AN INVESTIGATION OF SURFACE CHARACTERISTIC EFFECTS IN MELODY RECOGNITION

LIM WEE HUN, STEPHEN

(B.Soc.Sci. (Hons.), NUS)

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF PSYCHOLOGY

NATIONAL UNIVERSITY OF SINGAPORE

2009

Acknowledgements

To the following persons I am truly grateful

Associate Professor Winston D. Goh, whose dedication made my stint as a doctoral student a most memorable one.

My parents and siblings Dr. Eldin Lim and Miss Lim Wan Xuan, for loving and accepting me as who I am.

Ms. Khoo Lilin and Mr. James Ong, whose prayers and encouragement kept me persevering, and Mr. Stephen Tay, for making the additional difference.

Ms. Loh Poh Yee and Mr. Eric Chee, whose kindness in providing extensive administrative advice and support warmed my heart.

Every volunteer, who cared to come and participate in my study.

Poohly, Tatty, and Lambmy-Hondi, for being there.

My Lord Jesus, for His grace and faithfulness.

Stephen Lim
17 August 2009

Table of Contents

Summary

Music comprises two types of information – *abstract structure* and *surface characteristics*. While a representation of the abstract structure allows a melody to be recognized across different performances, surface characteristics shape the unique expression of the melody during each performance. Very often, these surface characteristics grab our attention, but to what extent are they represented and utilized in memory?

Four main experiments were conducted to determine if information about surface characteristics, specifically timbre and articulation attributes, is encoded and stored in long-term memory, and how these performance attributes influence discrimination performance during melody recognition. The nature of timbre effects in recognition memory for melodies played by multiple instruments was investigated in Experiments 1 and 2. The first experiment investigated whether timbre-specific familiarity processes or instance-specific matching processes, or both types of processes, govern the traditional timbre effects found in melody recognition memory. Melodies that remained in the same timbre from study to test were recognized better than were melodies that were presented in a previously studied but different, or previously unstudied (new) timbre at test. Recognition for melodies that were presented in a different timbre at test did not differ reliably from recognition for

melodies in a new timbre at test. Timbre effects appear to be attributed solely to instance-specific matching processes.

The second experiment assessed the contribution of timbre similarity effects in melody recognition. Melodies that remained in the same timbre from study to test were recognized better than were melodies that were presented in a distinct timbre at test. But when a timbre that was different from, but similar to, the original timbre played the melodies at test, recognition was comparable to that when the same timbre played them. A similar timbre was effective to induce a close match between the overlapping timbre attributes of the memory trace and probe. Similarities between music and speech processing were implicated.

Experiments 3 and 4 assessed the influence of articulation format on melody recognition. In Experiment 3, melodies that remained in the same articulation format from study to test were recognized better than were melodies that were presented in a distinct format at test. Additionally, when the melodies were played in an articulation format that was different from, but similar to, the original format, performance was as reliable as that when they were played in the same format. A similar articulation format, akin to a similar timbre, used at test was effective to induce matching.

Experiment 4 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 instance-specific matching observed in melody recognition was defined: Whether instance-specific matching obtains depends absolutely on the quantitative amount of

match between the memory trace and the recognition probe, suggesting a global matching advantage effect. Implications for the nature of melody representation are discussed.

List of Tables

List of Figures

CHAPTER 1

General Introduction

Fodor (1983) describes perception as making the external environment accessible to central cognitive systems like belief, memory, and decision-making. In short, to perceive is to render the world accessible to thought. Perception begins when the world impinges on the sense organs (or transducers). However, while the transducers respond to stimulation by electromagnetic wavelengths and acoustic frequencies, our beliefs, memories, and decisions are about faces and objects. In Fodor's terms, while the transducers deliver representations of proximal stimulation patterns, central processes typically operate on representations of the distal objects. How does one get from the former to the latter – from proximal stimulations to mental representations of faces and objects? Clearly, higher level representations of the distal world must be constructed or inferred based on the transducer outputs. Fodor's view is that input systems interpret transducer outputs in a format that central processing can understand. Thus, what we have is a tripartite scheme of transducers, input systems, and central cognitive systems, which is roughly akin to the classic triptych of sensation, perception, and cognition.

