Title
Haunting melodies: Specific memories distort beat perception

Author
Creel, Sarah C

Publication Date
2022-08-01

DOI
10.1016/j.cognition.2022.105158

Peer reviewed

Haunting melodies: specific memories distort beat perception

Sarah C. Creel

screel@ucsd.edu

Department of Cognitive Science

UC San Diego

9500 Gilman Drive Mail Code 0515

La Jolla CA 92093

2997 words (abstract + main text)

Abstract

How much does specific previous experience shape immediate perception? Top-down perceptual inference occurs in ambiguous situations. However, similarity-based accounts such as exemplar theory suggest that similar memories resonate with the percept, predicting that detailed previous experiences can shape perception even when bottom-up cues are unambiguous. The current study tests whether specific musical memories influence beat perception only under ambiguity, or more pervasively—that is, even when clear bottom-up beat cues are present. Listeners were exposed to 16 melodies, half in one meter, half in another. Later, each listener's perception of a specific melody's meter was tested when that melody occurred in either its original meter or another meter. Ratings of metrical probes were influenced not only by fit with the current (test) meter, but also by fit with the meter previously experienced with that melody. Findings suggest that perception is routinely influenced by detailed top-down perceptual imagery.

Keywords: exemplar theory, perceptual imagery, noisy inference, beat perception, music

To what extent are perception and processing controlled by generic memory representations, vs. veridical, highly-detailed memory representations? This answer to this question is critical to understanding of how perception takes place. A wealth of findings in domains of inquiry such as language processing (DeLong, Urbach, & Kutas, 2005; Kamide, Altmann, & Haywood, 2003; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) and motion processing (Saygin, Driver, & de Sa, 2008; Saygin, Ishiguro, Chaminade, Driver, & Frith, 2012; Shiffrar & Freyd, 1990) explicitly acknowledge the role of detailed knowledge in prediction. Relatedly, detailed memories can bias ambiguous percepts. For example, viewers seeing a grayscale picture of a banana perceive it as slightly yellow (Hansen, Olkkonen, Walter, & Gegenfurtner, 2006); color is modulated by perceived light sources (Purves, Lotto, & Nundy, 2002); listeners illusorily perceive a missing phoneme that has been replaced by noise (Samuel, 1981; Warren, 1970); beat perception for ambiguous-meter melodies is shaped by specific details of recently-heard musical memories (Creel, 2011, 2012, 2020).

Like prediction, perceptual inference depends on memory: the perceiver must activate preexisting knowledge to infer the underlying structure of a noisy perceptual signal. Yet it is not known to what extent, or under what circumstances, *specific* memories control perceptual inference. This theoretical gap is particularly large in cases where the bottom-up input is high-fidelity: if bottom-up input is strong enough, perhaps top-down information is not needed, suggesting that perceptual inference has effects primarily in high-noise situations, rather than pervading all of perceptual processing.

The pervasiveness of detailed memory activation in perception has not been explored extensively, particularly in music perception. While recent findings suggest that listeners form highly-specific auditory memories (Gjerdingen & Perrott, 2008; Krumhansl 2010; Palmer, Jungers, & Jusczyk, 2001; Schellenberg, Iverson, & McKinnon, 1999), no current account *links* highly-detailed memories to processing of critical structural aspects of music like timing. The current study tests how strongly detailed memories shape processing, through the lens of musical beat detection.

Advantages of storing specific memories

The work here favors a theoretical view of perception as a process of similarity-based memory activation. One set of approaches within this view construes memory as a collection of experienced instances or *exemplars* (Goldinger, 1998; Hintzman 1986; Pierrehumbert, 2001). On an exemplar account, perception itself is a process of obligatorily activating previously-learned instances to a degree proportional to their similarity to the input. The summed activation of exemplars might be thought of as simulation, auditory imagery (see Janata, 2012), or, in Hintzman's (1986) terms, a collective "echo" of experience. This detailed echo shapes the inferred percept. While the general notion of similarity-based activation is widespread in language processing, particularly word recognition (e.g., Goldinger, 1996; McClelland & Elman, 1986), it is less widespread in music perception. Further, even within word recognition, some aspects of specific memories, such as talker-specific detail, have not been deemed relevant by classic similarity-based models (McClelland & Elman, 1986). However, more recent evidence indicates talker-specific detail can impact real-time word recognition (e.g., Creel, Aslin, & Tanenhaus, 2008; Creel & Tumlin, 2011; Kapnoula & Samuel, 2019).

