
The Quarterly Journal of Experimental Psychology

ISSN: 1747-0218 (Print) 1747-0226 (Online) Journal homepage: http://www.tandfonline.com/loi/pqje20

# Articulation effects in melody recognition memory

Stephen Wee Hun Lim & Winston D. Goh

**To cite this article:** Stephen Wee Hun Lim & Winston D. Goh (2013) Articulation effects in melody recognition memory, The Quarterly Journal of Experimental Psychology, 66:9, 1774-1792, DOI: 10.1080/17470218.2013.766758

**To link to this article:** https://doi.org/10.1080/17470218.2013.766758

Accepted author version posted online: 16 Jan 2013.
Published online: 14 Feb 2013.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY, 2013
Vol. 66, No. 9, 1774–1792, http://dx.doi.org/10.1080/17470218.2013.766758

Routledge
Taylor & Francis Group

# Articulation effects in melody recognition memory

**Stephen Wee Hun Lim and Winston D. Goh**

Department of Psychology, National University of Singapore, Singapore

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 processes to enhance recognition. Experiment 2 revealed that melodies rated as perceptually dissimilar on the basis of the location of the articulation mismatch did not impair recognition performance, suggesting an important boundary condition for articulation format effects on memory recognition—the matching of the memory trace and recognition probe may depend more on the overall proportion, rather than the temporal location, of the mismatch. The present findings are discussed in terms of a global matching advantage hypothesis.

*Keywords*: Melody recognition memory; Articulation format effects; Global matching advantage.

When we hear a piece of music, we detect and occasionally remember phrases, motifs, themes, syncopations, suspensions, tonic chords, and cadences. 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 nonstructural 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., Lim & Goh, 2012; Peretz, Gaudreau, & Bonnel, 1998) and tempo (e.g., Halpern & Müllensiefen, 2008), are also encoded into LTM during a melody recognition task. For instance, Peretz et al. (1998), in Experiment 3 of their study, investigated the effects of surface features on melody recognition,

Correspondence should be addressed to Stephen Wee Hun Lim, National University of Singapore, Faculty of Arts & Social Sciences, Department of Psychology, Block AS4, 9 Arts Link, Singapore 117570. E-mail: psylimwh@nus.edu.sg

by modifying the instruments that were used to present the melodies. Their goal was to manipulate the surface characteristics of melodies while preserving their structural identities. During the study phase, half the melodies were presented on the piano while the remaining half were presented on the flute. During the test stage, the melodies were repeated either in the same timbre (e.g., piano–piano) or with a different timbre (e.g., piano–flute). Timbre appears to be critical to music identity because participants recognized melodies significantly better when the same timbre was used during both the study and the test phases.

The encoding specificity principle (Tulving & Thompson, 1973), which posits that the effectiveness of a retrieval cue depends on its degree of relatedness to the initial encoding of an item, can explain why melodies were recognized better when they were repeated in the same timbre. Timbre information is first encoded and stored in the memory traces of the melodies and is later used to retrieve or recover the melodies. Because a same-timbre repetition is, at the same time, an exact match with the memory trace for the old melody, that trace becomes more prominent than the other competing traces. On the other hand, a melody presented by a different timbre will match the memory trace for the melody only in terms of its structural properties, and not in terms of its surface (i.e., timbre) properties. As a result, this melody should be less discriminable at test than the melody that is repeated in the same timbre. The view is that the match between the episodic memory trace and the probe determines whether a recognition advantage would emerge; the more precise the match is, the more sizeable the recognition effect would be.

The influence of surface features—such as timbre or tempo—on recognition memory for auditory stimuli has been reported outside of the melody recognition literature. For instance, in the speech perception domain, the analogous view is that nonlinguistic properties of speech, such as the speaker's voice, are not separate from linguistic content, but rather constitute an integral component of the speech and language perception process (Nygaard & Pisoni, 1998; Pisoni & Lively, 1995). These voice attributes are retained in episodic memory along with lexical information and are found to later facilitate recognition memory. Speaker information is not discarded through normalization in speech. Instead, variation in a speaker's voice actually forms part of a rich and elaborate representation of speech. The assumption is that the end product of speech perception consists of—along with abstract, context-free linguistic units—nonlinguistic units such as the speaker's voice, and both kinds of content contribute to the identification and recognition of speech.

Evidence that supports this view in speech processing has been found in several memory studies (e.g., Goldinger, 1996; Pilotti, Bergman, Gallo, Sommers, & Roediger, 2000; Sheffert, 1998), which showed that recognition accuracy at test for words or sentences repeated in the same voice surpassed recognition accuracy when words or sentences were repeated in a different voice. Although a handful of researchers did not observe this difference (e.g., Church & Schacter, 1994; Luce & Lyons, 1998), the general trend favours the position that surface feature (i.e., voice) information, along with lexical information, is encoded into LTM (see Goh, 2005, for a review).

## Articulation effects in melody recognition memory

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 integrates characteristics such as timbre and tempo with melody identity. Specifically, we were curious 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 abstract 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 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 interonset interval (IOI), as compared to when the note was played in *legato* form (i.e., IOI = 100%). 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 formats *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

