The role of working memory in melodic perception

Maegen Walker (maegenw@hawaii.edu)
Department of Psychology, University of Hawai‘i at Mānoa
2530 Dole Street, Honolulu, HI 96822 USA

Ahnate Lim (ahnate@hawaii.edu)
Department of Psychology, University of Hawai‘i at Mānoa
2530 Dole Street, Honolulu, HI 96822 USA

Scott Sinnett (ssinnett@hawaii.edu)
Department of Psychology, University of Hawai‘i at Mānoa
2530 Dole Street, Honolulu, HI 96822 USA

Abstract

We explored the extent to which working memory underpins the processing of relational information in melodies. Using a between subjects design, one group of participants was primed with a melodic stream while performing a concurrent 2-back task while the other group was also primed with the melodic stream but did not perform a concurrent task. Participants were then given a melodic relational categorization task where relations (melodic contour and intervals) could either match or not match the primed melody. Reaction times on the categorization task for primed melodies tended to be faster than for non-primed melodies in the no-task condition, suggesting that relational information in melodies could influence behavior more under conditions where working memory resources were not being used in concomitant tasks. Given the marginal results, more data should be collected to ascertain the full extent to which working memory is involved in the processing of relational melodic content.

Introduction

Melody is one of the most fundamental and salient aspects of music. Simple melodies consist of discrete units or notes, with each note characterized by a pitch, or fundamental frequency (i.e., Hertz value). Pitch is a property of sound related to the rate of vibration that produces the sound, and is characterized by descriptions such as “lowness” or “highness.” The two most common ways in which the pitch sequence of a melody can be encoded are *absolute* and *relative* pitch. Encoding a melody via absolute pitch involves storing the notes according to the fundamental frequencies (i.e., featural aspects) of each pitch, whereas encoding in terms of relative pitch involves storing the melody in terms of the *relations* or intervals (specific frequency differences) between each note.

Relative pitch encoding is considered to be the core strategy humans use to categorize and store familiar melodies (Attneave & Olson, 1971; Page, 1994). For example, the song *Happy Birthday* is immediately recognizable due to the unique intervals between each of the notes regardless of whether or not the song starts on a low or high pitch relative to the key in which the song is traditionally played. There is much evidence on the use of relative pitch information in adults through both behavioral (Dowling, 1978, 1984, 1988) as well as neuroimaging studies (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004; Trainor, McDonald, & Alain, 2002).

In addition to relative pitch, the contour and the intervallic sequence are two other characteristics that can be used to categorize melodies. Contour refers to the general shape, or sequence of up and down frequency shifts, as the melody progresses from note to note, while the intervallic sequence refers to the tonal distance from one pitch to another. For example, a melody with an identical contour to *Happy Birthday*, but with a different intervallic sequence would be perceived as a completely different song though it would still have the same general “shape” (i.e., contour), or up and down pattern (compare A or B to C in Figure 1).

Although the intervallic pattern may be the most overtly salient and representative feature of a melody to humans, studies have shown that, at least for a short duration after being exposed to a melody, human adults are also sensitive to absolute pitch and melodic contour, (Bartlett & Dowling, 1980; Dowling, 1978). Even though the intervallic and contour properties of melodies may characteristically differ in the type of information they carry, what is perhaps more important is that the nature of the information they carry is fundamentally *relational*. Meaning that this type of information depends on the relationship (whether it is the precise intervallic distance or the general contour shape) between each pitch, and not on the actual pitch frequencies themselves. Thus, it is within this relational capacity that melodic perception can be said to share a cornerstone property with many other cognitive processes.