Specific memory effects on perceptual inference are most evident when the input is ambiguous. This leaves open the possibility that top-down knowledge supports perceptual inference only weakly, trading off to bottom-up input when the latter is relatively noise-free. Yet, if detailed representations are *routinely* activated during listening, then they should be active even when bottom-up perceptual information is strong.

Some recent models of music perception (see Huron, 2006; Huron & Margulis, 2010; Pearce & Wiggins, 2012) begin to account for specific pitch sequence knowledge in pitch prediction. However, current models of beat-finding and meter perception are heavily bottom-up, driven by event onsets (Large, Herrera, & Velasco, 2015; Scheirer, 1998; Tomic & Janata, 2008). That is, they take as input features of the current auditory signal and compare those features to representations of periodicities (e.g., comb filters, Scheirer, 1998; damped linear oscillators, Large et al., 2015, Tomic & Janata, 2008). Importantly, input is *not* compared to a register of richer musical information, such as specific musical memories. Thus, models currently incorporate highly schematized top-down knowledge, making it impossible for them to account for effects of memory detail. One exception is deep neural net approaches to beat detection, which can incorporate extensive musical training data, though even these may use data-reducing transformations (Schreiber & Müller, 2018).

Models' lack of top-down knowledge is problematic because bottom-up cues are not always sufficient to explain listeners' behavior. For example, Western listeners have difficulty perceiving *complex meters*, where beats are unevenly spaced in time (such as 2:3 duration ratios), while listeners whose cultures have complex meters have no such difficulty (Hannon, Soley, & Ullal, 2012; Hannon & Trehub, 2005a, b; Kalender, Trehub, & Schellenberg, 2013; Snyder, Hannon, Large, & Christiansen, 2006; see also Drake, 1993, and Parncutt, 1994, on binary vs. ternary patterns). This finding suggests that cultural musical knowledge (presence/absence of a meter in one's culture), not just the bottom-up auditory signal, influences meter processing. In addition to general music knowledge, models do not specify any role for memory detail, yet recent work (Creel 2011, 2012, 2020) suggests that detailed memories shape perception of melodies that on their own are metrically ambiguous. However, Creel's work leaves unanswered whether memory detail influences meter processing routinely, even when bottom-up cues are strong, or only influences meter processing in cases of metrical ambiguity.

The current study presents a stronger test of the role of detailed memory in perception by testing whether memory affects beat perception *even when the bottom-up signal clearly favors* *one beat structure or another*. This work builds on recent findings (Creel 2011, 2012, 2020) that beat perception is influenced by specific top-down knowledge when beat cues are ambiguous. If specific top-down knowledge influences beat perception even when bottom-up beat cues are *un*ambiguous, it suggests that activation of specific memories is a routine element of perception.

Method

Participants were 96 individuals from the UCSD human participants pool. This large sample size was suggested by earlier work, where 48 participants and 8 items were sufficient for by-participants but not quite by-items effects, so both participants and items were doubled. Research was approved by the UCSD Human Research Protections Program, and informed consent was obtained. Two additional participants were excluded for failing to complete the experiment.

Stimuli. Melodies from Creel (2011, Experiment 2) and Creel (2020, Experiment 2) were combined to yield 16 melodies. All melodies contained six sub-beats per measure and were ambiguous between 3/4 time—three groups of two sub-beats (repeating pattern of XxXxXx, where capital X's indicate beats and lower-case x's indicate sub-beats, such as the underlying beat pattern of *Happy Birthday*), and 6/8 time—two groups of three sub-beats (repeating pattern of XxxXxx, e.g., *Greensleeves*). Melodies differed in instruments (e.g., flute, violin), keys (C, F#), modes (major, minor, other), and sub-beat duration (see Creel, 2011, Table 5, and Creel, 2020, Table 4 for details).

The Creel (2011) melodies originally were roughly 16 measures long, while the Creel (2020) melodies were roughly 8 measures long. To match the sets, the Creel (2011) melodies were shortened to 8 measures. Music files were exported from Finale (2009, MakeMusic Inc.)

software as .aiff files, which were then were normalized for RMS amplitude and exported as .wav files using Praat 5.4.01 (Boersma & Weenink, 2014).

Table 1. Correlations of note onsets with idealized 34 and 68 metrical patterns. For each pair of columns, the stronger correlation is bolded. In all cases, the contexts and the melody+context combinations correlated more strongly with the intended meter. Melodies 1-8 were drawn from Creel (2020), and Melodies 9-16 were drawn from Creel (2011, Experiment 2).