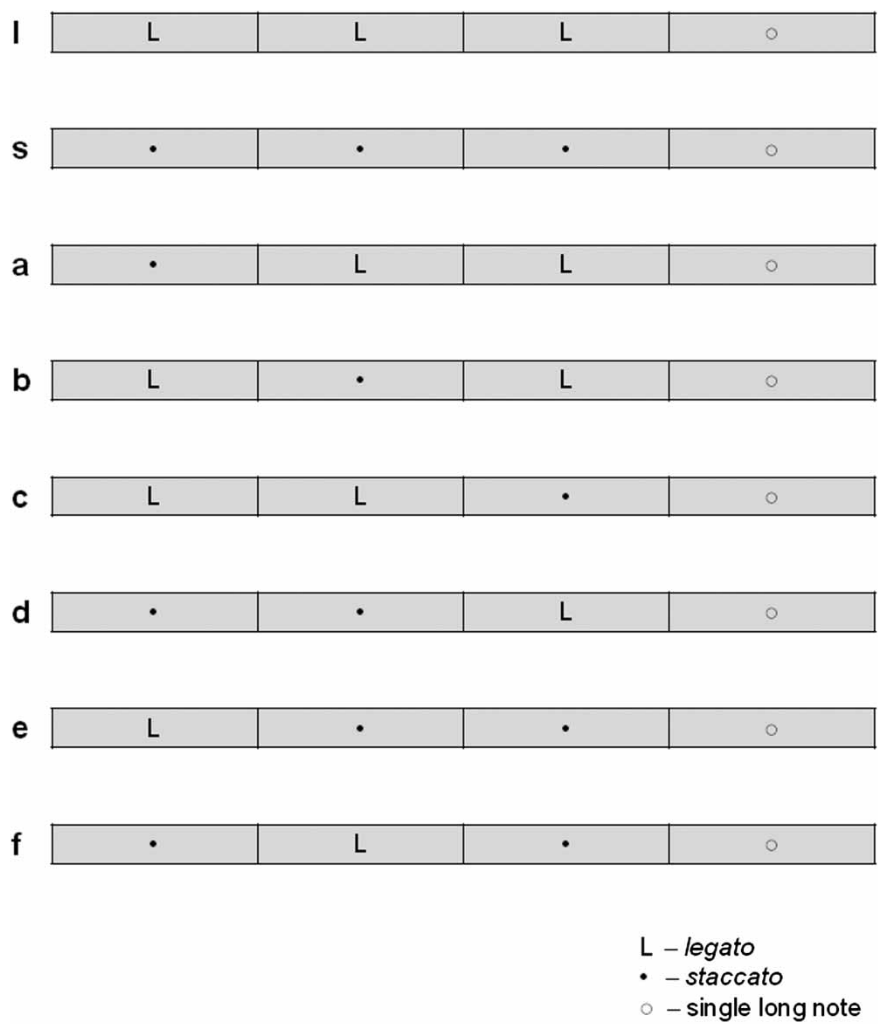


**Figure 1**. *Schematic of the eight different articulation format manipulations.*

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), 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.

## EXPERIMENT 1

In Experiment 1, we asked two questions: (a) Is articulation feature information retained in LTM, and (b) 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 of either the same or a 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. Our earlier work suggests that timbre (dis)similarity influenced melody recognition performance (Lim & Goh, 2012). A second goal of the present study was thus to ascertain 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[1] with varying music training experience (15 with at least four years of formal music training versus 32 without or with less than four years of music training) from the National University of Singapore 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 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 double or simple triple time, lasting approximately 6 seconds or 7.2 seconds, respectively. The melodies were constructed using the *Finale 2009* software and were saved as .wav sound files.

Among the eight different articulation formats designed for this study (see Figure 1), it is logical that a melody presented fully in *legato* form (format *l*) ought to be perceived as maximally distinct from the same melody presented fully in *staccato* form (format *s*), since these are two contrasting articulation formats. In other words, we

[1] Undergraduate students at the National University of Singapore are aged between 18 and 25 years, although we did not collect these data explicitly.


**Figure 2**. *Samples of the melodies used for the present study.*

assumed perceived articulation similarity to be a function of the absolute amount of match in articulation format between the two instances of the melody. For example, articulation formats *d*, *e*, and *f* would be perceived as similar to each other because they each contained the same quantity (i.e., two bars) of the *staccato* component, whereas articulation format *a* would be perceived as somewhat different from format *d* because the latter contained a greater measure of *staccato* than the former. Prior to conducting Experiment 1, this assumption was verified by first constructing a multidimensional "articulation map" 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 use in the subsequent main experiments can be based on objective measures of the degree of perceived similarity among different articulation formats. The steps taken to collect similarity ratings and the generation of the "articulation map" using multidimensional scaling (MDS) techniques (Kruskal & Wish, 1978) are described in the Appendix. The two-dimensional MDS solution for the eight articulation formats appears in Figure 3.

After the scaling solution was derived, object coordinates in the space provided the basis to approximate[2] the perceptual distances between all

[2] The approximations were derived with the Euclidean geometric equation for distance between two points in a plane: distance = $\sqrt{[(x_1 - x_2)^2 + (y_1 - y_2)^2]}$, in which $(x_1 y_1)$ and $(x_2 y_2)$ are planar coordinates for Points 1 and 2, respectively.

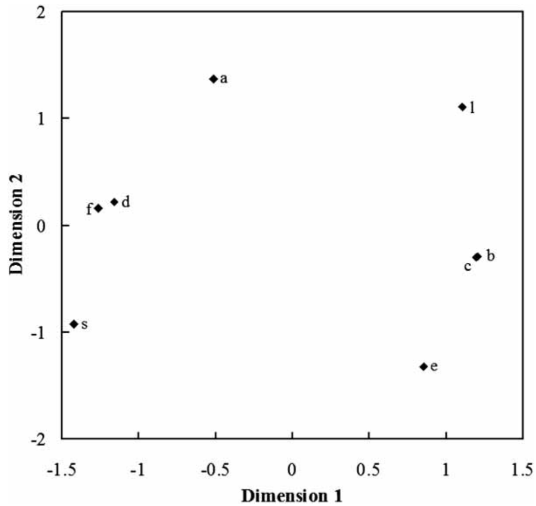


**Figure 3**. *Two-dimensional multidimensional scaling (MDS) solution for the eight articulation formats.*

**Table 1.** *Two set combinations of articulation formats derived for melody presentation at test in Experiment 1*

| | *Articulation format context* | | |
|---|---|---|---|
| *Set combination* | *Same* | *Similar* | *Distinct* |
| 1 | l | b | s |
| 2 | s | f | l |

