Melody Recognition: Effects of Articulation Format

Stephen Wee Hun Lim (gmslwhs@nus.edu.sg)
Winston D. Goh (psygohw@nus.edu.sg)
Department of Psychology, National University of Singapore
Block AS4, 9 Arts Link, Singapore117570

Abstract

Various surface features – timbre, tempo, and pitch – influence melody recognition memory, but articulation format effects, if any, remain unknown. For the first time, these effects were examined. In Experiment 1, melodies that remained in the same, or appeared in a different but similar, articulation format from study to test were recognized better than were melodies that were presented in a distinct format at test. A similar articulation format adequately induced matching. Experiment 2 revealed that initial perceptual (dis)similarity as a function of the location of articulation (mis)match between two instances of the melody did not accurately determine discrimination performance. An important boundary condition of the matching process was defined: Whether matching occurs depends on the physical quantity, rather than location, of fit between the memory trace and the recognition probe, suggesting a global matching advantage effect.

Keywords: Melody recognition memory; articulation format effects; global matching advantage

Introduction

When we hear a piece of music, we detect and occasionally remember phrases, motifs, themes, syncopations, suspensions, tonic chords, cadences, and so on. We recognize the instrument playing the melody, or even identify with the emotions of the specific musician performing the work. To this end, what exactly is the nature of mental representations that underlie the music experience? To address this question, it is useful to first recognize that there are two kinds of information in music, namely *abstract structure* and *surface characteristics* (Trainor, Wu, & Tsang, 2004). The *abstract structure* entails the relative pitches and ratios of the durations between adjacent musical notes, regardless of the individual note's absolute pitch level or length *per se*. *Surface characteristics*, in contrast, contain the non-structural aspects of the music, such as absolute pitch, tempo, and timbre. Both the abstract structure and surface characteristics contribute towards musical interpretation. Representing the abstract structure enables recognition of a melody across different performances, and musical variations of a motif within a musical composition (Large, Palmer, & Pollack, 1995). For example, *Happy Birthday* retains its identity and is readily recognized even when it is played or sung in various keys and tempos, or by different voices or instruments. Yet, these very surface characteristics lead us to identify the specific musician and unique performance of the work, defining the emotional interpretation of that rendition. While Raffman (1993) has suggested that only the abstract structural information is encoded into long-term memory (LTM), others have reported that surface features, such as timbre (e.g., Peretz, Gaudreau, & Bonnel, 1998) and tempo (e.g., Halpern and Müllensiefen, 2008), are also encoded into LTM during a melody recognition task.

In music, the way a melody is articulated shapes its surface appearance. In the extant literature that examined the effects of surface characteristics on melody recognition performance, it is surprising that no study has explored the effects of articulation format, even though it is a feature that is commonly manipulated by both composers and performers. Trained musicians commonly define articulation as whether the music (e.g., melody) is played in a *legato* (i.e., continuous) or *staccato* (i.e., detached) format. Because no one has studied the influence of articulation on melody recognition, our initial motivation was to add to that literature. Thus far, memory representations that subserve explicit recognition of melodies appear to be formed by a highly specialized association that binds together characteristics such as timbre and tempo with melody identity. It is thus attractive to ask whether the articulation feature is tied to a melody's identity and computed during the perceptual analysis of the melodic input. By addressing this question, we hope to explicate more fully the central idea that variability in surface features, along with the idealized canonical structure of music, is important in music perception and processing.

To examine the effects of articulation format on melody recognition, we designed the melody to occur either fully in *legato* form, fully in *staccato* form, or in mixed articulation format (i.e., a combination of *legato* and *staccato* components). When the melody was played in *staccato* form, the duration of each note in the melody was manipulated to last 10% of the full duration when the note was played in *legato* form. The schematic of the eight different articulation formats is shown in Figure 1. These formats are coded as *l*, *s*, *a*, *b*, *c*, *d*, *e*, and *f*: The *legato* and *staccato* formats are abbreviated as format *l* and *s*, respectively, while the six mixed-articulation formats follow an alphabetical system of coding for ease of reference. Each set of four boxes represents sequentially the four bars of the melody respectively.

Taking format *f* for instance, the melody opens in *staccato* form (i.e., the notes of the melody are articulated by the instrument in a disjointed fashion) for the first bar, switches to *legato* form (i.e., the notes are now articulated smoothly in a continuous manner) by the second bar, returns to *staccato* mode in the third bar, and finally closes with a long-sounding note in the final bar.

l	L	L	L	○
s	•	•	•	○
a	•	L	L	○
b	L	•	L	○
c	L	L	•	○
d	•	•	L	○
e	L	•	•	○
f	•	L	•	○

L – *legato* • – *staccato* ○ – single long note

Figure 1: Schematic of the eight different articulation format manipulations.