	Melody		34 context		68 context		Melody + 34 context		Melody + 68 context	
Melody	34	68	34	68	34	68	34	68	34	68
1	**0.59**	0.41	**1.00**	0.59	0.39	**0.96**	**0.95**	0.58	0.56	**0.96**
2	0.26	**0.56**	**1.00**	0.63	0.62	**1.00**	**0.89**	0.74	0.55	**0.94**
3	**0.70**	0.65	**0.91**	0.24	0.37	**0.91**	**0.91**	0.37	0.49	**0.89**
4	**0.75**	0.68	**0.91**	0.26	0.32	**0.93**	**0.91**	0.47	0.55	**0.90**
5	0.57	**0.77**	**1.00**	0.64	0.63	**0.99**	**0.93**	0.76	0.64	**0.96**
6	**0.62**	0.46	**0.89**	0.22	0.31	**0.91**	**0.87**	0.31	0.47	**0.86**
7	**0.70**	0.65	**1.00**	0.56	0.33	**0.92**	**0.98**	0.62	0.47	**0.93**
8	0.80	**0.88**	**0.99**	0.67	0.59	**1.00**	**0.98**	0.70	0.62	**1.00**
9 (a)	0.35	**0.55**	**0.92**	0.28	0.37	**0.95**	**0.90**	0.47	0.40	**0.90**
10 (b)	0.51	**0.66**	**0.92**	0.31	0.47	**0.98**	**0.89**	0.51	0.53	**0.97**
11 (d)	**0.59**	0.43	**0.98**	0.57	0.59	**1.00**	**0.98**	0.58	0.62	**0.99**
12 (e)	**0.70**	0.34	**0.98**	0.45	0.37	**0.95**	**0.99**	0.45	0.53	**1.00**
13 (f)	0.56	**0.76**	**0.89**	0.22	0.31	**0.92**	**0.93**	0.48	0.44	**0.95**
14 (g)	0.61	**0.84**	**0.93**	0.30	0.38	**0.95**	**0.96**	0.55	0.48	**0.96**
15 (h)	**0.83**	0.64	**0.95**	0.70	0.47	**0.93**	**0.94**	0.70	0.58	**0.91**
16 (i)	**0.65**	0.44	**0.80**	0.18	0.40	**0.96**	**0.94**	0.34	0.59	**0.94**
M	0.61	0.61	**0.94**	0.42	0.43	**0.95**	**0.93**	0.54	0.53	**0.94**
SD	0.15	0.16	0.06	0.19	0.11	0.03	0.04	0.14	0.07	0.04

To verify that contexts and melody+context combinations indicated each meter unambiguously, onset times of notes in melodies and contexts were correlated with idealized 3/4 and 6/8 patterns (2 0 1 0 1 0 and 2 0 0 1 0 0 respectively). For each voice (melody, context upper voice, context lower voice) a value of 1 was assigned if there was an onset at that time, and a 0 otherwise. Sixteenth notes (1/12 of measure) were not counted. Values summed across measures and voices were then correlated with idealized meter patterns. Correlations (Table 1) indicated that the intended meter for a given context had a stronger correlation than the other meter in every case.

Figure 1. Example of musical contexts presented. A listener might hear the top context (3/4; second and third staves) playing along with this melody during exposure, but might hear both of the top and bottom contexts (3/4, second and third staves, 6/8, fifth and sixth staves) on different trials during testing.

Procedure. Before the main experiment, listeners completed brief demographic and music experience questionnaires. The main experiment had two phases: exposure and test, as in Creel (2011, 2012, 2020). During exposure, each listener heard all 16 melodies: 8 melodies in a metrically clear 3/4 musical context, 8 in a metrically clear 6/8 musical context (examples in Figure 1). Each melody was heard once per 16-trial subblock x 8 subblocks = 128 exposure trials. Trials in a subblock were randomly ordered. A given listener heard a particular melody in a single meter during exposure. Across participants, 8 different lists ensured that each melody occurred equally often in each meter. For example, one participant's exposure phase might include Melody 1 in 3/4 and Melody 9 in 6/8, while another participant heard Melody 1 in 6/8 and Melody 9 in 3/4. Thus, no metrical properties of any melody itself could drive overall results.