How, then, would Fodor describe music perception? The lower-level psychoacoustic correlates to frequency and intensity are presumably inferred from the transducer outputs via the input systems, and eventually understood as pitch and loudness by central processing. In the same way, a sequence of pitch-time events (or musical notes) is recovered based on lower-level temporal information about the durations of events. But surely, when we hear a piece of music, we hear more than undifferentiated events. We hear, 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?

The general goal of this dissertation is to examine the nature of representational entities that are used in music perception and melody recognition. The series of experiments will examine how melodies are represented in memory and whether surface characteristics, along with abstract structures, are encoded into long-term memory (LTM). More specifically, these experiments will investigate whether information about timbre and articulation is represented in memory, and how this information is used during the retrieval and recovery of previously studied melodies.

In a recent review by McMullen and Saffran (2004), the authors suggest that there might be similar mechanisms of learning and memory that govern music and language processing. In the forthcoming sections of this chapter, I will first highlight these common mechanisms, which provide the initial motivation to investigate the specific issues raised in this dissertation. This will be followed by a critical review of extant work that has examined the nature of representational entities that are used in speech perception and spoken word recognition, and a consideration of the possible nature of representation in music perception and melody recognition. Finally, the specific goals of this project will be elaborated in greater detail.

SIMILAR MECHANISMS FOR MUSIC AND LANGUAGE

By sheer appearance, music and language are grossly different. No audience would ever confuse Mozart's sonata with a politician's speech, because we possess elaborate and distinct categories of knowledge about each of these two domains. Yet, scientists who are interested in the nature of music and language continue to be intrigued by possible connections between these two types of knowledge. For this dissertation, of particular interest is that from a developmental perspective, similar mechanisms already appear to subserve learning and memory for music and language from a young age.

Learning Mechanisms

Once the learner has been sufficiently exposed to musical and linguistic systems, he must in some way derive structure across the specific experiences represented in memory. Different learning mechanisms have been implicated in this process. Here, I focus on one particular mechanism: statistics.

Statistical learning, i.e., the detecting of sounds, words, or other units in the environment that cue underlying structure (see Saffran, 2003a), has become a topic of much interest. In the environment, statistical information, which is roughly correlated with different levels of structure, is plentiful. An example is that the probabilities with which syllables follow one another serve as cues to word boundaries. In other words, syllable sequences that recur consistently are more likely to be words than sequences that do not. To illustrate, the likelihood that "pre" is followed by "ty" exceeds the likelihood that "ty" is followed by "ba" in the sequence "pretty baby". Several studies (e.g., Aslin, Saffran, & Newport, 1992; Saffran, Aslin, & Newport, 1996) have shown that eight-month-old infants can capture these statistics when given just two minutes of exposure time, discovering word boundaries in speech based solely on the statistical properties of syllable sequences.

It seems that similar statistical learning abilities exist for sequences of musical tones. Several studies (e.g., Saffran, 2003b; Saffran & Griepentrog, 2001; Saffran, Johnson, Aslin, & Newport, 1999) have shown that infants can identify boundaries between "tone words" by tracking the probabilities with which some notes occur. Taken together, the results suggest that at least some aspects of music and language may be learned through the use of a common learning mechanism. Considering other facts about music and language, this assertion is probably not far-fetched. Pitch, for instance, plays a central role in many languages. In "tone languages" such as Mandarin, Thai, and Vietnamese, the same syllable spoken in a different pitch or pitch contour results in a completely different meaning and interpretation. The recent view is that people who speak tone languages are more likely to maintain highly specific pitch representations for words than those who speak nontone languages, such as English (see Deutsch, 2002).

Memory Mechanisms

In order for learning to take place, one must first be able to represent musical experiences in memory, so that the knowledge can be subsequently accumulated and manipulated. Jusczyk and Hohne (1997) investigated the LTM abilities of 7-month-old infants by exposing them to brief stories repeatedly. After that, the infants did not hear the stories for two weeks. They were later tested to see if the words were retained in LTM. The infants showed a preference in listening to the words taken from the stories compared to new, unstudied words. This finding suggests that the words have actually been retained in LTM.