Experiment 1

In Experiment 1, we asked two questions: (1) Is articulation feature information retained in LTM, and (2) what is the role of feature similarity in melody recognition memory? Our first goal was to investigate the effects of manipulating articulation context on melody recognition. The hypothesis was that to the extent that articulation format information is not erased from, but is in fact preserved in, LTM, discrimination performance ought to improve when old melodies are repeated in the same articulation format, as compared to when the melodies appeared in a distinct articulation format during the recognition stage.

In addition, we recognized that extant studies that examined surface feature effects have used test stimuli that were denoted as either of the same or different format, neglecting effects that could arise from varying magnitudes of intermediate perceptual differences. For instance, Peretz *et al.* (1998) presented melodies in timbres at test that were either the same as, or distinct from, those used at study; Halpern and Müllensiefen (2008) made the tempo changes in altered tunes "large enough to be perceptible" (p. 1378). Effects of fine-grained perceptual details of surface features, such as tempo or timbre, have been somewhat overlooked, so it is unclear whether these details actually contributed to the disparate surface feature effects observed in the literature. As such, a second goal was to assess the contribution of fine perceptual details in melody recognition memory, by including a similar-articulation-format condition. We speculated that to the extent that articulation similarity constitutes an integrated part of the matching and retrieval processes involved in melody recognition, performance ought to improve even when old melodies are tested with a different *but similar* articulation format, as compared to when the melodies appeared in a distinct articulation format.

Method

Participants Forty-seven introductory psychology students participated for course credit.

Materials The stimulus set contained 48 novel monophonic melodies (see Figure 2 for samples). An equal number of four-bar melodies were composed in the tonality (key) of C major, C minor, G major, or G minor. The melodies started either on the tonic, mediant, or dominant, but always ended with a single long note on the tonic of their home key. Each melody was written in simple triple or simple quadruple time, lasting approximately six seconds or 7.2 seconds respectively. The melodies were constructed using the *Finale 2009* software, and saved as .wav sound files.
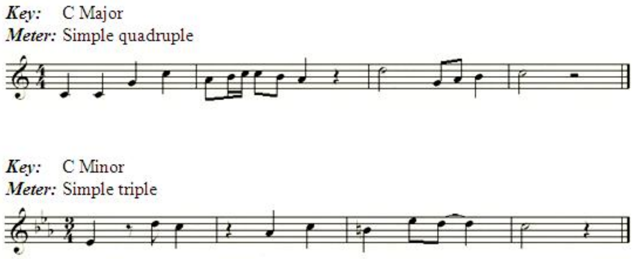
Figure 2: Samples of the 48 melodies used.

Prior to conducting Experiment 1, we first derived a multidimensional "articulation map" using MDS techniques (Kruskal & Wish, 1978) that shows the similarity relations between the individual articulation formats that will be used as the stimulus materials. This procedure was necessary to ensure that the selection of specific articulation formats for

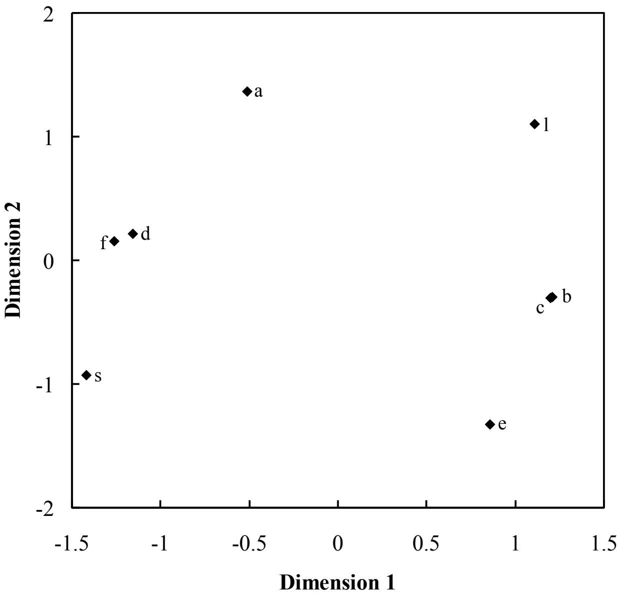

Figure 3: Two-dimensional MDS solution for eight articulation formats.