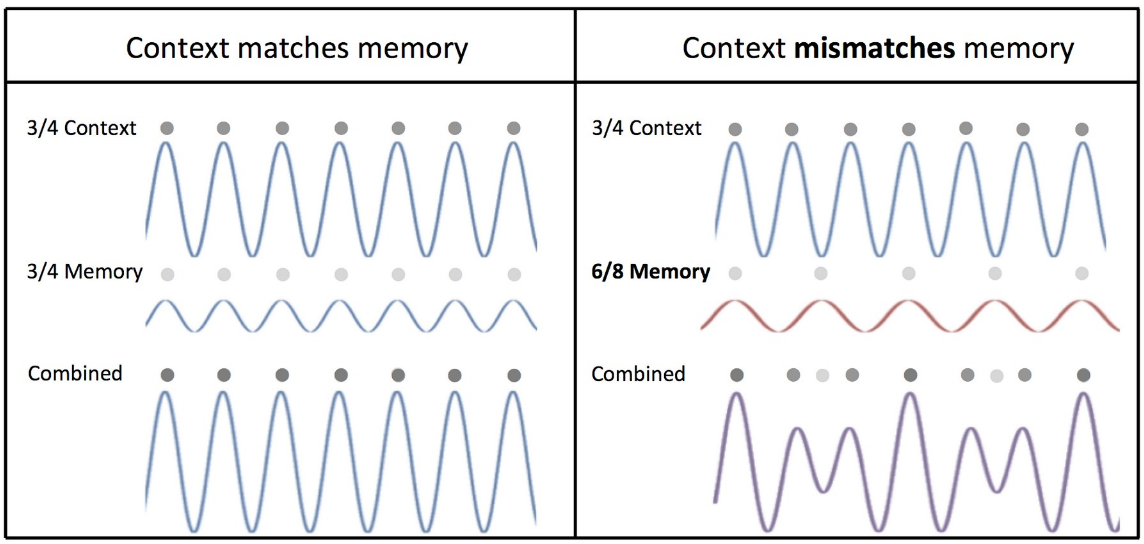

Figure 2. On an exemplar account, hearing a specific melody reactivates information associated with that melody, including metrical information ("Memory," middle line). When hearing clear metrical information with a particular melody ("Context," top line), the reactivated metrical information combines with it to yield the metrical percept ("Combined," bottom line). When the current (context) meter and the memory meter match, they reinforce each other (left side). When the current (context) meter and the memory meter mismatch, they interfere with each other (right side). Filled circles indicate beats; darkness of circles indicates perceptual strength.

On each exposure trial, listeners rated perceived affect (happy or sad) and their preference (like or dislike) by clicking in a square with two axes, affect and liking. This cover task aimed to maintain participant attention without alerting them to the upcoming test or its contents, so ratings were not examined.

The test consisted of two 16-trial blocks (total 32 trials), with a brief, self-paced break between blocks. Each block of 16 presented each melody once; melodies within a block were presented in random order. On each trial, a listener heard a melody in a clear metrical context—either the one they had heard at exposure, or the other one. Right after the melody, a series of probe drumbeats played in a woodblock timbre, either in 3/4 time or in 6/8 time. Listeners then clicked on a graphical ruler to rate how well the probe beats fit with the preceding music. The ruler was 384 pixels long, with the label "Bad fit" at the left end, "Okay" in the center, and "Good fit" at the right end. Ratings were quantified as the x-coordinate clicked on the ruler (range 0-384). For display purposes this value was scaled using the formula (rating – 192)/192, so that it ranged from -1 (bad fit) to +1 (good fit), with 0 indicating the center of the ruler.

For half of test trials, the probe meter matched the test metrical context, and for the other half, it mismatched. Crossed with this, the probe equally often matched or mismatched the *exposure* meter—the context heard during exposure with a given melody. Following the test, listeners completed a brief questionnaire about the experiment.

Predictions. Figure 2 outlines a general account of how activated, specific memories might shape meter perception by combining the bottom-up beat percept with beat features from top-down memory. If listeners are not influenced by specific musical memories (from the exposure phase), then goodness-of-fit ratings should be determined completely by the test metrical context—whether the probe meter matches the context heard with the melody at test (Figure 3, left). However, if listeners *are* influenced by musical memories, then goodness-of-fit ratings should be influenced by the exposure meter (Figure 3, center) or, more likely, by both the test meter and the exposure meter (right side of Figure 3), with the highest probe ratings when the probe matches both test meter and exposure meter, lower when it matches only one, and lowest it mismatches both. A specific example (Figure 4): suppose a listener previously heard Melody 10 during exposure with a 3/4 meter context. At test, if they hear Melody 10 with a 3/4 context, a 3/4 probe will sound very good, and a 6/8 probe will sound quite poor. If they hear Melody 10 with a 6/8 meter context, a 3/4 probe will sound less bad despite mismatching the immediate context, and a 6/8 probe will sound less good despite matching the immediate context.

