Supramodal Representations in Melodic Perception

Ahnate Lim (ahnate@hawaii.edu)
Department of Psychology, University of Hawaii at Manoa
2530 Dole Street, Honolulu, HI 96822 USA

Leonidas A. A. Doumas (alex.doumas@ed.ac.uk)
Department of Psychology, University of Edinburgh
7 George Square, Edinburgh, Midlothian EH8 9JZ, UK

Scott Sinnett (ssinnett@hawaii.edu)
Department of Psychology, University of Hawaii at Manoa
2530 Dole Street, Honolulu, HI 96822 USA

Abstract

Music is highly relational and in this manner shares much in common with other characteristically human behavior. While this may suggest that the processes used in music perception could be domain general, the nature and flexibility of these representations remain less understood. If the underlying representations and manipulations required for perceiving music overlap with those in other cognitive domains, it should be fairly easy to map such representations across domains. This hypothesis was tested and supported using a novel experiment with melodic stimuli in the auditory modality and analogous visual sequential stimuli (Gabor sequences) in the visual modality. Testing for transfer across the two modalities and for the two types of representations (contour and intervallic) was done through four counterbalanced conditions. Cross-modal mapping was successful in three out of the four conditions, implying general flexibility of representational transfer. Implications for representational flexibility, sequential learning and future studies are discussed.

Keywords: Melodic perception; relation learning; cross-modal mapping; representations; categorization.

Introduction

As research techniques for studying human behavior and the brain have evolved, the underlying processes of music perception have continued to fascinate cognitive scientists and spur much research. Recent advances in neuroscience for example, have shown that the simple act of perceiving music involves distributed activity throughout the brain, including diverse regions such as Broca's area (Fadiga, Craighero, & D'Ausilio, 2009), the pre-frontal cortex (Bengtsson, Csíkszentmihályi, & Ullén, 2007), as well as the amygdala (Limb, 2006). Indeed, many if not all of these regions are used in other tasks such as speech processing, for example (Koelsch et al., 2004). Given the integrated and distributed neurological underpinnings of music, it should come as no surprise that a host of relationships has been made between music and other areas of cognition. Aside from the common comparisons made between music and language, many have suggested that music may also help to aide us in understanding other behavior such as domain-general aesthetic preferences (Marcus, 2012), implicit and explicit learning of grammatical structures (Rohrmeier & Rebuschat, 2012), and complex event sequencing (Tillmann, 2012), to name but a few. Other approaches have sought to uncover links between musical training and performance in other domains such as mathematics, language, spatial-temporal abilities, and verbal memory (for a summary, see Rauscher, 2003). Thus it is clear that gaining a better understanding of the underlying processes and building blocks in music perception can aid towards understanding the underlying mechanisms and resources used in other cognition domains as well.

The Melody

One of the most fundamental and salient aspects of music is the melody. Simple melodies consist of discrete units or notes, where each note is characterized by a pitch, or fundamental frequency (e.g., Hertz value). The core strategy humans use to process and store familiar melodies is through a *relative pitch* code (Attneave & Olson, 1971; Page, 1994). This relative pitch code stores the pitch sequence of a melody in terms of the *relations* or intervals (specific frequency differences) between each note. For example, the song *Happy Birthday* is immediately recognizable due to the unique intervals between each of the notes. That is, one immediately recognizes this song regardless of whether it starts on a low or high note due to the unique intervallic pattern between all subsequent notes. There is considerable evidence on the use of relative pitch information in adults through both behavioral (Dowling, 1978, 1988) as well as neuroimaging studies (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004; Trainor, McDonald, & Alain, 2002).

In addition to relative pitch, the contour (general shape, or sequence of up and down movements in frequencies from note to note) is another characteristic upon which melodies can be categorized. Given the existence of these characteristics, the question remains as to how they contribute to a listener's mental representation of a melody. Note that while a melody with the same contour as *Happy Birthday* but with a different intervallic sequence would be perceived as a completely different song, it would still have the same general "shape", or up and down pattern. Although the intervallic pattern may be the most overtly salient and representative feature of a melody to humans, studies have shown that human adults are also sensitive to melodic contour, at least in the short term (Bartlett & Dowling, 1980; Dowling, 1978). Furthermore, sensitivity to intervallic and contour properties is present even in young infants (Trehub, 2001; Trehub, Bull, & Thorpe, 1984; Trehub, Trainor, & Unyk, 1993).