Saffran, Loman, and Robertson (2000) conducted an analogous study using musical materials which suggests that similar abilities exist in infant's memory for music. Infants were exposed daily to CD recordings of Mozart's piano sonatas for two weeks. After that, they did not hear these musical selections for two weeks. They were later tested on passages from the familiar pieces compared with novel passages drawn from other piano sonatas by Mozart performed by the same pianist. These infants were compared with a control group of infants who did not hear any of the selections previously. The observation was that the infants from the experimental group preferred the familiar selections compared to the novel ones, while the infants from the control group showed no preference. Subsequent experiments revealed that the infants did not just remember random fragments of the music, but had in fact represented aspects of the overall structure of the piece, showing expectations regarding where particular passages should be placed (Saffran *et al.*, 2000). Taken together, these findings suggest that infants' memory for music may be as refined as their memory for language.

There have been other recent studies that investigated infants' LTM for music which demonstrate that infants' mental representations are very detailed. For instance, Ilari and Polka (2002) showed that infants can represent more complex pieces of music, such as Ravel's compositions, in LTM. Ten-month-old infants can represent acoustic patterns drawn from the specific performances which they were previously exposed to (Palmer, Jungers, & Jusczyk, 2001). Six-month-old infants can remember the specific tempo and timbre of music which they were exposed to, such that when the music was played at new tempos or with new timbres, recognition was hampered. These findings suggest that infants' representations for music are as specific as to include even tempo and timbre information. There have been similar observations for representations of linguistic materials. Houston and Jusczyk (2000) showed that 7.5-month-old infants displayed difficulty in recognizing words when the words are spoken in new voices. This suggests that talker-specific cues are not discarded in their representations of spoken words.

Mainstream research on speech perception and the effects of talker variability on learning and memory has in fact indicated that variation in speech signals is actually encoded and utilized during subsequent processing. We will now turn to review the results of these learning and memory paradigms in talker variability research because they are relevant to the nature of the representational entities used in speech perception and spoken word recognition. We will then proceed to consider the nature of the representational units utilized in music perception and melody recognition, on the basis that common learning and memory mechanisms appear to be at work in both language and music.

SPEECH PERCEPTION AND RESEARCH ON TALKER VARIABILITY

Traditionally, the perception of the linguistic content of speech – the words, phrases, and sentences – has been studied separately from the perception of voice (talker) identity (Pisoni, 1997). Variation in the acoustic realization of linguistic components due to differences in individual talkers has been considered a source of noise that obscures the underlying abstract symbolic linguistic message. The proposed solution to this "perceptual problem" is that there is a perceptual normalization process in which voice-specific acoustic-phonetic properties are evaluated in relation to prototypical mental representations of the meaningful linguistic constituents. Variation is presumably abstracted, so that canonical representations underlying further linguistic analysis can be obtained. Under this view of perceptual normalization, one assumes that the end product of perception consists of abstract, context-free linguistic units that are independent of the identification, recognition, and storage of nonlinguistic properties of speech, such as the talker's voice.

A contrasting approach to the traditional abstractionist approaches proposes that representations of spoken language include nonlinguistic or surface characteristics of speech (Goldinger, 1998; Pisoni, 1997). Under this view, nonlinguistic properties of speech are not separate from linguistic content, but rather constitute an integral component of the speech and language perception process. These voice attributes are retained in episodic memory along with lexical information, and are found to later facilitate recognition memory. The view is that talker information is not discarded through normalization in speech. Instead, variation in a talker's voice actually forms part of a rich and elaborate representation of the talker's speech. Under this view, the assumption is that the end product of speech perception consists of, along with abstract, context-free linguistic units, nonlinguistic (indexical) units such as the talker's voice, and both kinds of content contribute to the identification and recognition of speech.

Talker Variability and Learning

In learning paradigms, one is primarily concerned with whether participants can retain information about the perceptual properties of voices studied during a familiarization phase, and whether the acquired indexical information is utilized in the analysis and recovery of linguistic information during speech perception. If a systematic relationship exists between perceptual learning of indexical information and subsequent performance in speech perception, it would mean that the indexical properties of speech are retained during perception.

