in phrases that contained six or more words (see Table 1). Even when we controlled for the length of the phrase, by adding the number of words in SP-Model 2, there was still no overall effect of the serial position of words within phrases, suggesting that the singer forgot phrases in their entirety (as seen in Figure 1).

Discussion

Although serial recall has been studied for many years, our study is the first to show how the ability to recall long musical sequences from memory depends on recovering from mistakes and omissions. As predicted by the segmentation hypothesis, gains and losses occurred at the starts of segments (sections and phrases), rarely within a segment. Phrases were recalled from long-term memory in all-or-none fashion, remembered or forgotten in their entirety, similar to the way skilled typists retrieve keystrokes from memory, a word at a time (Logan, 2018). When recall failed, content-addressable retrieval cues at the start of a subsequent segment enabled recovery. This pattern of forgetting and recovery is consistent with theories of serial recall that allow for the direct cuing of segments of a longer sequence (e.g., Burgess & Hitch, 2006; Farrell, 2012; Logan, 2018; Page & Norris, 2009). The effects of segmentation on gains and losses add to the long list of effects of musical structure on music practice, performance, and appreciation (respectively, Williamon & Valentine, 2002; Demos et al., 2020; Bisesi & Windsor, 2016). These effects of musical structure are, in turn, examples of the pervasive effect of temporal grouping on both action (Rosenbaum et al., 2001) and experience (Farrell, 2012).

As predicted by PC theory (Chaffin & Imreh, 2002), PCs provided points of recovery where recall continued after mistakes and omissions. Initially, the singer was able to write out her part from memory with only trivial errors. When she first began to make mistakes, 14 months after the performance, she recovered quickly, leaving only small gaps in her written recall. As the singer's memory for the piece faded over time, the gaps in her recall lengthened. She increasingly forgot phrases that did not start with a PC, resulting in losses at phrase starts without a PC. This lengthened gaps between forgetting and recovery because points of recovery became fewer and further between. Thus, recovery slowed as memory faded.

This supports the main conclusion of the earlier studies of the Presto and Prelude, with evidence from gains and losses rather than errors: Musicians' thoughts about the music during performance (PCs) serve as retrieval cues that help to elicit upcoming passages from long-term memory (Chaffin & Imreh, 2002; Chaffin et al., 2010; Lisboa et al., 2018). The Presto and Prelude studies did not examine gains and losses directly. Instead, they examined error rates and found primacy effects in phrases that started with a PC. From this, the authors inferred that recoveries (gains) at phrase starts with PCs were followed by forgetting (losses) at non-structural locations later in the phrase. In contrast, we identified recovery and forgetting more directly as gains and losses. Consistent with the earlier studies, we found that PCs promoted recovery at the start of a new segment, at phrase starts with PCs, and at section starts, most of which also started with a PC.

Our study provides a new understanding of the relationship between PCs and musical structure. The discovery that PCs on non-structural beats did not elicit recoveries (gains) in the same way as PCs on structural beats requires modification of the implied claim of PC theory that *all* PCs function as retrieval cues (Chaffin & Imreh, 20002). In our study, PCs functioned as retrieval cues only when they occurred on structural beats, not on non-structural beats. This may be because the singer repeated structural locations more often during practice, or because she used them as starting and stopping places, as previously reported by Ginsborg and Chaffin (2011; also see Chaffin & Imreh, 2002; Chaffin et al., 2010). Together, the practice and recall data support the claim of PC theory that PCs become retrieval cues when a performer repeatedly thinks about a particular location in the music during practice; this establishes a link between thought and action so that, later, thinking of the location elicits playing or, in this case, singing (Chaffin et al., 2010).

Additional studies are needed to determine the generality of our findings across performers, instruments, musical styles, types of performance, and levels of expertise. Two of the present findings seem likely to generalize because they are consistent with previous studies. First, the main finding of our study, that gains occurred on structural beats with PCs, may generalize because it is consistent with previous studies reporting primacy effects at PCs (Chaffin & Imreh, 2002; Chaffin et al., 2010) and at structural boundaries (Finney & Palmer, 2003; Mishra, 2010); and, more broadly, with the well documented benefits of elaboration on many kinds of memory (Schacter & Graf 1986), including memory for performance (Lisboa et al., 2015; Timperman & Miksza, 2019). Second, the primacy effect at the top level of musical structure (sections within the piece) is similar to primacy effects reported for many types of material (Brown et al., 2007), including music performance (Crowder & Greene, 2010; Finney & Palmer, 2003; Timperman & Miksza, 2019).[13] Previous studies provide no information about the generality of our third finding, that recoveries did not occur at PCs in non-structural locations, because, unlike the present study, there were no expressive/interpretive PCs in non-structural locations in the Presto and Prelude studies.