use in the subsequent main experiments can be based on objective measures of the degree of perceived similarity among different articulation formats. Sixteen students from the same population sample but who did not participate in the main experiments rated the pairwise similarity of the eight articulation formats across four different melodies using a 7-point Likert scale. The two-dimensional MDS solution (Kruskal's stress = .15, R^2 = .85) for the eight articulation formats appears in Figure 3. The interpretation is that the further away two articulation formats are positioned from each other in space, the more perceptually distinct they are. Two different combinations of articulation formats were selected for melody presentation. For each combination, the articulation formats are listed in the order that constitutes the same-, similar-, and distinct-articulation context conditions, respectively: (1) *l*, *b*, *s* and (2) *s*, *f*, *l*. These sets were created for counterbalancing purposes described in the procedure.

Apparatus Computers equipped with 16-bit sound cards were used for the experiment. Participants received the signals through a pair of Beyerdynamic DT150 headphones at approximately 70 dB SPL. The stimuli were presented using E-prime 1.2, and data were collected using the PST Serial Response Box (Schneider, Eschman, & Zuccolotto, 2002), with the left- and right-most buttons of the button-box labelled *No* and *Yes* respectively.

Design The 48 melodies were divided equally into two lists. One list was designated to consist of old melodies while the other to consist of new melodies. At study, all the 24 old melodies were presented using a single articulation format. In the test phase, the 24 new melodies were divided among three articulation formats, where eight melodies were assigned to be presented in the same format, eight in a similar format, and the remaining eight in a distinct format. For the 24 old melodies, likewise, eight were assigned to the same-articulation context condition, eight to the similar-articulation context condition, and the remaining eight to the distinct-articulation context condition (see Table 1).

Procedure Half of the participants were randomly assigned to listen to melodies played by the clarinet, while the other half were randomly allocated to listen to melodies played by the violin. The session consisted of two parts – the memorization phase and the recognition phase. The forthcoming recognition test was made known to participants before the memorization phase started. Participants were told to silently memorize each melody that was played through the headphones. At the start of each trial, a ready prompt was displayed on the monitor for one second, after which it was deleted. One second later, a melody was played over the headphones; the melody was repeated 800 ms following its first presentation. Participants then pressed the space key to proceed to listen to the next melody. This sequence persisted until all 24 melodies had been presented. The melody presentation sequence was randomized across participants.

Following the memorization phase, participants were first presented with versions of two well-known melodies – *Mary had a little lamb* and *London bridge is falling down* – that varied in their articulation formats to clarify the definition of "form". After which, the recognition test began. On each trial, the ready prompt appeared for one second and disappeared. 800 ms later, the question *Did you hear this melody in Part 1?* was displayed, and a single melody was played through the headphones. Participants were told to press the *Yes* button on the Serial Response Box if they thought they had heard the melody earlier, regardless of the original "form" (i.e., articulation format) that the melody was presented in. Otherwise, they were told to press the *No* button. Participants were told to respond as accurately as possible. No feedback was provided on any of the trials. A new trial was started after a button response.

Results and Discussion

Table 2 presents the pattern of results for d' performance across the three articulation-context conditions. There was a reliable main effect of articulation context, $F(2, 90) = 3.94$, $MSe = 0.36$, $p < .05$. Pairwise comparisons revealed that participants were significantly better at discriminating melodies presented with the same articulation format than they were at discriminating melodies presented with a distinct articulation format, $t(46) = 2.42$, $p < .05$; participants also performed better when melodies appeared in a similar articulation format than they did when melodies appeared in a distinct format, $t(46) = 2.03$, $p < .05$.

Table 1: Summary of Experiment 1's Design

Memorization	Recognition					
Study melodies	Test melodies (Old)			Test melodies (New)		
	Articulation format context					
	Same	Similar	Distinct	Same	Similar	Distinct
	Set combination 1 articulation formats					
l	l	b	s	l	b	s
24	8	8	8	8	8	8
	Set combination 2 articulation formats					
s	s	f	l	s	f	l
24	8	8	8	8	8	8

Discriminability did not differ between the same- and similar-articulation context conditions, $t < 1.05$. This pattern of results indicates that discriminability increased significantly so long as melodies were tested in at least a similar articulation format.

Table 2: Discrimination Performance (d') Across Articulation-Context Conditions in Experiment 1.