Nygaard and Pisoni (1998) and Nygaard, Sommers, and Pisoni (1994) reported a series of perceptual learning studies in which participants were trained to identify a set of 10 voices during the study phase. The participants were later given an intelligibility test in which they had to identify novel words spoken by either familiar talkers or unfamiliar talkers. The results revealed that familiarity with the talker improved the intelligibility of novel words produced by that talker. Nygaard and Pisoni (1998) extended these findings by showing a similar effect when participants were trained and tested on sentences. It appears that when one acquires indexical knowledge about a talker, perceptual sensitivity to linguistic information increases. This suggests that indexical and linguistic properties are integral in terms of the underlying processing mechanisms involved in speech perception. In other words, speech perception appears to be a talker-contingent process (see Goh, 2005). The view is that familiarity with voices may be stored in some form of procedural memory about specific aspects of the talker's voice that later helps in the processing of that particular talker (see Kolers, 1973; Pisoni, 1997).

Talker Variability and Memory

In memory paradigms, one is mainly concerned with whether the encoding of voice details would subsequently enhance or impede the recovery and discrimination of words or sentences presented during study. In most studies, voice information is manipulated and regarded as surface details of the token (see Pisoni, 1997). The task was to retrieve and respond to the linguistic content of the token while ignoring these surface details. Whether systematic effects of the voice manipulations on participants' performance are observed would determine whether memory for words and sentences is dependent on memory for voices.

Many studies (e.g., Goldinger, 1996; Pilotti, Sommers, & Roediger, 2000; Sheffert, 1998) have shown 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)[1], the general trend favours the position that voice information, along with indexical information, is encoded into LTM.

[1] A detailed discussion on the possibilities as to why null effects were observed in these reports is beyond the plan of this dissertation. See Goh (2005) for a review of these possibilities.

This view is compatible with exemplar-based models of LTM which assume that a new representation of a word or an item is stored in LTM every time it is encountered. These memory models, such as search of associative memory (Gillund & Shiffrin, 1984; Raaijmakers & Shiffrin, 1981), MINERVA 2 (Hintzman, 1988), and retrieving effectively from memory (Shiffrin & Steyvers, 1997), all incorporate the storage of detailed memory traces that include multiple aspects of the memory episode such as item, lexical, associative, and contextual information. In contrast to abstractionist assumptions made by traditional symbolic theorists, the position here is that information is not lost due to any normalization process. Instead, both general and contextual information are integrated in a holistic fashion, and these details are encoded and stored in memory. Under this view, memory is a dynamic and interactive process, where the processes underlying perception are not decoupled from the processes underlying memory.

Goldinger (1998) has applied this theory, using Hintzman's MINERVA 2 model (Hintzman, 1988), to an exemplar-based lexicon for speech perception and spoken-word recognition. By successfully modeling extant word-recognition data with a framework that affords that indexical information is preserved in memory, the implication is that variation and variability in speech are as important as the idealized canonical entities in spoken language processing.

MUSIC PERCEPTION AND RESEARCH ON SURFACE FEATURE VARIABILITY

As reviewed, the perception of the linguistic content of speech has traditionally been treated separately from the perception of talker identity, because talker variability has been regarded as noise that obscures the main underlying linguistic message. Yet, a contrasting approach proposes that representations of spoken language include nonlinguistic or surface characteristics of speech (Goldinger, 1998; Pisoni, 1997), where nonlinguistic aspects of speech, such as talker variability, are not separate from linguistic content, but rather constitute an integral component in memory for speech.

There is a similar dichotomy in the music domain. While there are linguistic and nonlinguistic content in speech, two kinds of information exist in music, namely *abstract structure* and *surface characteristics* (see Trainor, Wu, & Tsang, 2004). The *abstract structure* consists of the relative pitches and relative durations of the tones in the music, which refer to the pitch durations between tones regardless of their absolute pitch level, and the ratios between durations, regardless of their absolute length, respectively. A normalization process must occur to capture this structural information. During this extraction, information about performance features, such as absolute pitch, tempo, and timbre, is discarded. The *surface (or performance) characteristics*, on the other hand, consist of the non-structural aspects of the music, such as the exact pitch level, tempo, timbre, and prosodic rendering.