Other features of our data that differed from previous studies may be less general. First, the low error rate (0.36 after five years) and the absence of an increase in losses over time may, in part, be products of the multiple recall events and/or the singer's role as a researcher in our study. Second, unlike Mishra (2010) and the Presto and Prelude studies (Chaffin & Imreh, 2002; Chaffin et al., 2010), we did not find primacy effects at lower levels of musical structure. Instead, we found essentially no forgetting on non-structural beats within a phrase; the probability of forgetting was constant across serial positions for words within phrases and phrases within sections. The reason for the difference is unclear. The presence of song lyrics in our study may have provided additional constraints during recall that strengthened item-to-item associations within phrases. Or, the difference could be due to one of the many other ways in which in the present study differed from the earlier studies in experimental procedure (length of preparation, multiple recall events, and active participation of the singer as a researcher,) and type of music (composer, musical style, and musical forces).

A third difference from the Presto and Prelude studies is that we found no evidence that basic PCs affected recall in the opposite direction to other types of PC. In our study, there were only three basic PCs on structural beats and they appeared to elicit gains in the same way as other

---

[13] The Presto and Prelude studies did not examine serial position at this top level of music structure.

types of PC. In contrast, in the earlier studies, errors increased at locations with basic PCs (Chaffin & Imreh, 2002; Chaffin et al., 2010). A likely explanation for the difference is that, unlike the instrumentalists in the earlier studies, the singer had her instrument available as she wrote out her part. She hummed and conducted as she wrote, as she often did during practice, thereby generating some of the sensory features to which basic PCs refer. Thus, basic PCs may have been more helpful to the singer in the present study than to the instrumentalists in the earlier studies, who did not have their instruments available when writing out the score (Chaffin & Imreh, 2002; cf. Godden & Baddeley, 1975). Future studies could explore this by comparing singers with instrumentalists, and written with played recall, examining gains and losses, in addition to errors, as in the present study.

In conclusion, our study shows how memory for performance depends on recovery from mistakes and omissions. When recall failed, the singer was able to recover at the start of a new phrase or section where a PC provided a content-addressable retrieval cue to provide renewed access to the memorized sequence of the performance. Thus, the singer's thoughts about musical goals during performance (PCs) were the key to her ability to remember the piece. We suggest that PCs play a similar role during performance. When something goes wrong on stage, experienced performers recover at the start of a new section or phrase where there is a PC.

Acknowledgements

We thank Mary Crawford, Dakota Lindsey, Ian Neath, and an anonymous reviewer for their helpful comments, and Andrea Schiavio who fulfilled the role of Acting Editor-in-Chief for this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Roger Chaffin https://orcid.org/0000-0001-7627-8787
Jane Ginsborg https://orcid.org/0000-0003-0297-1751
Alexander P. Demos https://orcid.org/0000-0001-7964-7808

Supplemental material

Supplemental material for this article is available online.

References

Anderson, J.R. (1983). *The architecture of cognition*. Harvard University Press.

Bisesi, E., & Windsor, W. L. (2016). Expression and communication of structure in music performance: Measurements and models. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed.; pp. 359–371). Oxford University Press.

Bonini, L., Serventi, F. U., Simone, L., Rozzi, S., Ferrari, P. F., & Fogassi, L. (2011). Grasping neurons of monkey parietal and premotor cortices encode action goals at distinct levels of abstraction during complex action sequences. *Journal of Neuroscience*, *31*, 5876–5886. https://doi.org/10.1523/JNEUROSCI.5186-10.2011

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H. J., Maechler, M., & Bolker, B.M. (2017). glmmTMB balances speed and flexibility among packages for zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, *9*(2), 378–400. https://doi.org/10.3929/ethz-b-000240890

Brown, G.D.A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*, 539–576. https://doi.org/10.1037/0033-295X.114.3.539

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*(4), 627–652. doi:10.1016/j.jml.2006.08.005

Chaffin, R. (2007). Learning Clair de Lune: Retrieval practice and expert memorization. *Music Perception*, *24*, 377–393. https://doi.org/10.1525/mp.2007.24.4.377

Chaffin, R., & Imreh, G. (2002). Practicing perfection: Piano performance as expert memory, *Psychological Science*, *13*, 342–349. https://doi.org/10.1111/j.0956-7976.2002.00462.x

Chaffin, R., Demos, A.P., & Logan, T. (2016). Performing from memory. In S. Hallam, I. Cross, & M. Thaut (Eds.), The Oxford Handbook of Music Psychology, 2nd Ed. (pp. 359–371). Oxford University Press.