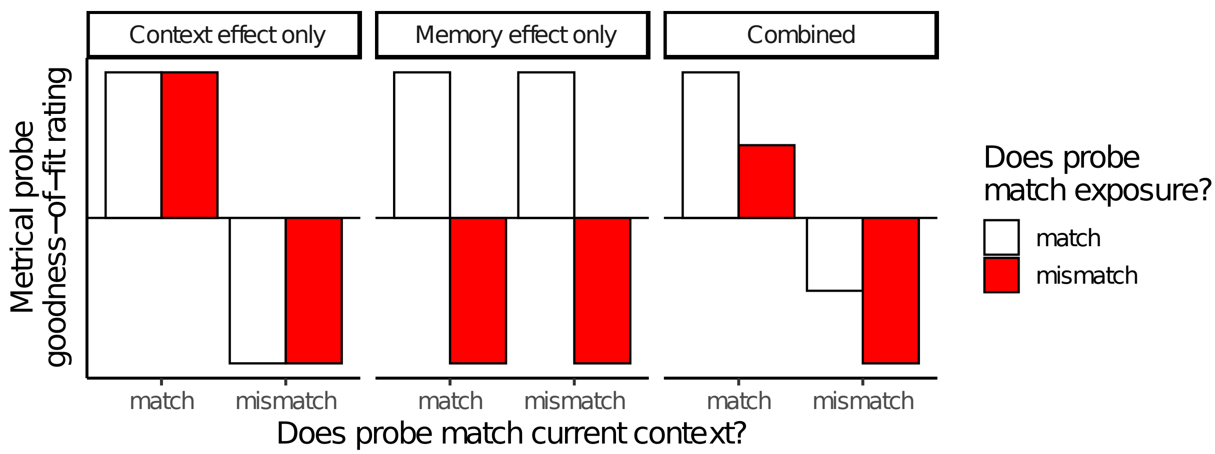

Figure 3. Predictions of goodness-of-fit ratings of metrical probes, depending on whether or not the metrical context heard during exposure affects meter perception during the test phase. Left: if previous exposure to a particular meter does not matter when a clear meter is heard during the test phase, main effect of context match to probe; center: predictions if previous exposure to a particular meter completely overrides a clear meter heard during the test phase, main effect of exposure match to probe (unlikely); right: predictions if previous exposure to a particular meter affects meter perception even when hearing a clear meter during the test phase, main effects of both context match to probe and exposure match to probe.