Despite the difference in informational content between intervallic and contour properties, there is a fundamental similarity in the *relational* nature of this information. That is, these representations depend 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. It is within this relational capacity that melodic perception can be said to share a cornerstone property with many other cognitive processes.

The Role of Relations in Cognition

The ability to process relational properties has been proposed as a fundamental mechanism underlying a wide range of cognitive phenomena. This includes not only higher level reasoning skills such as analogy-making (Gentner, 1983; Holyoak & Thagard, 1995), language (Kim, Pinker, Prince, & Prasada, 1991), and rule based learning (Lovett & Anderson, 2005), but also extend to perceptual processes such as the detection of similarities (Medin, Goldstone, & Gentner, 1993).

Given that melodic processing appears to require extracting relational information from melodies, it is reasonable to hypothesize that the same mechanisms used in other relational tasks may also operate when processing melodies. Common to the approaches (e.g., intervallic and contour) that both adults and infants use to encode melodies is the underlying relations between individual notes. That is, the ability to recognize a melody (or its shape) involves the processing of the relationships between pitches, and not just the specific frequencies of individual pitches.

Domain General or Domain Specific?

When speaking of mechanisms that enable the perception and processing of music, a central question is whether these may be specific to music, or if they underpin other cognitive mechanisms (i.e., domain-specific or domain-general) as well. One of the approaches to answering this question has been to look at musical behavior in infants, since abilities that are present at infancy are less likely to be acquired through experience or specialization and more likely attributable to innate core cognitive functions. The brunt of the work on infant music perception suggests that many of the cognitive subsets used for processing music are indeed domain-general mechanisms (Hannon & Trainor, 2007; Patel, 2008; Trehub & Hannon, 2006).

Despite the existing body of evidence pointing towards domain-general mechanisms for music, the true underlying nature of the mechanisms and representations of music perception remain elusive and much less understood. While listening to music, how are representations stored and subsequently manipulated? Evidence from behavioral studies with adults have shown that musical pitch can be mapped to a wide variety of representations, including vertical space (Melara & O'Brien, 1987), luminosity and loudness (Hubbard, 1996; McDermott, Lehr, & Oxenham, 2008), as well as words related to emotion, size, sweetness, texture and even temperature (Eitan & Timmers, 2010; Nygaard, Herold, & Namy, 2009; Walker & Smith, 1984). For instance, a recent study using video clips of singers performing different types of hand motions (as primes for musical stimuli) demonstrated that pitch processing shares representations with spatial processing (i.e., higher spatial movements in the visual modality primed the perception of higher pitches; see, Connell, Cai, & Holler, 2013). Evidence of such cross-modal mappings is interesting and could perhaps be explained from an associative learning perspective (where learning context and environmental regularities may promote such cross-domain associations). To gain a better understanding of the representational content of melodies, a more direct and precise approach may be helpful. For instance, are the representations of contour (up or down direction between notes) and intervals (frequency distance between notes) unique to music, or are they used in other areas of cognition as well? Although evidence of crossmodal correspondences between pitch and spatial frequency has been previously demonstrated (Evans & Treisman, 2010; for a review of correspondence in general, see Spence, 2011), the full automaticity of such correspondence is still under debate (Spence & Deroy, 2013). Furthermore, to our awareness there are no studies that have utilized *sequences* of stimuli (as direct analogs to melodies) to examine correspondences in representational content[1], an important point considering the inherently sequential nature of music (all music unfolds through time).

In this experiment we examined whether the representations used for melodic perception can be mapped to visual sequences. We define domain-generality here as the extent to which melodic representations can be flexibly mapped to other senses (e.g., Doumas, Hummel, & Sandhofer, 2008). Although other mapping tasks (e.g., Connell et al., 2013) have been used previously, such findings are more speculative due to the possibility of indirect priming. In order to examine the transfer of melodic sequences to visual spatial frequency sequences, as well as the crossmodal processing of melodic representations, visual stimuli with analogous dimensions (i.e., frequency and duration) were created for this study.