	Articulation context		
	Same	Similar	Distinct
M	0.97	0.90	0.64
SD	0.66	0.56	0.67

The present data revealed an advantage in melody recognition for same-articulation repetitions over distinct-articulation presentations. There was also an advantage in melody recognition for similar-articulation presentations over distinct-articulation presentations. An interpretation based on the the now-classic encoding specificity framework (Tulving & Thompson, 1973) is apt. Under this framework, the effectiveness of a retrieval cue depends on its degree of relatedness to the encoding of an item at first. Our view is that surface (articulation) and structural attributes of a melody are stored together in the LTM trace. Melody recognition is reliable when a specific match between the episodic memory trace and the probe occurs, but is hampered when there is a mismatch.

The comparison of shared properties between the memory trace and the probe implies that item similarity *per se* constitutes an integral part of the retrieval process. In fact, the degree of similarity among the features of the exemplar traces in memory and the target probe forms a central aspect in exemplar models of memory and categorization (Gillund & Shiffrin, 1984; Hintzman, 1988). Memory theorists have assumed that memory for a stimulus is really memory for features contained in that stimulus. The global matching approach (see Clark & Gronlund, 1996) suggests that these features in a test item, when matched with the features that have earlier been stored in memory, evoke a familiarity signal. Specifically, the greater the degree of match is, the stronger the signal will be. In our case, when a melody was re-played in the same or in a similar articulation format at test, there are many overlapping features between the articulation formats of the two melody instances from study to test. These overlaps presumably contribute to a strong sense of familiarity signal evoked by resemblance to the studied melody (see Cleary, 2004). In contrast, when the melody appeared at test in a distinct format, there are few overlapping features with the melody's original format. As such, the familiarity signal is presumably weaker, which hinders melody discrimination.

The present experiment suggests that when matching occurs, melody recognition performance is reliable at test. Experiment 2 was designed to establish an important boundary condition which determines whether this matching process would prevail (or fail).

Experiment 2

A first examination of the articulation similarity scaling solution shown in Figure 3 reveals that the greater the amount of *physical* articulation match between two instances of a melody, the more similar they were perceived to be. For instance, formats *d* and *f*, each containing two bars of *staccato* component, were perceived as similar to each other. But a closer look at the scaling solution reveals that only when the articulation format of two instances of the melody matched *at the melody's onset* would the two instances of the melody be perceived as similar to each other. This interpretation can explain why format *e* was perceived as rather different from formats *d* and *f* even though each of these formats contained two bars of *staccato* component. This observation is intriguing because two articulation formats, given the same quantitative amount of articulation match, could in fact be perceived as different from each other due to the fact that the match did not occur at the melody's onset.

We therefore pursued a third question here: Would this perceptual dissimilarity between two instances of the melody (e.g., in formats *d* and *e*) due to the location of the (mis)match hamper discrimination performance during the test stage, even when both instances contain the exact same quantity of articulation match (e.g., two bars of *staccato* component)? The goal was to illuminate the underlying nature of the matching process in melody recognition memory, and we hypothesized that to the extent that perceptual dissimilarity, as a function of the location of (mis)match in format, affects matching between study and test, discrimination performance ought to be hampered when old melodies that were originally played in, say, format *s* are repeated in format *e* (i.e., perceptually dissimilar format) at test, as compared to when the melodies are repeated in format *d* or *f* (i.e., perceptually similar format) at test, although formats *d*, *e*, and *f* each contains the exact same quantity (i.e., two bars) of *staccato* component.

Method

Participants Sixty-four psychology undergraduates participated. None had participated in Experiment 1.

Materials, Apparatus, Design, and Procedure The materials and procedures were essentially the same as those of Experiment 1, with a slight modification in materials. Based on Figure 3, four different combinations of articulation formats were selected for melody presentation. For each combination, the articulation formats are listed in the order that constitutes the same-, similar-, and distinct-articulation context conditions respectively: (1) *s*, *d*, *e*, (2) *s*, *f*, *e*, (3) *l*, *b*, *a*, and (4) *l*, *c*, *a*. Set combination was counterbalanced across participants.

Results and Discussion

Table 3 presents the pattern of results for *d'* performance across the three articulation-context conditions. There was no reliable main effect of articulation context, $F < 1.23$.

Discriminability between the same-, similar-, and distinct-articulation context conditions did not differ reliably. Articulation format did not influence performance.

Table 3: Discrimination Performance (d') Across Articulation-Context Conditions in Experiment 2.