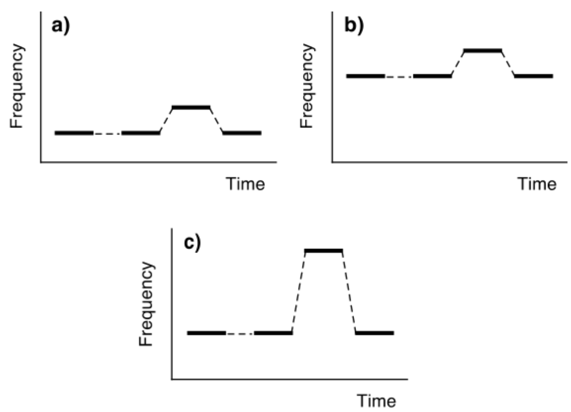

Figure 1: First four notes of Happy Birthday in low (a) and high (b) pitches, and a different melody with similar contour (c).

The ability to explicitly and implicitly process relational properties in stimuli has been proposed as a fundamental mechanism underlying a wide range of cognitive phenomenon. This includes not only higher level reasoning skills such as analogy-making (Gentner, 1983; Gick & Holyoak, 1980; Holyoak & Thagard, 1995), language (Kim, Pinker, Prince, & Prasada, 1991), and rule based learning (Lovett & Anderson, 2005), but also extends to perceptual processes such as the detection of similarities (Medin, Goldstone, & Gentner, 1993). For example, there is evidence to suggest that we recognize objects due to the specific relationships that exists between component shapes (Biederman, 1987).

Given that melodic processing appears to require extracting relational information from melodies, it is a reasonable and parsimonious starting point to propose that the same mechanisms used in other relational tasks could also operate when processing melodies. That is, the representation of melodies as relations between individual notes may be the common underlying mechanism to the approaches that humans employ to encode melodic information (e.g., intervallic and contour). Thus, the strength of relational reasoning lies in the ability to reason beyond the specific features of an object; it is the ability to extract the relationship that an object has with others. Similarly, the ability to recognize a melody (or its shape) rests on appreciating the relationship between the pitches, and not just the specific frequencies of each note.

Another unique and defining aspect of melodies is their inherent temporal and sequential nature. Given this sequential nature, the question remains: how are the elemental pitches within melodies bound together over time such that relationships can be extracted and processed in a listener's mind? Different theories have put forward explanations for such a binding mechanism. Anne Treisman's feature integration theory (Treisman, 1998) posits that there are several different stages in the process. For example, in the initial stage an object's features are processed separately and attention might be likened to a kind of "glue" that binds the various features together. Other researchers have proposed that working memory is responsible for this binding process, wherein the ability to simultaneously hold different features or objects in memory while they are being processed could be likened to a more integrative mechanism (Allen, Baddeley, & Hitch, 2006).

The role of working memory and its interaction with attention is a widely studied and debated phenomenon (e.g., Feng, Pratt, & Spence, 2012; Postle, 2006). Working memory is a dynamic form of memory that is manipulated quickly (in seconds) and used to temporarily store information for further analysis (Baddeley, 2003). In fact, working memory is often associated with objects in attention, and the two concepts are somewhat interconnected. For example, a functional magnetic resonance imaging (fMRI) study strongly implicated their overlap (LaBar, Gitelman, Parrish, & Mesulam, 1999). It should be noted that resolving these opposing theories is beyond the scope of this paper. However, the present study does utilize the conceptualization that working memory may operate as an integrative mechanism to facilitate the binding process when processing melodic information.

Research in the visual domain has shown that working memory may function as a binding mechanism for sequential visual events. Yet this binding mechanism may not be entirely impervious to cognitive strain, as experiments that place participants under dual task conditions with heavy memory (and attentional) demands have shown (Allen et al., 2006; Lavie, 2005). While in the auditory domain studies have looked at the binding of spatial and verbal features through sequential exposure (Maybery et al., 2009), to date no study has systematically and directly investigated the relationship between working memory and binding mechanisms in the specific context of melodic perception.

By examining the extent to which melodic perception depends on working memory resources, the role of working memory in relational processing can be inferred. Whether working memory resources are used similarly across different types of sequential processing, or whether there may be a bias towards musical processing due to a predisposition for musical stimuli is also an open question.