The representational characteristics of melodic perception were investigated using a novel paradigm consisting of matched sequential auditory and visual stimuli analogous on several basic dimensions. Gabor patches (i.e., sinusoidal gratings) were used as the visual analogue to auditory pitch

---

[1] Although auditory pitch and visual gabor sequences have been previously used to study asynchrony and temporal recalibration mechanisms (Heron, Roach, Hanson, McGraw, & Whitaker, 2012).

stimuli, as they can be defined by both frequency and duration (how long stimuli is presented). Thus, any learning transfer of relational properties from melodies to visual sequences (and vice versa) could be observed. In theory, if learning relations in one modality could transfer to stimuli in a separate modality, this would be evidence for the flexible use (or mapping) of relations in a domain-general manner (Doumas et al., 2008; Gentner, 1983; Holyoak & Thagard, 1995). In this study we tested for the transfer of both contour and intervallic properties important for processing melodies. Due to the fact that an intervallic relation may carry with it more information (magnitude) than contour (direction), it is hypothesized that the transfer of intervallic relations may be harder than for contour relations. In light of considerable evidence on the domain general characteristics of music, it is also predicted that simple melodic representations should map fairly easily to the visual domain, and vice versa.

Methods

Participants

Fifty-two participants (mean age = 21±5; 32 females) were recruited from undergraduate courses at the University of Hawaii at Manoa and offered course credit for their participation. Each participant was randomly allocated into one of the four conditions for a total of 13 participants in each condition. Each group did not significantly differ from one another on either age ($F(3,41) = 2.0$, $p = 0.1$) or sex statistics ($\chi^2 = 6.9$, df = 6, p = 0.3). All participants were naïve to the purpose of the experiment and ethical approval was obtained from the University's Committee on Human Subjects.

Stimuli and apparatus

Stimuli were presented on a 21inch Core2Duo 2.4 GHz iMac computer using the Psychophysics Toolbox extension for Matlab. Participants were seated at an eye to monitor distance of approximately 60cm. From this distance, all presented auditory stimuli occurred at approximately 75 decibels, as measured by a sound meter. Responses were made via key presses to one of two buttons on a keyboard.

Auditory stimuli consisted of sinusoidal sound waves at different frequencies determined by the equation

$$F = f \cdot 2^{\left(\frac{n}{12}\right)}$$

where n could vary from 0 to 11, producing the range of 12 semitone pitches contained within an octave, and f was set to 440 Hz (A4).

Visual stimuli consisted of Gabor patches generated with Gaussian envelopes and rotated at a 45° angle. Selection of the spatial frequency followed an analogous process determined by the equation

$$F = s \cdot 2^{\left(\frac{n}{10}\right)}$$

where s represents the spatial frequency.

Each melodic and visual sequence consisted of three items. The relationship between these three items would vary depending on whether the participant was placed under the contour or intervallic discrimination condition (described below). The presentation stream was continuous, with each pitch and Gabor pattern being presented for one second (with no pauses in between pitches). Thus each melodic or Gabor sequence would have a total duration of three seconds.

Contour Discrimination In this condition, the property that differentiated the two types of sequences to be categorized was the contour. One type of sequence would have a steadily increasing contour, where if the first note was n, the second note would be $n + 2$, and the third would be $n + 4$ (see Figure 1). The other type of sequence would have an up-down frequency relationship, where if the first note was n, then the second note would be $n + 2$, and the third would be $n - 2$.

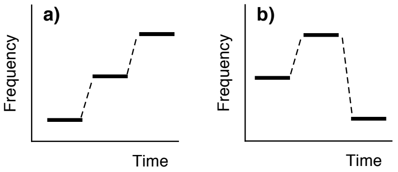

Figure 1: Example stimuli for the contour discrimination condition. The solid lines represent individual notes in the melody. In a) the contour is up-up, and in b) up-down.

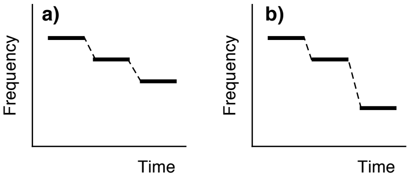

Figure 2: Example stimuli for the intervallic discrimination condition. Note the different intervals (but similar contours) for the last notes in the sequences.

Intervallic Discrimination For intervallic discrimination, the property that differentiates the two types of sequences is the interval or distance between each element in the sequence. One sequence would have a steadily decreasing contour, where if the first note was n, the second note would be $n - 2$, and the third would be $n - 4$. The other sequence would have a non-steady (exponential) decrease, where if the first note was n, the second note would be $n - 2$, and the third would be $n - 8$ (see Figure 2).