articulation formats. 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: (Combination 1) *l*, *b*, *s* and (Combination 2) *s*, *f*, *l* (see Table 1). 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 presented through a pair of Beyerdynamic DT150 headphones at approximately 70 dB sound pressure level. The stimuli were presented using E-prime 1.2, and data were collected using the PST Serial Response Box (Schneider, Eschman, & Zuccolott, 2002), with the left- and rightmost 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 was designated to consist of new melodies. At study, all the 24 old melodies were presented using a single articulation format, enabling participants to study and memorize (the abstract structure of) the melodies effectively. In the test phase, for each individual participant, the 24 new melodies were randomly divided among three articulation formats, where 8 melodies were assigned to be presented in the same format, 8 in a similar format, and the remaining 8 in a distinct format. For the 24 old melodies, likewise, 8 were randomly assigned to the same-articulation context condition, 8 to the similar-articulation context condition, and the remaining 8 to the distinct-articulation context condition (see Table 2 for a summary).

### *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. Figure 4 shows the schematic of the sequence of a trial. Participants were tested individually or in small groups of seven or fewer. 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. The memorization phase lasted approximately five minutes. 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.

**Table 2.** *Summary of the design used in Experiment 1*

| *Articulation formats* | *Memorization stage: Study melodies* | *Recognition stage* | | | | | |
|---|---|---|---|---|---|---|---|
| | | *Test melodies (old)* | | | *Test melodies (new)* | | |
| | | *Articulation format context* | | | | | |
| | | *Same* | *Similar* | *Distinct* | *Same* | *Similar* | *Distinct* |
| Set Combination 1 | l | l | b | s | l | b | s |
| | 24 | 8 | 8 | 8 | 8 | 8 | 8 |
| Set Combination 2 | s | s | f | l | s | f | l |
| | 24 | 8 | 8 | 8 | 8 | 8 | 8 |

*Note:* Numbers indicate the quantity of melodies played in the respective formats in each classification.

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 "articulation format". After this, the recognition test began. On each trial, the ready prompt appeared for one second and disappeared. After 800 ms, 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. The computer recorded response accuracy. No feedback was provided on any of the trials. A new trial was started after a button response. It took approximately 10 min to complete all 48 randomly presented trials.

After the recognition test, information on the participants' music training experience was captured. The question *How many years of formal music training, in total, have you undergone?* was displayed; participants pressed *1* on the keyboard if they had none, or less than four years of training, and they pressed *2* on the keyboard if they had undergone at least four years of training. At the end of the session, participants were debriefed.

## Results

The measure $d'$ was used to assess the accuracy of melody discrimination performance. Participants discriminated between old and new melodies, regardless of their articulation format at test.

The effects of musical training were peripheral to the study's goals, which was why we had not systematically controlled the number of participants with or without formal musical training. Nevertheless, we always examined potential effects of musical training before proceeding to the main analyses for the sample. In addition, each participant heard the melodies played in either of two instruments (clarinet versus violin); our goal was to demonstrate that articulation context effects, if any, ought to persist across instrument types. Accordingly, a three-way mixed-design analysis of variance (ANOVA), with instruments (clarinet, $n = 24$, versus violin, $n = 23$) and musical training (participants who had at least four years of formal music training, $n = 15$, versus those without or with less than four years of music training, $n = 32$) as the between-subjects factors, and articulation context as the