In order to address these questions, participants listened to a melodic stream while either performing a concurrent 2-back task on a visually presented letter stream, or passively watching the letter stream. Following each task, participants were again presented with a melodic stream, but now were required to match the relational attribute (shape) of the melodies to one of two categories.

Thus it is hypothesized that if working memory resources are a prerequisite for relational processing, as some have suggested (Doumas, Hummel, & Sandhofer, 2008; Morrison, Doumas, & Richland, 2011; Morrison, Holyoak, & Truong, 2001), then the ability to perceive melodic content should falter as these resources become depleted. That is, as relational processing falters, perception of the relational content of melodies should consequently suffer.

To minimize possible confounds, an indirect priming approach was used to measure relational processing[1]. During the testing phase the task was to listen to a three-note melody and to categorize the contour of the melody (as "Up-down" or "Down-up"). In this test, the melody could either match or not match the contour and intervallic pattern of the primed melody heard during the exposure phase (see Figure 3). We hypothesized that when the test melody had the same intervals and contour (relations) as the primed melody, reaction times on the categorization task should be faster compared to when the test melody did not have the same contour as the primed melody. This hypothesis is in itself a novel prediction; as to our awareness no previous studies have examined such priming effects on an orthogonal relational categorization task.

[1] A priming design was used instead of more direct approaches such as asking participants which of two choices sounded "the most familiar" to avoid possibly biasing participants' responses. That is, the inherent subjectivity in how to interpret such test prompts may not ensure that relational processing of the exposed melody is in fact measured, and thus was avoided here.

Crucial to this study—and relevant to our proposal that working memory is required for both relational and melodic processing—we also hypothesized that under the 2-back condition, the priming effect would go away. That is, if working memory resources are required for the processing of relations, the depletion of these resources under the 2-back task should prevent relational priming.

Methods

Participants

Sixty participants were recruited from the University of Hawaiʻi at Mānoa, for a total of 30 in the baseline condition (age 20.7 ± 3.1, 22 females) and 30 in the 2-back condition (age 20.8 ± 2.5, 19 females). Participants' musical experience (M = 4 years, SD = 4) and self-reported perfect pitch abilities (6 participants) did not differ across the conditions ($p > 0.1$). All participants were naïve to the experiment and had normal or corrected to normal hearing and vision.

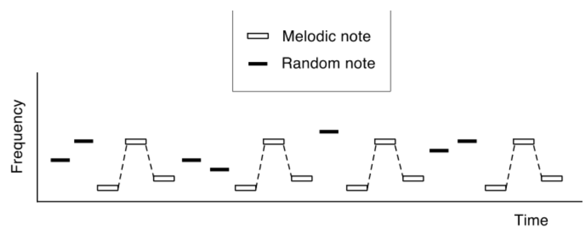

Figure 2: Exposure phase stimuli consisting of auditory stream with interleaved melody.

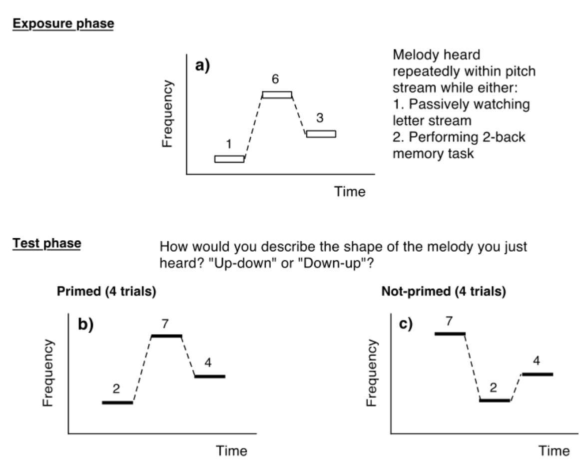

Figure 3: Test phase stimuli. During the test phase participants were asked to categorize melodies as either "Up-down" or "Down-up".