Procedure

Participants were randomly allocated into one of two groups and received either the melodic or the visual sequence training. Within each of those groups, half (n=13) would perform the contour discrimination task, and the other half would perform the intervallic discrimination task.

Within these four sub-groups, the procedure was identical and started off with an introduction screen on the computer providing participants with instructions for the task. Next, participants would either hear or see a sequence of three auditory pitches or Gabor patches (respectively, see Figure 3), depending on the condition they were placed in. Stimuli were presented in the same modality throughout the learning phase. Participants were then instructed to categorize each stimuli sequence into one of two categories using the keyboard. Since the purpose was for participants to discover the categories for themselves via feedback provided at the end of the trial, they were instructed to guess on the initials trials and to then use the received feedback to guide their categorization on subsequent trials. This learning phase continued until the participant answered correctly 12 times in a row, whereupon the experiment would proceed to the testing phase.

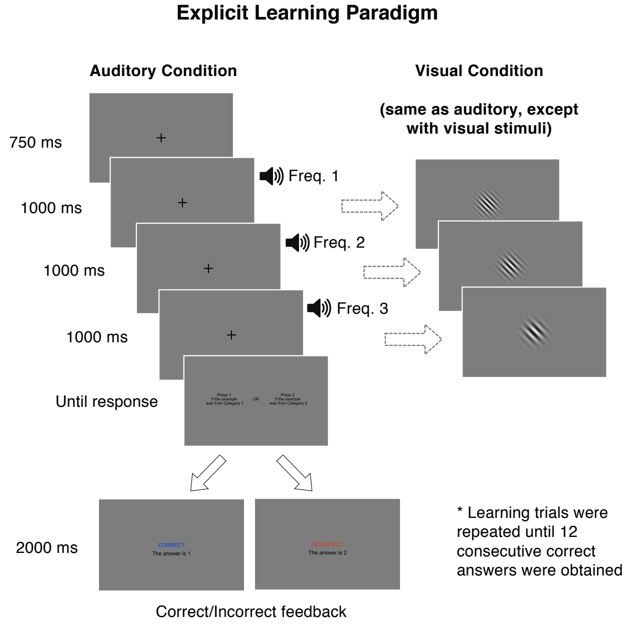

Figure 3: Time course schematic of the experiment.

In the testing phase, participants were presented with 20 exemplars in the opposite sensory modality than the learning phase. Half of these exemplars were from one category, and the other half was from the other category (previously learned, albeit in a separate modality). Presentation order of the 20 exemplars was randomized. No feedback was provided during the testing phase. Participants were instructed to categorize in the same manner as they had done during the learning phase, but were provided no other directions.

Results

Of the 13 participants allocated into each of the four conditions, one participant failed to reach training criterion in the intervallic discrimination condition with Gabor patch training and four participants failed to reach criterion in the intervallic discrimination with melodic training. The rest of the analyses were conducted only on participants that completed the training section. The number of training trials until criterion was reached did not significantly differ across the four conditions (overall $M = 31$, $SD = 22$, $F(3,43) = 1.5$, $p = 0.2$)

To determine whether participants subsequently categorized the test stimuli better than chance (detection of any difference between the two stimuli categories, regardless of correct mapping of the corresponding categories from one modality to the other), one-sample t-tests were conducted on the absolute accuracy difference from chance ($|accuracy - 0.5|$) of each of the four conditions. Results indicate that in all testing conditions participants detected differences between the two types of categories significantly better than chance (overall discrimination accuracy $M = 0.83$, $SD = 0.15$, all $p < .001$).

Of crucial interest to this study was whether participants could map categorical representations from one modality to another. One-sample t-tests were conducted on mean accuracy scores to determine if accuracy was greater than chance (0.5). In the contour discrimination condition, mapping accuracy was significantly better than chance for transfer from Gabor patches to melodies ($M = 0.78$, $SD = 0.18$, $t(12) = 5.7$, $p < .001$), as well as for transfer from melodies to Gabor patches ($M = 0.72$, $SD = 0.37$, $t(12) = 2.1$, $p = .055$). In the intervallic discrimination condition, mapping accuracy was significantly better than chance only for transfer from Gabor patches to melodies ($M = 0.70$, $SD = 0.29$, $t(11) = 2.4$, $p = .03$), and not for transfer from melodies to Gabor patches ($M = 0.62$, $SD = 0.33$, $t(8) = 1.1$, $p = .3$). To compare performance across the different conditions, a two by two ANOVA was conducted. No significant main effects were found for either factors of category type ($F(1,43) = 1.1$, $p = 0.3$) or modality ($F(1,43) = 0.7$, $p = 0.4$).