	Articulation context		
	Same	Similar	Distinct
M	1.13	0.94	1.09
SD	0.67	0.78	0.70

Experiment 1 suggested that articulation properties are bound with the melody's structural identity. Surface feature information of the melody is first encoded and stored in the memory trace, and later used to retrieve the melody. Because a same- or similar-feature repetition constitutes an exact, or at least a close, match with the memory trace for the old melody, the trace becomes more salient than the other competing traces, enhancing discrimination performance. On the other hand, a distinct-feature presentation would not match with the trace for the old melody, thus performance is hampered. The interpretation is that given a retrieval cue that coincides with the initial encoding of the melody in terms of its surface properties, the cue would help the melody to be recovered at test.

But Experiment 2 revealed that initial perceptual (dis)similarity, as a function of the location of feature (mis)match between two instances of the melody, did not accurately determine discrimination performance. When two instances of the melody are perceived as different from each other from study to test, matching presumably would not occur. Yet, some form of matching must have occurred despite the perceptual mismatch because the overall discrimination performance (in the distinct articulation condition) was good, average $d' = 1.09$.

Values of d' between 1 and 2 usually represent good yes-no recognition performance (Neath & Surprenant, 2003, p. 202). To further justify that this was good performance, we conducted three planned comparisons on the d' data. The first and second comparisons established that the data sets between Experiments 1 and 2 were comparable: Performance in the same-articulation conditions, as well as performance in the similar-articulation conditions, across both experiments did not differ, $ts < 1.28$, $ps > .21$. The third comparison used performance in Experiment 1's distinct articulation condition as baseline, and revealed that performance in Experiment 2's distinct-articulation condition reliably exceeded performance in this baseline condition, $t(109) = 3.44$, $p < .01$, implicating good discrimination performance in this case.

Thus, the logical inference is that whether matching would occur is likely to be contingent on the absolute physical quantity of match between the memory trace and the recognition probe per se, rather than the perception of dissimilarity due to the location of (mis)match in the feature attributes. These data defined an important boundary condition of matching observed in melody recognition under which matching would (or would not) be successful.

General Discussion

Several studies have demonstrated that the alteration of the initial part of a sound can affect the recognition of musical instruments (e.g., Berger, 1964; Grey & Moorer, 1977). These findings suggest that temporal features are important in timbre perception and music processing at large. Yet, Experiment 2 suggests that altering the initial part of the articulation format (i.e., at the onset of a melody) did not influence discrimination performance. In explaining these data, we offer a global matching advantage interpretation which finds its roots in Gestalt psychology. A basic position of the Gestalt view is that a whole is qualitatively different from the complex that one might predict by considering only its parts. Under this view, wholes are organized prior to perceptual analysis of their properties and components in perceptual organization. Navon (1977) proposed that perceptual processing starts with global structuring and later moves towards more fine-grained analysis. This proposal was termed as the *global precedence hypothesis*. This hypothesis has been tested by studying the perception of hierarchical patterns in which larger figures are constructed by suitable arrangements of smaller figures.

An example is a set of large letters constructed from the same set of smaller letters having either the same identity as the larger letter or a different identity (see Figure 4). The larger letter is considered a higher-level unit relative to the smaller letters, which are, in turn, lower-level units. Properties of the higher-level unit are considered more

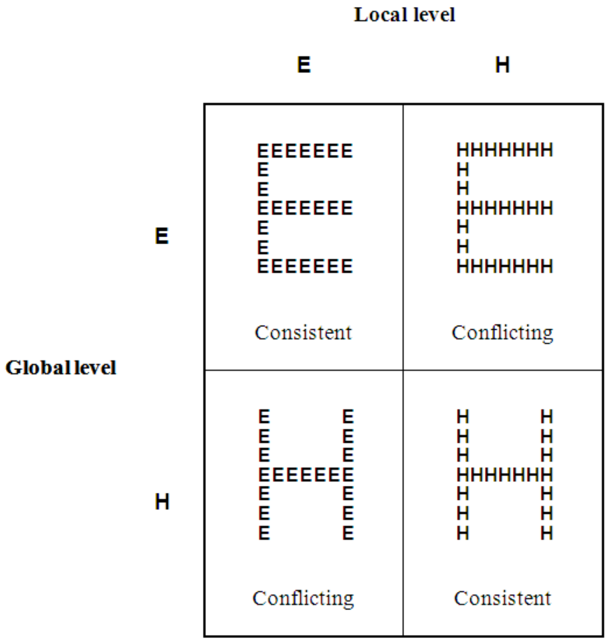

Figure 4: An example of Navon's (1977) type hierarchical stimuli. Large Es and Hs are composed using small Es and Hs.

global than properties of the lower-level units by virtue of their position in the hierarchical structure. In a typical experiment, observers are presented with such stimuli and are required to identify the larger (i.e., global) or the smaller (i.e., local) letter in different trials. *Global advantage* is observed, where the global letter is identified faster than the local letter.