Stimuli

The auditory pitch stream in the exposure phase was constructed using randomly determined pitches from a five-note whole-tone scale[2]. The pitch stream was assembled by a paradigm script using the following procedure: 1) first a random melody was constructed, 2) next, this melody was played and repeated, while 3) interspersing random notes of random quantities (between 0-2 notes) in between each repeated melody (see Figure 2). In addition, a visual letter stream was concurrently presented on a computer screen (participants were only required to respond to the letter stream in the 2-back condition). This letter stream was constructed from randomly chosen non-repeated letters (from the following set: B, C, D, E, F, J, K, L, M, N, P, R, S, T, Y, X, Z). Each letter event was presented for 700 ms, with 16.7 ms of silence and a blank gray screen as separation between events. Each sound pitch was played for 500 ms. In the 2-back condition the letters would repeat after one intervening letter (e.g., B-A-B) at randomly allocated positions.

The test phase consisted of eight two-alternative-forced-choice (2AFC) questions asking participants to categorize the shape of each of the three-note melodies as either "down-up" or "up-down" (see Figure 3).

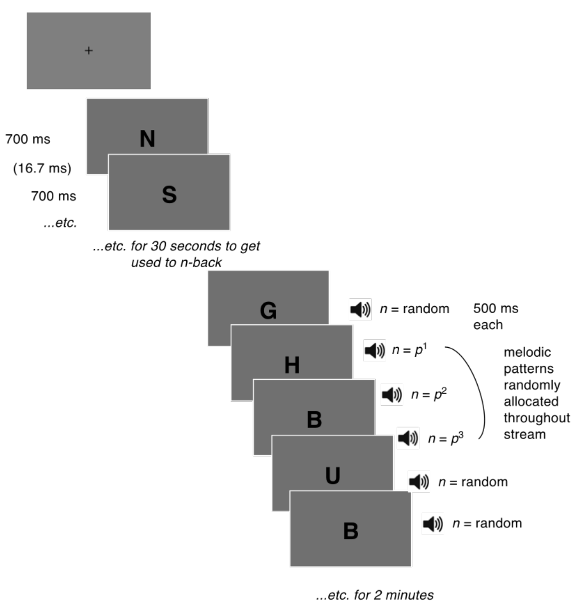

Figure 4: Exposure phase.

Procedure

Both conditions consisted of two phases (with the extra addition of 2-back training prior to the 2-back task). For the first exposure phase, participants listened to the pitch stream. Concurrent to the pitch stream was a visual letter stream that participants were required to monitor (see Figure 4). In the baseline condition participants were not required to respond to the visual stream. During the 2-back visual task participants responded with the spacebar each time a 2-back repetition occurred (e.g. A-B-A, G-Y-G, etc.).

The auditory pitch stream lasted for two minutes (the repeated melody was played approximately 100 times during this period[3]). After the pitch stream ended, the letter stream continued for one minute. Following this exposure phase, participants were then presented with the test phase where they heard eight three-note melodies and were asked to categorize the shape of each melody as either "up-down" or "down-up."

To ensure that participants were familiar with the tasks and could perform the 2-back task, participants in the 2-back group were trained on the 2-back test prior to the experiment. In order to acclimatize participants to the 2-back task, during the actual exposure phase, a lead-in period of 30 seconds for the visual letter stream was used prior to the onset of the pitch stream (see Figure 4).

Results

For the 2-back condition, data from one participant was excluded due to accuracy on the exposure-phase memory task being lower than 70%. Reaction times within and across the two conditions were analyzed using a 2x2 ANOVA containing within factors of priming status (primed vs non-primed) and between factors of exposure-phase task (no task vs 2-back).

Due to the large proportion of trials discarded when using only correct trials (36% for no task, and 48% for 2-back), the ANOVA was conducted on two datasets consisting of 1) all trials and 2) correct trials only.

Both correct and incorrect trials The main effect of priming status approached significance indicating that reaction times tended to be faster for primed melodies compared to non-primed

[2] The whole-tone scale was used to avoid the possibility of having any harmonic or scale related information within the melody and pitch stream as possible confounds.