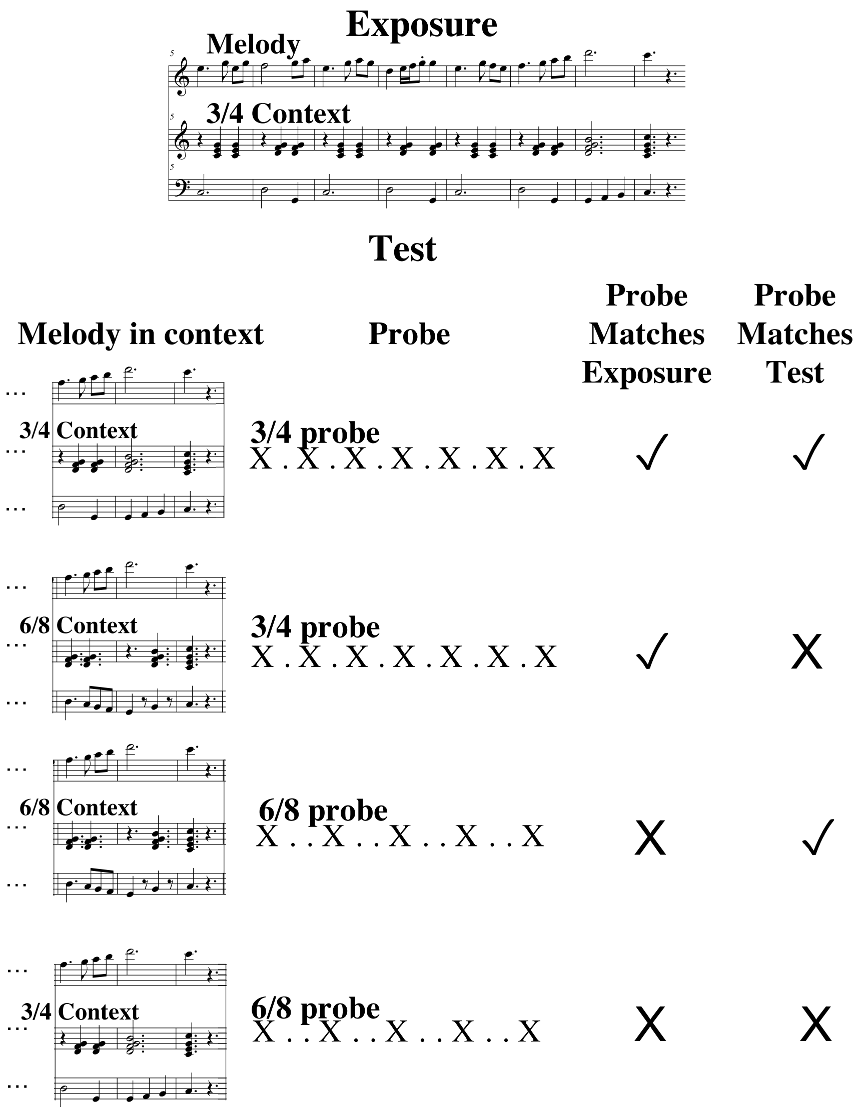

Figure 4. Example melody (Melody 10) with 3/4 exposure context and associated test trials for different combinations of test metrical contexts and probe meters.

Results

Ratings were averaged over participants and over items and were entered into by-subjects and by-items analyses of variance (ANOVAs). Factors were Test-Probe Match (does probe meter *match* or *mismatch* test metrical context), Exposure-Probe Match (does probe meter *match* or *mismatch* previously experienced metrical context), and Probe Meter (3/4, 6/8). All factors were within-subjects and within-items. If exposure does not affect goodness-of-fit ratings, but immediate metrical context does, there should only be a main effect of Test-Probe Match. However, if listeners are reactivating the exposure meter from memory, there should also be a main effect of Exposure-Probe Match.

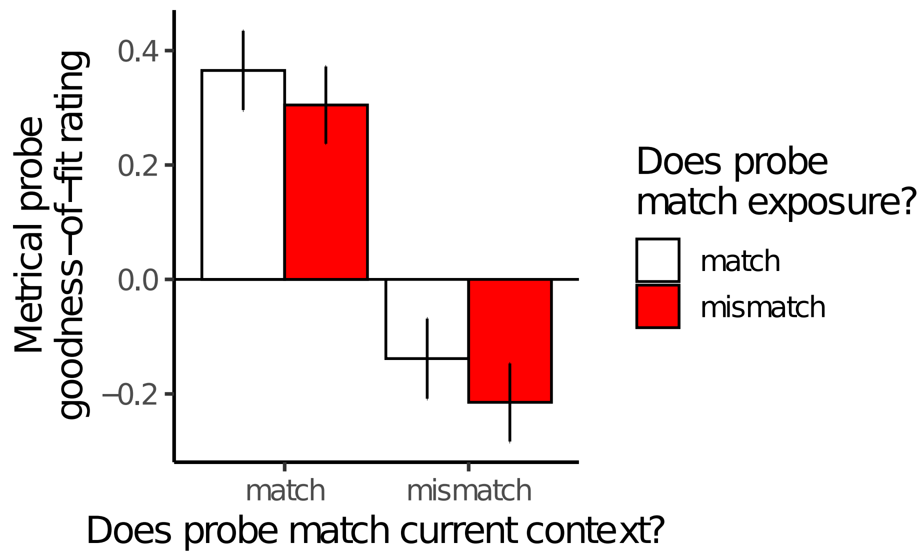

Figure 5. Metrical probe goodness-of-fit ratings (with standard errors) by the probe's metrical match to the test context and the probe's match to the exposure context.

The observed effects were as follows. An effect of Test-Probe Match ($F1(1,95) = 194.1$, $p < .0001$; $F2(1,15) = 100.9$, $p < .0001$; $\eta^2_P = .67$) indicated higher ratings when the test and probe meters coincided than when they mismatched (Figure 5, left bars vs. right bars). This large effect suggests the metrical contexts effectively indicated the intended meters. The effect of Exposure-Probe Match was also significant ($F1(1,95) = 12.98$, $p = .0005$; $F2(1,15) = 9.41$, $p = .008$; $\eta^2_P = .12$), such that probe ratings were higher when the probe meter matched the exposure meter

(Figure 5, light bars) than when the probe meter mismatched the exposure meter (dark bars). No other effects or interactions approached significance.