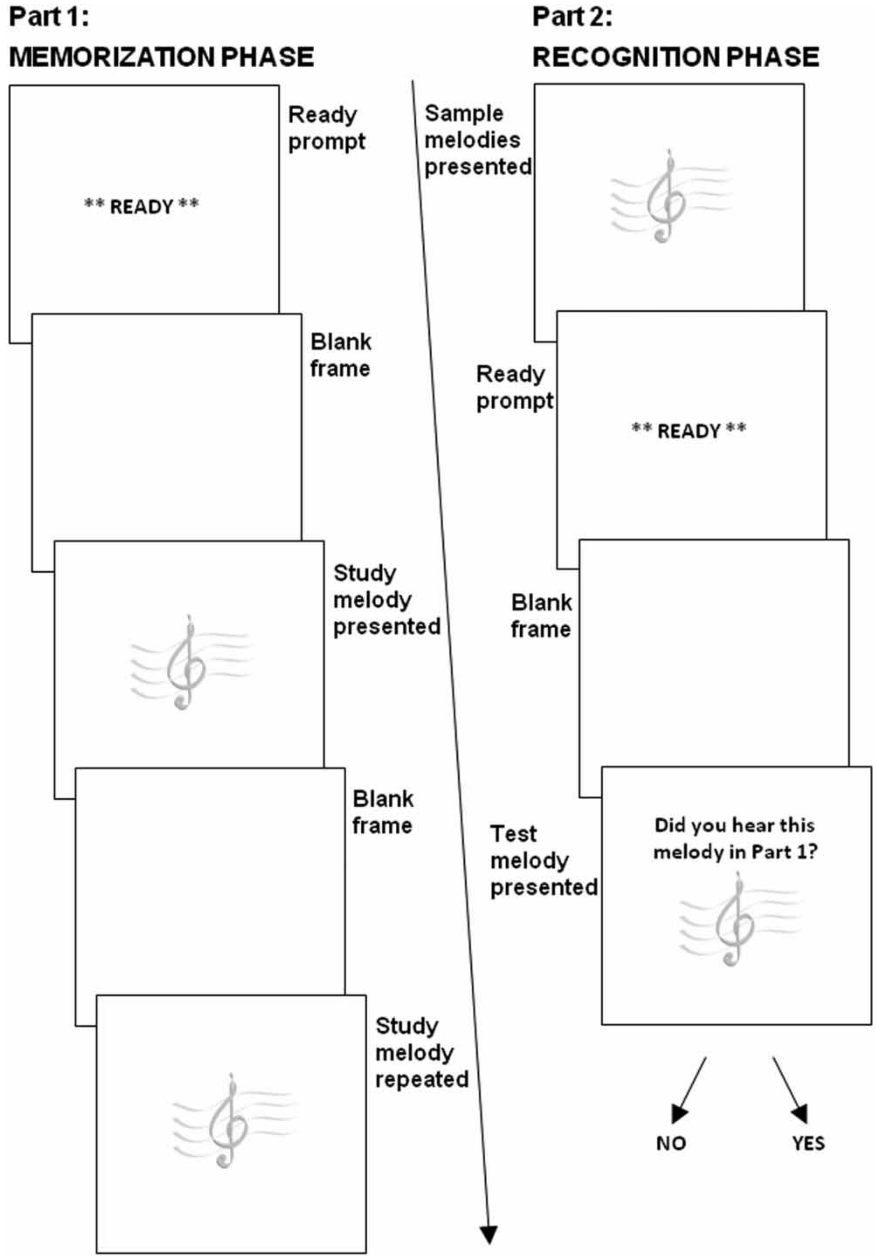


**Figure 4**. *Schematic of the sequence of a trial in Experiment 1.*

within-subjects factor, was conducted. The outcome of these analyses is reported before we consider the main findings involving the three articulation-context conditions.

*Influences of musical training and instruments*

No three-way interaction was observed. More important, there were no significant interactions between musical training and articulation context, and between instruments and articulation context, both $F$s < 1. This shows that neither musical training nor instruments influenced the articulation effects. There were also no significant main effects of musical training or instruments, both $F$s < 1. Since neither musical training nor instruments interacted with articulation context, the main findings for the articulation context conditions reported below can be generalized across all participants and instruments used within the sample, and all subsequent tabulations of results are collapsed across musical training and instruments.

### *Effects of articulation context*

Table 3 presents the pattern of results for $d'$ performance across the three articulation-context conditions by participants and items. There was a significant main effect of articulation context by participants, $F_p(2, 90) = 3.94$, $MSE = 0.36$, $p < .05$, and by items, $F_i(2, 94) = 3.18$, $MSE = 0.36$, $p < .05$. Pairwise comparisons (with Bonferroni correction) 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, $t_p(46) = 2.42$; $t_i(47) = 2.19$, $p$s < .05; participants also performed better when melodies appeared in a similar articulation format than they did when melodies appeared in a distinct format, $t_p(46) = 2.03$; $t_i(47) = 2.00$, $p$s < .05. Discriminability did not differ between the same- and similar-articulation context conditions, all $t$s < 1.05. This pattern of results indicates that discriminability increased significantly so long as melodies were tested in at least a similar articulation format.

## Discussion

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 encoding specificity principle (see introduction) offers a good fit with these data. Recall that based on encoding specificity, whether the memory probe serves as an effective retrieval cue depends on how specifically it coincides with the initial encoding of the melody. 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. This idea is compatible with the feeling-of-knowing (FOK) phenomenon (see

**Table 3.** *Discrimination performance ($d'$) across articulation-context conditions in Experiment 1*

| | *Articulation context* | | | | | |
|---|---|---|---|---|---|---|
| | *Same* | | *Similar* | | *Distinct* | |
| $d'$ | M | SD | M | SD | M | SD |
| By participants | 0.97 | 0.66 | 0.90 | 0.56 | 0.64 | 0.67 |
| By items | 1.17 | 0.86 | 1.14 | 0.79 | 0.80 | 0.86 |

Hart, 1965), which instantiates a dissociation between objective and subjective indices of knowing. For instance, while a person fails to recall an item, he/she is often confident that the particular item is really available in the memory store, which can be recalled or recognized in the (near) future. In a typical FOK judgement task, participants are given a memory cue and are asked to recall the corresponding target from memory. When they fail to retrieve the target, they proceed to make an FOK judgement regarding the likelihood of identifying the correct target among several distractors that appear in a subsequent recognition test. The basis of FOK judgements has been tied to the implicit application of global heuristics (see Koriat, 2000; Koriat & Levy-Sadot, 1999; see also Koriat & Levy-Sadot, 2001, for a fuller account of FOK processes).

In our case, when a melody was replayed 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, and melody discrimination is hindered.

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 that determines whether this matching process would prevail (or fail).

## EXPERIMENT 2

A first examination of the articulation similarity scaling solution shown in Figure 1 reveals that the greater the amount of 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. Similarly, formats *b* and *c*, each containing two bars of *legato* component, were perceived as similar to each other. Experiment 1 examined whether articulation, and articulation similarity, information is retained in melody recognition memory and demonstrated that melody recognition is effective when there is generally a fit between two formats. Specifically, when two instances of a melody were perceptually similar in terms of their *physical* match in articulation format from study to test, matching obtained. As a result, melody recognition performance was enhanced.

But a closer look at the scaling solution reveals that the *location* of the match (or mismatch) was apparently important in determining whether two instances of the same melody would be perceived as similar to each other. Specifically, 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)? Addressing this question would illuminate the underlying nature of the matching process in melody recognition memory. Our data based on Experiment 1 suggest that to the extent that two instances of a melody are perceptually similar to each other, matching occurs. While the absolute amount of match in articulation format appears to determine this perceptual similarity and, consequently, discrimination performance, the goal of Experiment 2 was to verify the influence of perception as a function of *location* of (mis)match on discrimination performance. The critical hypothesis was that to the extent that perceptual dissimilarity, as a function of the location of (mis)match in articulation 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, even though formats *d*, *e*, and *f* each contains the exact same quantity (i.e., two bars) of *staccato* component.