[3] Note that the melody in the exposure stream is not only priming general up or down relationships, but they are also priming specific relationship, (e.g. 5 whole steps up and 3 whole steps down, see Figure 2).

melodies, $F(1, 57) = 3.8$, $p = 0.06$. An interaction trend also suggests that the priming effect was stronger in the no-task condition compared to the 2-back condition, $F(1, 57) = 3.4$, $p = 0.07$, (see Figure 5).

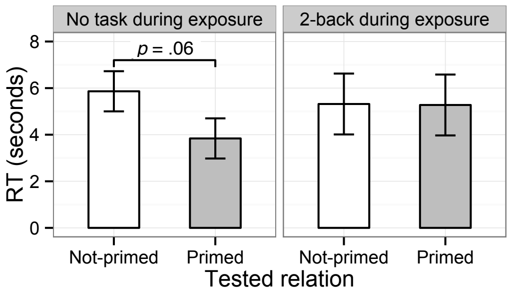

Figure 5: Reaction times for both correct and incorrect trials on the relational categorization task. Error bars = 95% confidence interval.

Correct trials only Although there were neither main effects of n-back task or priming status ($p > 0.1$), the interaction did trend towards significance, indicating that for correct trials only the priming effect also tended to be stronger in the no-task condition compared to the 2-back condition, $F(1, 48) = 3.0$, $p = 0.09$, (see Figure 6).

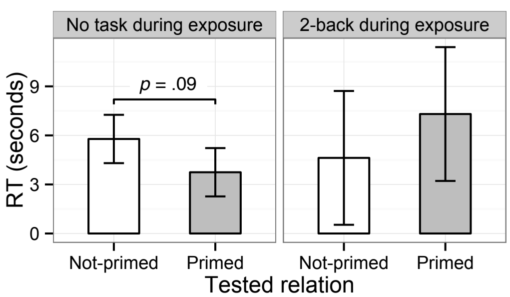

Figure 6: Reaction times for correct trials on the relational categorization task. Error bars = 95% confidence interval.

Discussion

This exploratory study on the relationship between working memory and melodic perception yielded several insights. First, these findings contribute to existing research suggesting that working memory may play a role in relational learning (Morrison et al., 2011; Morrison et al., 2001). Second, under conditions in which working memory was not depleted, reaction times[4] for categorizing melodies that shared relational content with primed melodies tended to be faster compared to non-primed melodies. This implicit learning effect suggests that the underlying mechanism for processing melodic information may involve a relational component. However, it should be noted that this finding must be interpreted cautiously given that this marginal difference failed to be below a conventional alpha level of .05.

Importantly, under conditions of working memory taxation, such a priming trend on the relational processing of melodies was not observed. That is, participants were no longer faster to categorize primed melodies compared to non-primed melodies. This finding could mean that 1) working memory may serve as the integrative mechanism for encoding melodic information, 2) processing of relational content of melodies may not be automatic or impervious to the concomitant side effects of working memory taxation, and 3) available working memory may be a prerequisite for melodic perception and relational learning. Note however, that until more data are obtained for corroboration, and in light of the non-significant trends, these conjectures are highly speculative.

A possible algorithmic level account for the linkage between working memory and relational learning exists in at least one neurally-plausible computational model, which defines working memory as dynamic binding operations occurring in the prefrontal cortex (Doumas et al., 2008). Thus, while researchers have suggested a link between relational learning and music, the three-way linkage between working memory, relational learning, and melodic processing is a novel one. However, in light of the evidence presented here, we believe this notion warrants further exploration. This could include varying the extent to which working memory is taxed during the exposure stage in subsequent experiments.