One might ask whether effects represented a subset of participants guessing the hypothesis and deliberately responding based on songs' initial meters. This seems unlikely for three reasons. First, participants did not know there would be a test, much less what they would be tested on, until after the exposure phase. Second, only three participants in the post-experiment questionnaire indicated any awareness of the critical manipulation: a change in meter relative to the exposure phase. One participant noted that this awareness started when they arrived at the test phase; none appeared to have guessed the meter manipulation at the outset. Third, when asked if they knew which time signatures were present (that is, what the notated musical meters would be), most participants replied that they did not know or had not been paying attention. Of the participants reporting anything construable as time signatures at all (e.g., "2/4," "4/6," "waltz," "renaissance,"; 20/94), only 9 indicated at least two time signatures, the minimum needed for using explicit labels to differentially encode melodies' meters. Importantly, when these participants were omitted from analyses, the Exposure-Probe Match effect was still significant. That is, results were still consistent with reactivation of metrical information from specific memories even when restricting analyses to listeners who could not have differentially verbally labeled melodies' meters.

Discussion

This study finds effects of specific, recent prior listening experience on beat perception even in the presence of strong bottom-up beat cues. This suggests that listeners' musical percepts are shaped by similarity-based activation of specific memory traces. Further, this effect does not appear to result from deliberate encoding or strategic memorization of meters. Future work should examine on-line measures of beat perception such as EEG or tapping to verify that similarity-based activation takes place in real time.

The study makes several novel contributions. Principally, it suggests that perception is influenced by memory reactivation even in the presence of clear bottom-up input. This lends support to the principle that similar memories are obligatorily reactivated during processing and shape perception, as similarity-based activation models predict. Second, findings unify recent demonstrations of detailed musical knowledge (Gjerdingen & Perrott, 2008; Krumhansl, 2010; Palmer et al., 2001; Schellenberg et al., 1999) and musical prediction via detailed memory activation (Huron, 2006; Huron & Margulis, 2010; Pearce & Wiggins, 2012) with listeners' perceptual experiences. That is, current results suggest that detailed musical knowledge subserves not just rapid recognition of songs (Schellenberg et al., 1999) or genres (Gjerdingen & Perrott, 2008; Krumhansl, 2010), but also *perception* (of meters; the current study). Third, given that detailed musical knowledge materially affects beat processing, the current study suggests that models of music processing and beat processing in particular (e.g., Large et al., 2015; Tomic & Janata, 2008) need to be modified in order to account for effects of detailed memory. Even for newer deep neural network models with detailed training data, the current results would not be reproducible without incorporating the particular musical materials used here, and perhaps giving them greater weight than other musical material.

More broadly, current findings raise the possibility that seemingly unambiguous perceptual information is routinely modified by prior perceptual experience. Some accounts view all of perception as a process of inference under uncertainty (e.g., Kleinschmidt & Jaeger, 2015; Purves et al., 2002), invoking a role for knowledge in perception. Sources of uncertainty include external noise, neural noise, attentional competition from sound sources, speech production errors, and expressive timing variations. Given so many noise sources, listeners should not depend completely on surface perceptual input. This perspective, along with the current data, is consistent with detailed top-down perceptual knowledge affecting perception in a wide variety of situations.

Conclusion

Recent, previous musical experiences influenced judgments of musical beat even when bottom-up cues clearly indicated a beat. This suggests that detailed memories shape perception even when bottom-up cues are strong, implying that coactivation of specific memories materially shapes perception. This provides evidence that memory-based perceptual imagery, rather than trading off with bottom-up input, is a routine element of recognition, prediction, and processing, consistent with similarity-based activation models of memory representation.

References

Boersma, P. & Weenink, D. (2014). Praat: doing phonetics by computer [Computer program]. Version 5.4.01, retrieved 9 November 2014 from http://www.praat.org/

Creel, S. C. (2011). Specific previous experience affects perception of harmony and meter. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1512–1526. doi:10.1037/a0023507

Creel, S. C. (2012). Similarity-based restoration of metrical information: Different listening experiences result in different perceptual inferences. *Cognitive Psychology*, *65*(2), 321–51. doi:10.1016/j.cogpsych.2012.04.004

Creel, S. C. (2020). Metrical restoration from local and global melodic cues: Rhythmic patterns and overall melodic form. *Music Perception, 38*(2), 106–135.

Creel, S. C., Aslin, R. N., & Tanenhaus, M. K. (2008). Heeding the voice of experience: The role of talker variation in lexical access. *Cognition, 106*, 633–664. https://doi.org/10.1016/j.cognition.2007.03.013

Creel, S. C., & Tumlin, M. A. (2011). On-line acoustic and semantic interpretation of talker information. *Journal of Memory and Language, 65*(3), 264–285. https://doi.org/10.1016/j.jml.2011.06.005

Delong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, *8*(8), 1117–1121.

Drake, C. (1993). Reproduction of musical rhythms by children, adult musicians, and adult nonmusicians. *Perception & Psychophysics*, *53*(1), 25–33.

Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The rapid recognition of music genres. *Journal of New Music Research*, *37*(2), 93–100. doi:10.1080/09298210802479268

Goldinger, S. D. (1998). Echoes of echoes?: An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279.

Hannon, E. E., & Trehub, S. E. (2005a). Metrical categories in infancy and adulthood. *Psychological Science*, *16*(1), 48–55.

Hannon, E. E., & Trehub, S. E. (2005b). Tuning in to musical rhythms: Infants learn more readily than adults. *Proceedings of the National Academy of Sciences*, *102*(35), 12639–12643.

Hannon, E. E., Soley, G., & Ullal, S. (2012). Familiarity overrides complexity in rhythm perception: a cross-cultural comparison of American and Turkish listeners. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(3), 543–8. doi:10.1037/a0027225

Hansen, T., Olkkonen, M., Walter, S., & Gegenfurtner, K. R. (2006). Memory modulates color appearance. *Nature Neuroscience*, *9*(11), 1367–1368. doi:10.1038/nn1794

Hintzman, D. L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review*, *93*, 411–428.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. New York: Oxford University Press.

Huron, D. & Margulis, E.H. (2010). Music, expectancy, and thrills. In P.N. Juslin & J.A. Sloboda (Eds.), *Handbook of Music and Emotion: Theory, Research, Applications* (pp. 575-604). New York: Oxford University Press.

Janata, P. (2012). Acuity of mental representations of pitch. *Annals of the New York Academy of Sciences*, *1252*(1), 214–221. doi:10.1111/j.1749-6632.2011.06441.x

Kalender, B., Trehub, S. E., & Schellenberg, E. G. (2013). Cross-cultural differences in meter perception. *Psychological Research*, *77*(2), 196–203. doi:10.1007/s00426-012-0427-y

Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*, 133–156. doi:10.1016/S0749-596X(03)00023-8

Kapnoula, E. C., & Samuel, A. G. (2019). Voices in the mental lexicon: Words carry indexical information that can affect access to their meaning. *Journal of Memory and Language, 107*, 111–127. https://doi.org/10.1016/j.jml.2019.05.001

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.

Krumhansl, C. L. (2010). Plink: "Thin slices" of music. *Music Perception*, *27*(5), 337–354.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, *106*(1), 119–159. doi:10.1037/0033-295X.106.1.119

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18(1), 1–86. http://www.ncbi.nlm.nih.gov/pubmed/3753912

Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language*, *45*(4), 526–545. doi:10.1006/jmla.2000.2780

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception*, *11*(4), 409–464.

Pearce, M. T., & Wiggins, G. A. (2012). Auditory expectation: the information dynamics of music perception and cognition. *Topics in Cognitive Science*, *4*, 625–652. doi:10.1111/j.1756-8765.2012.01214.x

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Typological studies in language 45: Frequency and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Purves, D., Lotto, R. B., & Nundy, S. (2002). Why We See What We Do. *American Scientist*, *90*, 236–243.

Samuel, A. G. (1981). Phonemic restoration: insights from a new methodology. *Journal of Experimental Psychology: General*, *110*(4), 474–94. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2939193

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, *7*, 413–422. doi:10.1093/scan/nsr025

Saygin, A. P., Driver, J., & de Sa, V. R. (2008). In the footsteps of biological motion and multisensory perception: judgments of audiovisual temporal relations are enhanced for upright walkers. *Psychological Science*, *19*(5), 469–75. doi:10.1111/j.1467-9280.2008.02111.x

Scheirer, E. D. (1998). Tempo and beat analysis of acoustic musical signals. *Journal of the Acoustical Society of America*, *103*(1), 588–601. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9440344

Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review*, *6*(4), 641–6. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10682207

Schreiber, H., & Müller, M. (2018). A single-step approach to musical tempo estimation using a convolutional neural network. *Proceedings of the 19th International Society for Music Information Retrieval Conference*, 98–105.

Shiffrar, M., & Freyd, J. (1990). Apparent motion of the human body. *Psychological Science*, *1*(4), 257–264.

Snyder, J. S., Hannon, E. E., Large, E. W., & Christiansen, M. H. (2006). Synchronization and continuation tapping to complex meters. *Music Perception*, *24*(2), 135–148.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., Sedivy, J. C., Series, N., & Jun, N. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*(5217), 1632–1634.

Tomic, S. T., & Janata, P. (2008). Beyond the beat: Modeling metric structure in music. *Journal of the Acoustical Society of America*, *124*(6), 4024–4041. doi:10.1121/1.3006382

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, *167*(3917), 392–393.