Both *abstract structure* and *surface characteristics* are useful for music interpretation. A representation of the abstract structure enables one to recognize a melody across different performances, and to recognize musical variations of a motif within a musical composition (Large, Palmer, & Pollack, 1995). For instance, *Happy Birthday* can be recognized even when it is presented at various pitches and tempos, or even when it is embellished and harmonized on various musical instruments. On the other hand, the surface characteristics allow one to identify the specific musician performing the work, and contribute to the emotional interpretation of that rendition. While Raffman (1993) has suggested that only the abstract structural information is encoded into LTM, others have reported that surface features are encoded into LTM as well (e.g., Halpern & Müllensiefen, 2008; Peretz, Gaudreau, & Bonnel, 1998; Radvansky, Fleming, & Simmons, 1995; Wolpert, 1990).

For instance, Peretz *et al.* (1998), in Experiment 3 of their study, investigated the effects of surface features on melody recognition, 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 piano while the remaining half were presented on 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 familiarization and test phases. In a sense, timbre attributes may be assumed, at this juncture, to be computed during the perceptual analysis of the musical input.

DISSERTATION OBJECTIVES

What are the representational units that are used in music perception and melody recognition? Are these units analogous to those that are utilized in speech perception and spoken word recognition? While voice information appears to play a substantive role in speech processing, to what extent are the surface features, such as timbre information, of melodies encoded, represented, and utilized in memory? Answering these questions constitutes the general goal of this dissertation. More specifically, this project seeks to investigate three key research issues – the role of (1) timbre-specific familiarity, (2) timbre similarity, and (3) articulation format – in music perception and melody recognition.

The Role of Timbre-Specific Familiarity

Extant studies that examined the effects of timbre information (e.g., Halpern & Müllensiefen, 2008; Peretz *et al.*, 1998; Radvansky *et al.*, 1995; Wolpert, 1990) have adopted the standard procedure to begin with a study list of melodies presented by different instruments, with each instrument presenting an equal number of melodies. After the study phase, the old melodies were randomly presented at the test phase, together with an equal number of new melodies. The task was to determine whether a melody presented at test was previously presented during the study phase, regardless of the instrument that originally played the melody. The critical manipulation was that at test, half of the old melodies were assigned to be played by the same instrument that originally played those melodies at study, whereas the remaining old melodies were played by a different instrument (i.e., an instrument that was used at study but which did not originally play that particular melody). The new melodies were distributed equally to be presented by the instruments used in the study set. Differences in recognition performance between the same-instrument and different-instrument conditions constitute a timbre repetition effect. The interpretation is that timbre information, together with structural information, has been encoded into LTM.

An alternative, and perhaps less popular, way of assessing timbre effects would be to compare performance between same-timbre repetitions and new-timbre, instead of different-timbre, presentations (see Trainor *et al.*, 2004). Rather than assigning half of the old melodies to a previously studied but different instrument at test, these melodies were presented with *completely new* instruments that never appeared before at study. Here again, differences in recognition performance between the same-timbre and new-timbre conditions constitute a timbre repetition effect, whereby timbre information has presumably been encoded into LTM.

Theoretically, at least two processes can explain why same-timbre repetitions offer an advantage over new-timbre presentations during melody recognition. First, the match between the episodic memory trace and the probe can determine whether a repetition advantage would obtain. The more precise the match is, the more sizeable the repetition effect would be. This assertion is based on the now-classic 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. Timbre information is first encoded and stored in the memory traces of the melodies, and 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 compared to the other competing traces.

On the other hand, a melody presented by a new 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 compared to the melody that is repeated by the same timbre.

Second, a timbre repetition effect can also be attributed to a greater familiarity with the timbre properties of the studied instruments, rather than to the extent of match between memory traces and probe *per se*. Global memory frameworks, such as search of associative memory (e.g., Gillund & Shiffrin, 1984) and MINERVA 2 (Hintzman, 1988) propose that all memory traces are probed concurrently, and that the relative activation strengths of each memory trace depend on the degree of matching attributes with the probe. A previously studied (i.e., familiar) timbre may induce heightened activation levels of the memory traces of all melodies that contain attributes of the studied timbre; an unstudied (i.e., unfamiliar) timbre will not constitute an effective cue because no memory trace will contain attributes of the unstudied timbre. Thus, melodies played by the unstudied timbre ought to be less discriminable than those that are repeated in the same timbre from study to test.