Given that Baddeley's formulation of working memory (Allen et al., 2006) contains multiple components (i.e., phonological loop, episodic buffer, visuo-spatial sketchpad, and central executive), the question remains as to which components are involved in melodic perception, and if some are more heavily used than others. For example, the n-back task may engage the phonological loop and episodic buffer more than other WM components, but which components does melodic perception itself engage? These are

[4] RTs were measured from the offset of last note in the melody. RTs were relatively higher since participants had to read the prompt, think about what was being asked (they were not trained on this categorization task before hand), and also reflect back on the melody.

all important theoretical considerations, and ones that future experiments should explore.

References

Allen, R. J., Baddeley, A. D., & Hitch, G. J. (2006). Is the binding of visual features in working memory resource-demanding? *Journal of experimental psychology: General, 135*(2), 298.

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *The American journal of psychology, 84*, 147-166.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: A key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception and Performance, 6*(3), 501.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review, 94*(2), 115-147.

Doumas, L. A. A., Hummel, J. E., & Sandhofer, C. M. (2008). A theory of the discovery and predication of relational concepts. *Psychological Review, 115*(1), 1.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review, 85*(4), 341.

Dowling, W. J. (1984). Assimilation and tonal structure: Comment on Castellano, Bharucha, and Krumhansl. *Journal of Experimental Psychology, 113*(3), 417-420.

Dowling, W. J. (1988). Tonal structure and children's early learning of music. In J. Sloboda (Ed.), *Generative Processes in Music*. Oxford: Oxford University Press.

Feng, J., Pratt, J., & Spence, I. (2012). Attention and visuospatial working memory share the same processing resources. *Frontiers in Psychology, 3*, 1-11.

Fujioka, T., Trainor, L. J., Ross, B., Kakigi, R., & Pantev, C. (2004). Musical training enhances automatic encoding of melodic contour and interval structure. *Journal of Cognitive Neuroscience, 16*(6), 1010-1021.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive science, 7*(2), 155-170.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12*(3), 306-355.

Holyoak, K. J., & Thagard, P. (1995). *Mental leaps: Analogy in creative thought*. Cambridge, MA: MIT Press.

Kim, J. J., Pinker, S., Prince, A., & Prasada, S. (1991). Why no mere mortal has ever flown out to center field. *Cognitive science, 15*(2), 173-218.

LaBar, K. S., Gitelman, D. R., Parrish, T. B., & Mesulam, M. (1999). Neuroanatomic overlap of working memory and spatial attention networks: a functional MRI comparison within subjects. *Neuroimage, 10*(6), 695-704.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in cognitive sciences, 9*(2), 75-82.

Lovett, M. C., & Anderson, J. R. (2005). Thinking as a production system. In K. J. Holyoak & R. Morrison (Eds.), *The Cambridge handbook of thinking and reasoning* (pp. 401–429). New York: Cambridge University Press.

Maybery, M. T., Clissa, P. J., Parmentier, F. B., Leung, D., Harsa, G., Fox, A. M., & Jones, D. M. (2009). Binding of verbal and spatial features in auditory working memory. *Journal of Memory and Language, 61*(1), 112-133.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*(2), 254.

Morrison, R. G., Doumas, L. A., & Richland, L. E. (2011). A computational account of children's analogical reasoning: balancing inhibitory control in working memory and relational representation. *Developmental Science, 14*(3), 516-529.

Morrison, R. G., Holyoak, K. J., & Truong, B. (2001). *Working memory modularity in analogical reasoning.* Paper presented at the Proceedings of the twenty-third annual conference of the Cognitive Science Society.

Page, M. P. A. (1994). Modelling the perception of musical sequences with self-organizing neural networks. *Connection Science, 6*(2-3), 223-246.

Postle, B. R. (2006). Working memory as an emergent property of the mind and brain. *Neuroscience, 139*(1), 23-38.

Trainor, L. J., McDonald, K. L., & Alain, C. (2002). Automatic and controlled processing of melodic contour and interval information measured by electrical brain activity. *Journal of Cognitive Neuroscience, 14*(3), 430.

Treisman, A. M. (1998). Feature binding, attention and object perception. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 353*(1373), 1295-1306.