Discussion

The purpose of this study was to 1) determine whether representations for short melodic sequences can be mapped to representations in the visual modality, and 2) examine whether mapping performance differs across different classes of melodic representations (contour and intervals) and their visual counterparts. A novel experimental paradigm involving active categorization with feedback and subsequent cross-modal testing was employed.

Overall, tentative evidence for shared representational resources was provided by successful cross-modal mappings in three out of the four conditions. In the contour discrimination condition, cross-modal mapping occurred in both directions (for visual sequences to melodies as well as for melodies to visual sequences). In the intervallic discrimination condition, on the other hand, successful mapping occurred only for transfer from visual sequences to melodies, and not from melodies to visual sequences. This finding, in conjunction to the higher rate of failure to reach training criterion and overall lower accuracy scores for the intervallic condition compared to contour (see Figure 4), may suggest that intervallic discrimination is harder than contour discrimination. These findings may also provide tentative support for the notion that discriminating change in frequency direction is easier than discriminating change in frequency magnitude (Lim, Doumas, & Sinnett, 2012; Trehub et al., 1984).

It is worth noting a trend in the current data suggesting that mapping transfer may not be bi-directional. That is, transfer could have been slightly easier from Gabor patches to melodies than the opposite. Although not significant, lower accuracy scores for transfer from melodies to visual sequences, compared to the opposite, could suggest a possible bias or difference in representational flexibility of melodies versus visual sequences. Whether this possible difference may arise during the acquisition or transfer stage of representations is an open question. Furthermore, the methodology used here would be strengthened by models demonstrating how such domain generalizability is achieved, given that the question remains as to what precisely is learned from exposure to such melodic and visual sequences.

In addition to previous studies indicating shared resources between pitch and space, this experiment extends such findings in a more experimentally controlled design to visual temporal sequences. Further, the present results supplement existing findings on the domain-general aspects of musical processing. When discussing the domain-general nature of music, perhaps the most frequent comparisons are made between music and language, since the two domains do share many commonalities. Relevant to this study, the role of sequential learning is fundamental to both music and language. It has been suggested that the processing of sound is an important opportunity during development that subsequently aids the brain in learning how to process sequential stimuli in general (Conway, Pisoni, & Kronenberger, 2009). This notion is consistent, for example, with a study that used an implicit learning task to measure visual sequencing abilities in deaf children with cochlear implants. It was found that deaf children exhibited general sequencing deficits that were correlated to deficits in language outcomes, leading the investigators to hypothesize that deprivation of early sequential learning opportunities in deaf children may explain their continued difficulties in language even after receiving cochlear implants (Conway, Pisoni, Anaya, Karpicke, & Henning, 2011). It has been proposed that temporality and sequential processes may also separate music from other art forms. Although music may rely on domain-general mechanisms, its unique appeal may lie in its inherently temporal nature that allows for the close interaction of prediction and novelty (Kivy, 1993; Marcus, 2012).

Given that most people can perceive music despite having little to no formal musical training (Bigand & Poulin-Charronnat, 2006; Koelsch, Gunter, Friederici, & Schröger, 2000), it has been suggested that musical experience and knowledge for most is acquired through *implicit learning* (for a recent review on the topic as it relates to music, see Rohrmeier & Rebuschat, 2012). That is, similar to language, mere exposure to music is adequate for the development and acquisition of knowledge about fairly complex sets of regularities and relationships. In the real world representations may be acquired through different means where learning could in fact occur either explicitly or implicitly (with or without awareness, supervision, or direct knowledge). Although this experiment used a more explicit learning routine through active categorization and feedback, future studies may provide a more ecological examination on musical behavior by incorporating *implicit* learning paradigms (i.e., passive observation), in order to more holistically account for domain generalizability of music, as well as any differences in representational content that may arise due to different learning approaches.

References

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

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.