Chaffin, R., Imreh, G., & Crawford, M. (2002). *Practicing perfection: Memory and piano performance*. Lawrence Erlbaum Associates.

Chaffin, R., Lisboa, T., Logan, T., & Begosh, K.T. (2010). Preparing for memorized cello performance: The role of performance cues. *Psychology of Music*, *38*, 3–30. https://doi.org/10.1177/0305735608100377

Christensen, W., Sutton, J., & McIlwain, D. J. (2016). Cognition in skilled action: Meshed control and the varieties of skill experience. *Mind & Language*, *31*(1), 37–66. http://doi.org/10.1111/mila.12094

Chueke, Z., & Chaffin, R. (2016). Performance cues for music 'with no plan': a case study of preparing Schoenberg's Op. 11, No. 3. In C. Mackie (Ed.), *New thoughts on piano performance* (pp. 255–268). London International Piano Symposium.

Cook, N. (2013). *Beyond the score: Music as performance*. Oxford University Press.

Crowder, R. G., & Greene, R. L. (2010). Serial learning. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford handbook of memory* (pp. 125–136). Oxford University Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321. https://doi.org/10.1037/0033-295X.93.3.283

Demos, A.P., & Chaffin, R. (2017). Removing obstacles to the analysis of movement in musical performance: Recurrence, mixed models, and surrogates. In M. Lesaffre, P-J. Maes, & M. Leman (Eds). *The Routledge Companion to Embodied Music Interaction* (pp. 341–349). Routledge.

Demos, A.P., Lisboa, T, Logan, T., Begosh, K. T., & Chaffin, R. (2020). A longitudinal study of the development of expressive timing. *Psychology of Music*, *48*, 50–66. https://doi.org/10.1177/0305735618783563

Diedrichsen, J., & Kornysheva, K. (2015). Motor skill learning between selection and execution. *Trends in Cognitive Sciences*, *19*, 227–233. http://dx.doi.org/10.1016/j.tics.2015.02.003

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). Teacher's College, Columbia University.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. https://doi.org/10.1037/0033-295X.102.2.211

Ettlinger, M., Margulis, E., & Wong, P. (2011). Implicit memory in music and language. *Frontiers in Psychology*, *2*, 211. https://www.frontiersin.org/article/10.3389/fpsyg.2011.00211

Farrell, S. (2012). Temporal clustering and sequencing in working memory and episodic memory. *Psychological Review*, *119*, 223–271. https://doi.org/10.1037/a0027371

Finney, S. A., & Palmer, C. (2003). Auditory feedback and memory for music performance: Sound evidence for an encoding effect. *Memory and Cognition*, *31*(1), 51–64. https://doi.org/10.3758/BF03196082

Fox, J. (2003). Effect displays in R for Generalised Linear Models. *Journal of Statistical Software*, *8*(15), 1–27. http://www.jstatsoft.org/v08/i15/

Ginsborg, J. (2009). Beating time: The role of kinaesthetic learning in the development of mental representations for music. In A. Mornell (Ed.), *Art in Motion* (pp. 121–142). Vienna: Peter Lang.

Ginsborg, J., & Chaffin, R. (2009). Very long term memory for words and melody: An expert singer's written and sung recall over six year. In C. Stevens, B. Kruitof, K. Buckley, & S. Fazio (Eds.), *Proceedings of the 2nd International Conference on Music Communication Science* (ICoMCS2; pp 24–28) , HCSNet, University of Western Sydney, Sydney, Australia

Ginsborg, J., & Chaffin, R. (2011). Performance cues in singing: evidence from practice and recall. In I. Deliège & J. Davidson (Eds.), *Music and the mind: Investigating the functions and processes of music (a book in honour of John Sloboda)* (pp. 339–360). Oxford: Oxford University Press.