## Method

### *Participants*

Sixty-four psychology undergraduates with varying music training experience (20 with at least four years of formal music training versus 44 without or with less than four years of music training) from the National University of Singapore participated in the experiment. 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 the Euclidean estimates of the articulation similarity scaling solution in 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: (Combination 1) *s*, *d*, *e*, (Combination 2) *s*, *f*, *e*, (Combination 3) *l*, *b*, *a*, and (Combination 4) *l*, *c*, *a* (see Table 4). Set combination was counterbalanced across participants. Table 5 shows a summary of the present design.

**Table 4.** *Four set combinations of articulation formats derived for melody presentation at test in Experiment 2*

| | *Articulation format context* | | |
|---|---|---|---|
| *Set combination* | *Same* | *Similar* | *Distinct* |
| 1 | s | d | e |
| 2 | s | f | e |
| 3 | l | b | a |
| 4 | l | c | a |

## Results

As with Experiment 1, $d'$ assessed the accuracy of melody discrimination performance, and before proceeding to the main findings involving the three articulation-context conditions, we first examined the potential effects of musical training and instruments for the sample, with musical training (participants who had at least four years of formal music training, $n = 20$, versus those

**Table 5.** *Summary of the design used in Experiment 2*

| | | *Recognition stage* | | | | | |
|---|---|---|---|---|---|---|---|
| | | *Test melodies (old)* | | | *Test melodies (new)* | | |
| | | *Articulation format context* | | | | | |
| *Memorization stage: Examples of study melodies* | | *Same* | *Similar* | *Distinct* | *Same* | *Similar* | *Distinct* |
| Set Combination 1 articulation formats | s | s | d | e | s | d | e |
| | 24 | 8 | 8 | 8 | 8 | 8 | 8 |
| Set Combination 3 articulation formats | l | l | b | a | l | b | a |
| | 24 | 8 | 8 | 8 | 8 | 8 | 8 |

*Note:* Numbers indicate the quantity of melodies played in the respective formats in each classification.

without or with less than four years of music training, $n = 44$) and instruments (clarinet, $n = 32$, versus violin, $n = 32$) as the between-subjects factors and articulation context as the within-subjects factor.

### *Influences of musical training and instruments*

No three-way interaction was observed. More important, there were no significant interactions between musical training and articulation context and between instruments and articulation context, $F\text{s} < 1.60$. This shows that neither musical training nor instruments influenced the articulation effects. No main effect of instruments was found, $F < 1$, although a main effect of musical training emerged, $F(1, 62) = 4.68$, $MSE\text{s} = 0.63$, $p < .05$. Participants with at least four years of formal musical training had higher discrimination scores ($M = 1.24$, $SD = 0.81$) than those without or with less than four years of musical training ($M = 0.97$, $SD = 0.66$). Since neither musical training nor instruments interacted with articulation context, the main findings for the articulation context conditions reported below can be generalized across all participants and instruments used within the sample, and all subsequent tabulations of results are collapsed across musical training and instruments.

### *Effects of articulation context*

Table 6 presents the pattern of results for $d'$ performance across the three articulation-context conditions by participants and items. There was no significant main effect of articulation context by participants and by items, all $F\text{s} < 1.23$. Discriminability between the same-, similar-, and distinct-articulation context conditions did not differ significantly. Articulation format did not influence performance.

## Discussion

The present data revealed that discrimination performance was comparable across the same-, similar-, and distinct-articulation conditions. Importantly, perceptual dissimilarity, as a function of the location of (mis)match in articulation format, did not appear to modulate melody recognition performance. For instance, even though formats *d* and *e* were *perceived* as different from each other due to the fact that they differed in articulation form at the melody's onset, discrimination *performance* was found to be comparable across both of these conditions. The present evidence suggests that so long as both formats of the melody contained the same *quantity* of articulation match (e.g., two bars of *staccato* component), discrimination performance would be comparable across both of these instances despite the *location* of the (mis)match per se.

The disparity in the effects of matching location between the similarity judgement task and the recognition memory task is intriguing. It appears that listeners pay attention to the temporal structure of the articulation pattern when making judgements of perceptual similarity. This may be possible because the task essentially requires immediate comparison in short-term memory (STM), given that the interval between the two instances of the melody is only 500 ms apart, well within the duration of auditory sensory memory (Baddeley, 1998, p. 20). As such, the STM traces are likely

**Table 6.** *Discrimination performance (d′) across articulation-context conditions in Experiment 2*

| | *Articulation context* | | | | | |
|---|---|---|---|---|---|---|
| | *Same* | | *Similar* | | *Distinct* | |
| *d′* | M | SD | M | SD | M | SD |
| By participants | 1.13 | 0.67 | 0.94 | 0.78 | 1.09 | 0.70 |
| By items | 1.21 | 0.72 | 1.12 | 0.64 | 1.15 | 0.69 |

to contain detailed information about the temporal order of the articulation pattern, and participants could be making perceptual similarity judgements on the basis of when a change in the melody is detected. This would be consistent with the MDS solution indicating that the onset of the melody is important.

On the other hand, the recognition task relies primarily on LTM traces of the studied melodies. The present findings suggest that the LTM traces may not encode the finer details of the temporal order of the articulation structure, and hence old/new judgements may be made only on the basis of the overall proportion of legato/staccato components in the melody.