Our view is that an analogous global advantage mechanism operates in the matching process found in melody recognition. The general articulation format of the melody (i.e., whether the melody is overall presented in a *staccato* or *legato* format) is considered a higher-level unit relative to the specific format of individual bars, which are, in turn, lower-level units, and properties of the higher-level unit are considered more global than properties of the lower-level (local) units based on their position in the hierarchical structure. In order for matching to occur, that there is a *global* match based on the *absolute quantity* of match between the memory trace and the recognition probe *per se* is more critical, as compared with whether there is a *local* match between the articulation format at the onset of the test melody and the format at the onset of the study melody. Once global matching attains, melody discrimination performance is enhanced.

The present global matching advantage hypothesis can be verified in a future study that manipulates the overall (global) and local matches in, say, timbre between two instances of a melody, by specifically altering the timbre at various temporal points (e.g., the onset) of the melody. Others could assess the effects of surface features that have yet to receive attention, such as the use of ornaments or phrase boundaries. More broadly, future investigations can extend to the domain of speech perception. There had been considerable work which argued for a commonality between music and speech processing (see Patel, 2003), and comparing these two processes can lead to an understanding of wider (and potentially shared) principles of perceptual categorization and temporal organization across brain areas (McMullen & Saffran, 2004; Patel, 2003). Thus, it is of interest whether the present effects would emerge in speech. There is a large body of data suggesting that talker's voice, a surface feature of spoken language, is encoded into LTM. Specifically, old words were recognized better when they were tested in a voice that matched with the original voice that originally spoke the word at study, than when the voices did not match (see Goh, 2005 for a review). Yet, the boundaries that permit (or prevent) this match in a speech context are not well defined. It is worthwhile to explore the extent to which speech recognition performance is driven by the absolute match in the physical properties of voice between two instances of speech and/or the location of match *per se* (e.g., in a sentence context).

References

Berger, K. W. (1964). Some factors in the recognition of timbre. *Journal of the Acoustical Society of America*, *36*, 1888–1891.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37-60.

Cleary, A. M. (2004). Orthography, phonology, and meaning: Word features that give rise to feelings of familiarity in recognition. *Psychonomic Bulletin & Review*, *11*, 446–451.

Gillund, G., & Shiffrin, R. M. (1984). A retrieval model for both recognition and recall. *Psychological Review, 91*, 1–67.

Goh, W. D. (2005). Talker variability and recognition memory: Instance-specific and voice-specific effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 40–53.

Grey, J. M., & Moorer, J. A. (1977). Perceptual evaluations of synthesized musical instrument tones. *Journal of the Acoustical Society of America*, *62*, 454–462.

Halpern, A. R., & Müllensiefen, D. (2008). Effects of timbre and tempo change on memory for music. *The Quarterly Journal of Experimental Psychology*, *61*, 1371–1384.

Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple trace memory model. *Psychological Review*, *95*, 528–551.

Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Newbury Park, CA: Sage.

Large, E. W., Palmer, C., & Pollack, J. B. (1995). Reduced memory representations for music. *Cognitive Science, 19*, 53–96.

McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, *21*, 289–311.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, *9*, 353–383.

Neath, I., & Surprenant, A. M. (2003). *Human memory: An introduction to research, data, and theory*. Toronto: Wadsworth.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*, 674–681.

Peretz, I., Gaudreau, D., & Bonnel, A. (1998). Exposure effects on music preference and recognition. *Memory & Cognition*, *26*, 884–902.

Raffman, D. (1993). *Language, music, and mind*. Cambridge, MA: MIT Press.

Schneider, W., Eschman, A., & Zuccolott, A. (2002). *E-Prime User's Guide*. Pittsburg: Psychology Software Tool Inc.

Trainor, L. J., Wu, L. & Tsang, C. D. (2004). Long-term memory for music: Infants remember tempo and timbre. *Developmental Science*, *7*, 289–296.

Tulving, E., & Thompson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373.