Ginsborg, J., & Sloboda, J. A. (2007). Singers' recall for the words and melody of a new, unaccompanied song. *Psychology of Music*, *35*(3), 421–440. https://doi.org/10.1177/0305735607072654

Ginsborg, J., Chaffin, R., & Demos, A.P. (2012). Different roles for prepared and spontaneous thoughts: A practice-based study of musical performance from memory. *Journal of Interdisciplinary Music Studies*, *6*, 201–231. http://doi.org/10.4407/jims.2014.02.005

Ginsborg, J., Chaffin, R., & Nicholson, G. (2006). Shared performance cues in singing and conducting: a content analysis of talk during practice. *Psychology of Music*, *34*(3), 167–194. https://doi.org/10.1177/0305735606061851

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of Psychology*, *66*(3), 325–331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

Halpern, A. R., & Bower, G. H. (1982). Musical expertise and melodic structure in memory for musical notation. *The American Journal of Psychology*, 31–50. http://dx.doi.org.ezproxy.lib.uconn.edu/10.2307/1422658

Lewandowsky, S., & Murdock, B.B.Jr. (1989). Memory for serial order. *Psychological Review*, *96*, 25–57. https://doi.org/10.1037/0033-295X.96.1.25

Lindsey, D. R. B., & Logan, G. D. (2019). Item-to-item associations in typing: Evidence from spin list sequence learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*, 397–416. http://dx.doi.org/10.1037/xlm0000605

Lindsey, D. R. B., & Logan, G. D. (in press). Previously retrieved items contribute to memory for serial order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Lisboa, T., Chaffin, R., & Demos, A.P. (2015). Recording thoughts while memorizing music: a case study. *Frontiers in Psychology*. 5:1561. https://doi.org/10.3389/fpsyg.2014.01561

Lisboa, T., Chaffin, R., & Logan, T.R. (2012). An account of deliberate practice: Thoughts, behaviour and the self in learning Bach's Prelude 6 for cello solo. In A. Cervino, M. Lettberg, C. Laws, & T. Lisboa (Eds.) *Practice of Practising* (pp. 9–31). Orpheus Research Centre in Music.

Lisboa, T., Demos, A.P., & Chaffin, R. (2018). Training thoughts and actions for virtuoso performance. *Musicae Scientiae*, *22*, 519–538. https://doi.org/10.1177/1029864918782350

Logan, G. D. (2018). Automatic control: How experts act without thinking. *Psychological Review*, *125*, 453–485. http://dx.doi.org/10.1037/rev0000100

Logan, G. D. (2021). Serial order in perception, memory, and action. *Psychological Review*, *128*(1), 1. http://dx.doi.org/10.1037/rev0000253

Lüdecke, D. (2020). sjPlot: Data Visualization for statistics in social science. https://doi.org/10.5281/zenodo.1308157), R package version 2.8.3, https://CRAN.R-project.org/package=sjPlot>.

Mishra, J. (2010). Effects of structure and serial position on memory errors in musical performance. *Psychology of Music*, *38*(4), 447–461. https://doi.org/10.1177/0305735609351919

Miller, G. A., Galanter, E., & Pribram, K. H. (1960). *Plans and the structure of behavior*. Henry Holt and Co. https://doi.org/10.1037/10039-000

Page, M. P. A., & Norris, D. (2009). A model linking immediate serial recall, the Hebb repetition effect and the learning of phonological word forms. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3737–3753.

Palmer, C., & Pfordresher, P. Q. (2003). Incremental planning in sequence production. *Psychological Review*, *110*, 683–712. https://doi:10.1037/0033-295X.110.4.683

Pfordresher, P. Q., Palmer, C., & Jungers, M. K. (2007). Speed, accuracy, and serial order in sequence production. *Cognitive Science*, *31*, 63–98. https://doi.org/10.1080/03640210709336985

Rosenbaum, D. A., Meulenbroek, R. J., Vaughan, J., & Jansen, C. (2001). Posture-based motion planning: Applications to grasping. *Psychological Review*, *108*, 709–734. http://dx.doi.org/10.1037/0033-295X.108.4.709

Rubin, D.C. (2006). The basic-system model of episodic memory. *Perspectives on Psychological Science*, *1*, 277–311. https://doi.org/10.1111/j.1745-6916.2006.00017.x

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 501–518. https://doi.org/10.1037/0278-7393.13.3.501

Schacter, D. L., & Graf, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(3), 432–444. http://dx.doi.org/10.1037/0278-7393.12.3.432

Singer, J. D., & Willett, J. B. (2003). *Applied Longitudinal Data Analysis: Methods for Studying Change and Event Occurrence*. Oxford University Press.

Timperman, E., & Miksza, P. (2019). Verbalization and musical memory in string players. *Musicae Scientiae*, *23*(2), 212–230. https://doi.org/10.1177/1029864917727332

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Williamon, A., & Valentine, E. (2002). The role of retrieval structures in memorizing music. *Cognitive Psychology*, *44*, 1–32. https://doi.org/10.1006/cogp.2001.0759