## GENERAL DISCUSSION

Experiment 1 of the present study makes a specific contribution towards the extant literature, by demonstrating that articulation properties are not erased from melody recognition memory, but are in fact integrated with the melody's structural identity. The interpretation is that when matching occurs between two instances of a melody, melody recognition is reliable. Surface feature information of the melody is first encoded and stored in the memory trace and is 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. As such, discrimination performance at test is enhanced. On the other hand, a distinct-feature presentation would not match with the trace for the old melody, and 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 property, the cue would be effective in allowing 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. Specifically, the data revealed that for explicit similarity ratings, the position of the articulation format is important, but for recognition memory, it is not. 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 $d'$ (by participants) = 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 by participants. 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, $t$s $< 1.28$, $p$s $> .21$. 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 significantly exceeded performance in this baseline condition, $t(109) = 3.44$, $p < .01$, implicating good discrimination performance in this case. These comparisons were repeated on the $d'$ data by items, and the same pattern emerged.

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.

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; Clark, Robertson, & Luce, 1964; Grey & Moorer, 1977; Saldanha & Corso, 1964; Wedin & Goude, 1972). These findings suggest that temporal features are important in musical timbre perception and processing. The present finding from Experiment 2 is intriguing because it appears 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 5). 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

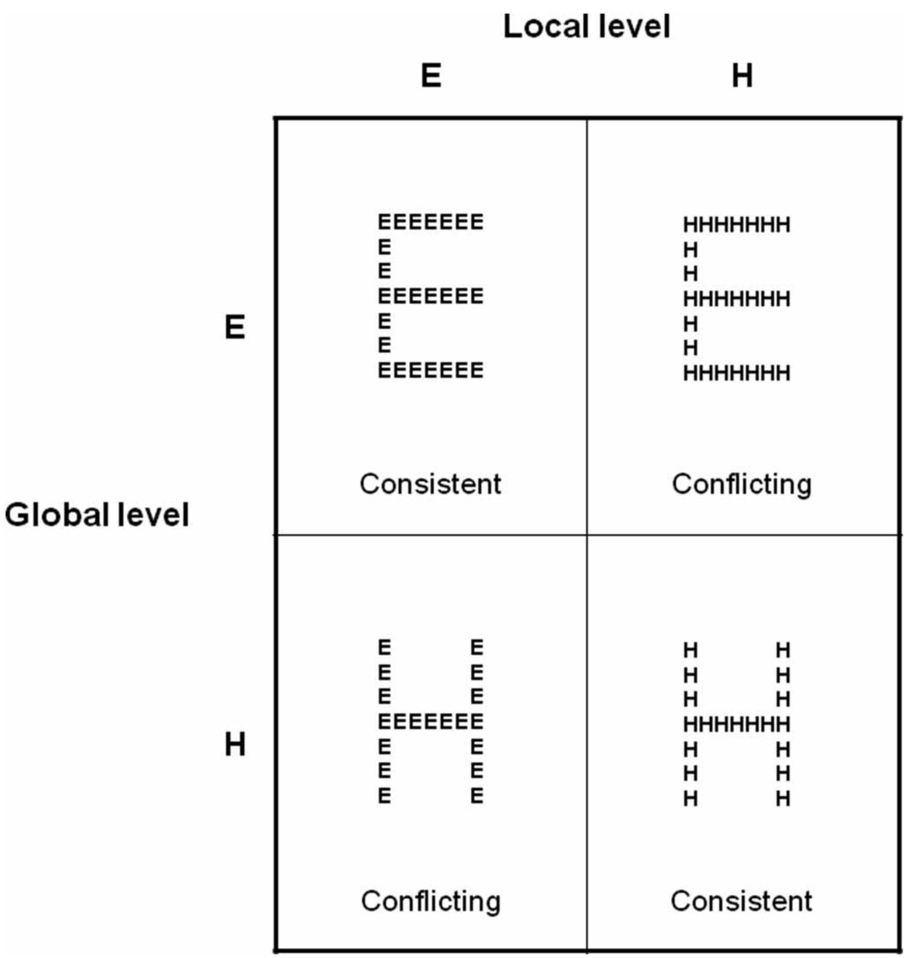


**Figure 5**. *An example of Navon's (1977) type hierarchical stimuli. Large Es and Hs are composed using small Es and Hs.*

more 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 than 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. Clearly, primacy of the (mis)match, while it plays a crucial role in influencing similarity judgements earlier, does not affect discrimination performance. Rather, once global matching attains, melody discrimination performance is enhanced.

## Global versus local aspects of melody recognition memory

The global advantage observed in the present study is well positioned within the context of that previously discovered regarding melody recognition that encompasses global and local aspects of memory (e.g., Dowling, 1972, 1978; Dowling & Fujitani, 1971). Two types of features or properties have been identified as functionally important in the processing of sequential pitch patterns: the pitch of individual tones or the interval size between two adjacent tones, or the contour that characterizes pitch directions independently of the previous pitch values (see Dowling, 1982, and Pick, 1979, for reviews). These two types of features have been classified as local and global, respectively (Dowling, 1982; Laberge, 1981).

For instance, Dowling and Fujitani (1971, Experiment 1) used a short-term recognition memory paradigm in which a randomly generated five-note melody with small pitch intervals between successive notes was used as the standard stimulus. The comparison stimulus followed shortly after the standard stimulus was played. Three types of comparison melodies were used. Specifically, the comparison melody was the same as the standard in both contour and pitch intervals, or had the same contours but different intervals, or was different in both contour and intervals (i.e., consisted of novel random sequences). While it was easy to identify (and reject) any comparison stimulus that did not contain exactly the same pitches as the standard, of particular interest is that when the comparison melodies were transposed, exact-same targets and same-contour comparisons were easily distinguishable from random ones, but almost impossible to tell apart. Listeners found it difficult to distinguish between two melodies with the same contour (global information) but different interval sizes (local information), performing at around chance level; they responded on the basis of the presence or absence of the "global" contour and were unable to recognize the sameness of "local" intervals, suggesting that global information takes precedence in melody recognition.