Bengtsson, S. L., Csíkszentmihályi, M., & Ullén, F. (2007). Cortical regions involved in the generation of musical structures during improvisation in pianists. *Journal of Cognitive Neuroscience, 19*(5), 830-842.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100*(1), 100-130.

Connell, L., Cai, Z. G., & Holler, J. (2013). Do you see what I’m singing? Visuospatial movement biases pitch perception. *Brain and cognition, 81*(1), 124-130.

Conway, C. M., Pisoni, D. B., Anaya, E. M., Karpicke, J., & Henning, S. C. (2011). Implicit sequence learning in deaf children with cochlear implants. *Developmental Science, 14*(1), 69-82.

Conway, C. M., Pisoni, D. B., & Kronenberger, W. G. (2009). The Importance of Sound for Cognitive Sequencing Abilities The Auditory Scaffolding Hypothesis. *Current Directions in Psychological Science, 18*(5), 275-279.

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

Eitan, Z., & Timmers, R. (2010). Beethoven’s last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition, 114*(3), 405-422.

Evans, K. K., & Treisman, A. (2010). Natural cross-modal mappings between visual and auditory features. *Journal of vision, 10*(1), 6.

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences, 1169*(1), 448-458.

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.

Hannon, E., & Trainor, L. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends in cognitive sciences, 11*(11), 466-472.

Heron, J., Roach, N. W., Hanson, J. V., McGraw, P. V., & Whitaker, D. (2012). Audiovisual time perception is spatially specific. *Experimental brain research, 218*(3), 477-485.

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

Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *The American journal of psychology*, 219-238.

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.

Kivy, P. (1993). *The fine art of repetition: Essays in the philosophy of music*. Cambridge, England: Cambridge Univ Press.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing:"nonmusicians" are musical. *Journal of Cognitive Neuroscience, 12*(3), 520-541.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience, 7*(3), 302-307.

Lim, A., Doumas, L. A. A., & Sinnett, S. (2012). *Modeling Melodic Perception as Relational Learning Using a Symbolic-Connectionist Architecture (DORA).* Paper presented at the 34th Annual Conference of the Cognitive Science Society, Sapporo, Japan.

Limb, C. J. (2006). Structural and functional neural correlates of music perception. *The Anatomical Record Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology, 288*(4), 435-446.

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.

Marcus, G. F. (2012). Musicality: Instinct or Acquired Skill? *Topics in Cognitive Science, 4*, 498-512.

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science, 19*(12), 1263-1271.

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

Melara, R. D., & O'Brien, T. P. (1987). Interaction between synesthetically corresponding dimensions. *Journal of experimental psychology: General, 116*(4), 323.

Nygaard, L. C., Herold, D. S., & Namy, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive science, 33*(1), 127-146.

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

Patel, A. D. (2008). *Music, language, and the brain*. Oxford, England: Oxford University Press.

Rauscher, F. (2003). Can Music Instruction Affect Children's Cognitive Development? *ERIC Clearinghouse on Early Education and Parenting.* Retrieved 10 October, 2011, from http://www.ericdigests.org/2004-3/cognitive.html

Rohrmeier, M., & Rebuschat, P. (2012). Implicit learning and acquisition of music. *Topics in Cognitive Science, 4*(4), 525-553.

Spence, C. (2011). Crossmodal correspondences: a tutorial review. *Attention, Perception, & Psychophysics, 73*(4), 971-995.

Spence, C., & Deroy, O. (2013). How automatic are crossmodal correspondences? *Consciousness and cognition, 22*(1), 245-260.

Tillmann, B. (2012). Music and Language Perception: Expectations, Structural Integration, and Cognitive Sequencing. *Topics in Cognitive Science, 4*(4), 568-584.

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.

Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences, 930*(1), 1.

Trehub, S. E., Bull, D., & Thorpe, L. A. (1984). Infants' perception of melodies: The role of melodic contour. *Child Development, 55*(3), 821-830.

Trehub, S. E., & Hannon, E. E. (2006). Infant music perception: Domain-general or domain-specific mechanisms? *Cognition, 100*(1), 73-99.

Trehub, S. E., Trainor, L. J., & Unyk, A. M. (1993). Music and Speech Processing in the First Year of Life. *Advances in Child Development and Behavior* (Vol. 24, pp. 1-35). New York: Academic Press.

Walker, P., & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception, 13*(1), 75-81.