Both instance-specific matching and timbre-specific familiarity can account for the advantage from using same-timbre repetitions. In the standard procedure of assessing timbre effects (see Halpern & Müllensiefen, 2008; Peretz *et al.*, 1998; Radvansky *et al.*, 1995; Wolpert, 1990), the timbres used at test would have previously appeared at study, and were therefore likely to be equally familiar to participants. Thus, it is logical that any timbre repetition effect obtained would be attributable to instance-specific matching processes *per se*, rather than timbre-familiarity processes. On the other hand, in the other paradigm that compared performance between the same-timbre repetitions and new-timbre presentations (see Trainor *et al.*, 2004), any timbre repetition effect would be attributed to either instance-specific matching or a global timbre-specific familiarity with a previously studied timbre, or to both of these processes.

However, it is apparent that both of these designs are inadequate to elucidate the role of timbre-specific familiarity processes *per se* in melody recognition. This project will systematically assess the unique contributions of both types of processes to the timbre repetition effect.

The Role of Timbre Similarity

Extant studies that examined timbre effects (e.g., Halpern & Müllensiefen, 2008; Peretz *et al.*, 1998; Radvansky *et al.*, 1995; Wolpert, 1990) have used test stimuli that were denoted as either of the same or different format, paying little attention to effects arising from varying magnitudes of intermediate perceptual differences. Such effects of fine-grained perceptual details of timbre have not been systematically examined, so one could not determine whether these details contributed to the disparate timbre effects observed in the extant literature.

Consider Experiment 3 of Peretz *et al.* (1998) for example. In their different-timbre condition, the timbres used to present the melodies at test (e.g., flute) appear to be completely distinct from those used during the study phase (e.g., piano). It can be argued that the two timbres are perceptually distinct from each other because the flute and the piano primarily belong to the woodwind and keyboard (i.e., different) orchestral family groups, respectively. It was reported that melody discrimination performance was impeded in this condition. The question I asked was: Would the same effect on melody recognition emerge if timbres that are different from, *but similar to*, those at study were now used to present the melodies at test? (Here, a candidate for testing could be the electric piano, if it can be established that this instrument is perceptually similar to the piano.) In this project, I will, in response to this question, assess the contribution of timbre similarity details to these timbre effects.

The Role of Articulation Format

According to Trainor *et al.* (2004), the surface or performance characteristics in music comprise of the non-structural aspects of the music, such as the exact pitch level, tempo, timbre, and prosodic rendering. The effects of these performance characteristics on melody recognition have been previously studied (see Trainor *et al.*, 2004). But to date, no one has examined the effects of a type of surface characteristics known as *articulation*. Articulation is commonly defined and understood by trained musicians as whether the music (e.g., melody) is played in a *legato* (i.e., continuous) or *staccato* (i.e., detached) format.

The significance of examining the effects of articulation on melody recognition is two-fold. First, this investigation is new. Second, it allows ease of manipulation control. It can be difficult to directly quantify the degree of similarity or match between two different voices during spoken word recognition, or between two different timbres during melody recognition. For instance, it has been reported that voice perception depends on a combination of multiple physical dimensions, such as gender and vocal pitch (see Goldinger, 1996). In a similar vein, musical timbre does not depend upon a single dimension. Attributes such as amplitude, phase patterns of components, the alteration of the initial part of a sound, as well as the brightness of the steady-state portion of the sound have been found to influence timbre perception (see Samson, Zatorre, & Ramsay, 1997). In contrast, the exact amount of match (or mismatch) between two instances of a melody varying in articulation format can be directly quantified and, therefore, systematically manipulated. This project will investigate the effects of varying articulation format on melody recognition.

SUMMARY OF PROJECT GOALS AND OVERVIEW OF EXPERIMENTS

Summarizing, this project has three specific goals. First, Experiment 1 of this project will systematically assess the unique contributions of both instance-specific matching and timbre-specific familiarity processes to the traditional timbre effects observed in previous research. Second, Experiment 2 will discover the contribution of timbre similarity to these timbre effects. Third, Experiments 3 and 4 will pioneer a new investigation of the effects of varying articulation format on melody recognition. An extensive discussion of the key findings from these experiments and their implications will be presented in the final chapter of this dissertation.