Notwithstanding, the importance of the global contour relative to the local intervals depends on factors such as time constraints, novelty of the material, and application of transposition (see Peretz, 1990, for a discussion). It appears, nevertheless, that the most immediate form in which melodies are encoded and/or retrieved from memory when first heard is in terms of its contour, implicating the crucial role of global information in aiding melody recognition (Edworthy, 1985).

## Future directions

The present global matching advantage hypothesis can be tested and verified further in a future study with a different design that utilizes different degrees of legato versus staccato articulation—for example, moving all the way from 10% (very staccato) to 100% (purely legato) of the IOI, with intermediate articulations at 40%, 60%, and 80%; all notes played with a duration equivalent to 40% of the IOI would be compared to another version of the same melody with all notes played with a duration equivalent to, say, 60%. This design would ascertain further whether the explanation that we offered in the present project based on the "absolute physical quantity of match" holds with these "intermediate articulations". A future study could also manipulate the overall (global) and local matches in, for instance, timbre between two instances of a melody, by specifically altering the timbre at various temporal points (e.g., the onset) of the melody. Studies henceforth could also assess the role of surface features that have yet to receive attention, including the use of accents, ornaments, melodic phrasing and phrase boundaries, or time manipulations such as *rubato* (i.e., free time), in influencing melody recognition. In addition, while the present melodies were tonal based with conjunct musical lines, future work could investigate whether the surface feature effects that emerged in this study are robust even with modal or atonal melodies, which consist of disconnected or disjointed intervallic leaps between adjacent notes. These extensions can potentially provide converging evidence to explicate more fully the principal finding that variability in surface attributes serves an indispensable function in music perception and processing.

A broader approach would involve an extension of future investigations to the domain of speech perception. There had been considerable work that 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 that suggest that the speaker'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 voice that originally spoke the word at study than when they were tested in a voice that did not match (e.g., Goh, 2005; Goldinger, 1996, 1998; Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994; Pilotti et al., 2000; Pisoni, 1997; Sheffert, 1998). Yet, the boundaries that permit (or prevent) this match in a speech context are not fully 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).

Original manuscript received 19 June 2012
Accepted revision received 4 January 2013
First published online 15 February 2013

## REFERENCES

Baddeley, A. D. (1998). *Human memory: Theory and practice* (Rev. ed.). Needham Heights, MA: Allyn & Bacon.

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

Church, B., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Memory for voice 0intonation and fundamental frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20,* 521–533.

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.

Clark, M., Robertson, J. P., & Luce, D. (1964). A preliminary experiment on the perceptual basis for musical instrument families. *Journal of the Audio Engineering Society*, *12,* 199–203.

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.

Dowling, W. J. (1972). Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics*, *12,* 417–421.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85,* 341–354.
Dowling, W. J. (1982). Melodic information processing and its development. In D. Deutsch (Ed.), *The psychology of music* (pp. 413–429). New York: Academic Press.
Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America*, *49,* S24–S31.
Edworthy, J. (1985). Melodic contour and musical structure. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 169–188). London, UK: Academic Press.
*Finale* [Software]. (2009). Eden Prairie, MN: MakeMusic, Inc.
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.
Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22,* 1166–1183.
Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105,* 251–279.
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.
Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, *56,* 208–216.
Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple trace memory model. *Psychological Review*, *95,* 528–551.
Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition*, *9,* 149–171.
Koriat, A., & Levy-Sadot, R. (1999). Processes underlying metacognitive judgments: Information-based and experience-based monitoring of one's own knowledge. In S. Chaiken & Y. Trope (Eds.), *Dual-process theories in social psychology* (pp. 483–502). New York, NY: Guilford Press.
Koriat, A., & Levy-Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27,* 34–53.
Kruskal, J. B., & Wish, M. (1978). *Multidimensional scaling*. Newbury Park, CA: Sage.
LaBerge, D. (1981). Perceptual and motor schemas in the performance of music. In J. A. Mason (Ed.), *Documentary report of the Ann Arbor Symposium* (pp. 68–76). Reston, VA: Music Educators National Conference.
Large, E. W., Palmer, C., & Pollack, J. B. (1995). Reduced memory representations for music. *Cognitive Science*, *19,* 53–96.
Lim, S. W. H., & Goh, W. D. (2012). Variability and recognition memory: Are there analogous indexical effects in music and speech? *Journal of Cognitive Psychology*, *24,* 602–616.
Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory and Cognition*, *26,* 708–715.
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, Canada: Wadsworth.
Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60,* 355–376.
Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5,* 42–46.
Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6,* 674–681.
Peretz, I. (1990). Processing of local and global musical information by unilateral brain-damaged patients. *Brain*, *113,* 1185–1205.
Peretz, I., Gaudreau, D., & Bonnel, A. (1998). Exposure effects on music preference and recognition. *Memory & Cognition*, *26,* 884–902.
Pick, A. D. (1979). Listening to melodies: Perceiving events. In A. D. Pick (Ed.), *Perception and its development: A tribute to Eleanor J. Gibson* (pp. 145–165). Hillsdale, NJ: Laurence Erlbaum Associates.

Pilotti, M., Bergman, E. T., Gallo, D. A., Sommers, M., & Roediger, H. L., III (2000). Direct comparison of auditory implicit memory tests. *Psychonomic Bulletin & Review*, *7,* 347–353.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego, CA: Academic Press.

Pisoni, D. B., & Lively, S. E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Ed.), *Speech perception and linguistic experience* (pp. 433–459). Baltimore, MD: York Press.

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

Saldanha, E. L., & Corso, J. F. (1964). Timbre cues and the identification of musical instruments. *Journal of the Acoustical Society of America*, *36,* 2021–2026.

Schneider, W., Eschman, A., & Zuccolott, A. (2002). *E-Prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Sheffert, S. M. (1998). Contributions of surface and conceptual information to recognition memory. *Perception & Psychophysics*, *60,* 1141–1152.

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.

Wedin, L., & Goude, G. (1972). Dimension analysis of the perception of instrumental timbre. *Scandinavian Journal of Psychology*, *13,* 228–240.

# APPENDIX

## Articulation similarity scaling

## Method

### *Participants*

Sixteen introductory psychology students from the National University of Singapore participated for course credit.

### *Materials*

The stimulus set comprised four melodies selected from the present database of newly composed 48 melodies used in Experiments 1 and 2. Each of the four melodies was composed in C major, C minor, G major, or G minor, respectively. Two of the melodies were written in simple triple while the other two were written in simple duple time. The instruments used to present the melodies were violin and clarinet.

### *Apparatus*

The equipment was the same as that used in the main experiments, except that the similarity ratings were collected using the computer keyboard. Keys 1, 3, 5, and 7 were labelled *very dissimilar*, *dissimilar*, *similar*, and *very similar*, respectively.

### *Design and procedure*

Participants were tested individually or in small groups of seven or fewer. The session was divided into two segments, each containing two parts. The primary purpose of this pilot study was to establish the degree of perceived similarity among different articulation formats; thus the same timbre was used to present the different formats throughout each segment. During the first segment, the first part was a two-minute familiarization phase to familiarize the participants with the eight different articulation forms that they would be rating. During this phase, participants were assigned to listen to the same melody presented in eight different articulation formats in a random order, played by either the violin or the clarinet. The allocation of the instruments for melody presentation in this segment and the subsequent segment was counterbalanced across participants.

No ratings were collected during the familiarization phase; participants were told to simply listen to the various forms of the melody. On each trial, a single melody was played by a particular instrument over the headphones, after which participants pressed the space key to proceed to listen to the next variation (in articulation format) of the same melody. This sequence continued until all eight articulation forms were presented. The articulation form presentation sequence was random across participants. Participants were informed of a forthcoming similarity rating task.

The second part was the similarity rating phase that took approximately 10 min to complete. At the start of each trial, the question *How similar are the two instances of the melody?* was displayed on the monitor. Two instances of the same melody that differed in articulation format were then presented in the same timbre, with an interval of 500 ms between the two instances. After participants pressed a button to indicate their similarity rating, the question on the monitor was erased, and a new trial began. The software controlling the experiment was written to ensure that button presses made before the onset of the second instance of each pair were not admissible. Presentation of the pairwise comparisons was randomized, and the instrument presentation order within each pair was counterbalanced across participants. Each participant was allowed to take a short break after 14 trials, after which they rated the remaining 14 trials for a total of 28 pairwise comparisons.

The procedure for the second segment of the session was virtually identical to that for the first, except that the other timbre (i.e., the timbre that was not used during the first segment earlier) was now used to present the melody throughout both Part 1 (familiarization phase) and Part 2 (rating phase), and an alternative melody that differed in metre and tonality was now used. The whole session lasted approximately 25 minutes. Participants were debriefed at the end of the session.

## Results and discussion

These perceptual similarity data were analysed by MDS using the ALSCAL routine of SPSS Version 16. The standard recommendation for MDS analyses is that the number of objects being scaled should be at least four times the number of dimensions to be derived (Kruskal & Wish, 1978). As such, because there were eight articulation forms, a two-dimensional scaling solution was derived.

In MDS, Kruskal's stress values, a goodness-of-fit statistic, range from 1.0 to 0.0, with smaller values indicating a good fit of the derived solution to the data. The stress value obtained here was .15. $R^2$, the amount of variance of the scaled data accounted for by their corresponding distances, was .85.

It should be noted that a definitive determination of the dimensions is not directly critical for the main experiments reported in this project. The primary objective of deriving the MDS solution of articulation similarity was to provide a principled basis for determining the articulation formats that would be used in the experiments. However, an interesting observation from the MDS solution was that perceived articulation similarity, contrary to the initial prediction, did not appear to be a mere function of the absolute quantity of articulation format match (or mismatch) between the two instances of the melody per se. The unanticipated finding was that format *a* was perceived as quite different from formats *b* and *c*, even though each of these three formats contains the exact same quantity of articulation match (e.g., two bars of *legato* component). In the same vein, format *e* was perceived as quite different from formats *d* and *f* even though each of these three formats contains two bars of *staccato* component (see Figure 1 and Dimension 1 of Figure 3).

We therefore attempted to determine the possible dimensions that participants could be using when making the similarity judgements. Primarily, the greater the amount of 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; formats *b* and *c*, each containing two bars of *legato* component, were perceived as similar to each other. Yet, perceived similarity seemed to be more than a simple function of the quantity of match. For instance, format *e* was perceived as somewhat different from formats *d* and *f* even though each of these formats contained two bars of *staccato* component. Specifically, it would appear that the *location* of the match (or mismatch) was in fact important in determining whether two instances of the same melody would be perceived as similar to each other. Location of the articulation match was mapped onto Dimension 1 of the articulation map. The interpretation is that to the extent that the articulation format of two instances of the melody matched *at the melody's onset*, the two instances of the melody tended to be perceived as similar to each other. Here, formats *l*, *b*, *c*, and *e*, and formats *a*, *d*, *f*, and *s*, were perceived as two groups of similar articulation formats, respectively: The former group consists of formats in which a melody would begin in *legato* style, whereas the latter consists of formats in which a melody would begin in *staccato* style. This interpretation accommodates format *e*'s perceptual dissimilarity from formats *d* and *f*. The nature of the second dimension appeared to be less definitive.