CHAPTER 2

Timbre Similarity Scaling and Melody Testing

This chapter describes two preliminary studies that were conducted. In the first study, the degree of perceived similarity among different timbres was established. The second study tested for an appropriate number of melodies to be used in the subsequent main experiments of the present project.

PRELIMINARY STUDY 1:

Timbre Similarity Scaling

Experiments 1 and 2 of the present project have been designed to investigate the nature of the traditional timbre effects observed in melody recognition. Prior to conducting these main experiments, it was first essential to construct a multidimensional "timbre map" that shows the similarity relations between the individual timbres that will be used as the stimulus materials. This was so that the selection of specific timbres for use in the subsequent main experiments can be based on objective measures of the degree of perceived similarity among different instruments that is independent of semantic categories of instruments, even though a trained musician might already assume that instruments within each of the orchestral family groups (e.g., strings, woodwind, brass, keyboard, etc.) would be similar sounding. This section describes the steps taken to collect similarity ratings and the generation of the "timbre map" using multidimensional scaling (MDS) techniques (Kruskal & Wish, 1978).

Method

Participants

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

Materials

The stimulus set comprised of monophonic C-major arpeggios[2] for the familiarization phase, and C-major diatonic scales[3] for the similarity rating phase, all of which were played by 12 different instruments. These 12 instruments have been selected on the basis that they are representative of four major orchestral family groups, as listed in Table 1. Each arpeggio and scale lasted approximately five seconds and eight seconds, respectively. These tunes were constructed using the *Finale 2009* software, and were recorded in .wav sound files.

---

[2] In western music context, an *arpeggio* can be understood in terms of a tonic triad that comprises the tonic, mediant, and dominant notes of a key. The *tonic* refers to the underlying key in which a melody is written (e.g., C for a melody written in the key of C major). Together with the *mediant* (E) and *dominant* (G), these three intervals are sounded simultaneously to form the melody's major chord called the tonic triad. An arpeggio is essentially a tonic triad with the three intervals played one at a time sequentially (i.e., C is sounded first followed by E, which is then followed by G). A basic form of arpeggio typically starts and ends on the tonic of the key.

[3] A diatonic scale in western music is made up of a succession of sounds ascending (or descending) from a starting note, usually the tonic. For instance, a C major scale, in its ascending form, comprises the following pitches to be played one at a time in sequence: C D E F G A B and C again.

Table 1

Twelve Instruments Classified by Orchestral Family Grouping

Orchestral family grouping			
Woodwind	Brass	Strings	Keyboard
Flute	Trumpet	Violin	Piano
Clarinet	French Horn	Viola	Harpsichord
Oboe	Trombone	Cello	Electric Piano[4]

Apparatus

Computers equipped with 16-bit sound cards were used to control the experiment. The signals were presented to participants via a pair of Beyerdynamic DT150 headphones at approximately 70 dB sound pressure level. E-prime 1.2 was used for stimuli presentation. The computer keyboard was used to collect the similarity ratings. Keys 1, 3, 5, and 7 were labeled *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 consisted of two parts. The first part was a short, three-minute familiarization phase to familiarize the participants with the 12 different timbres that they would be rating. During this phase, participants listened to a random order of 12 instruments

[4] Although the electric piano is not a standard member of orchestral instruments, it would be apt to classify this instrument under the *Keyboard* family on the basis of its functional similarity to the traditional piano.

playing the same arpeggio pattern. No ratings were collected during this phase; participants were told to simply listen to the instruments. On each trial, a single arpeggio was played by a particular instrument over the headphones, after which, participants pressed the space key to proceed to the next arpeggio. This sequence continued until all 12 timbres were presented. The timbre 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 15 minutes to complete. At the start of each trial, the question *How similar are the two instruments?* was displayed on the monitor. Two different instruments playing the same scale were then presented, 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 instrument 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 every 22 trials, and rated a total of 66 pairwise comparisons. Participants were debriefed at the end of the session.

Results and Discussion

MDS using the ALSCAL routine of SPSS version 16 was used to analyze these perceptual similarity data. Figure 1 shows the timbre map from the ALSCAL solution derived by collapsing across all participants. 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). Since twelve timbres were scaled, solutions with one through three dimensions were obtained, and the amount of variance accounted for and Kruskal's stress values were examined for each solution.

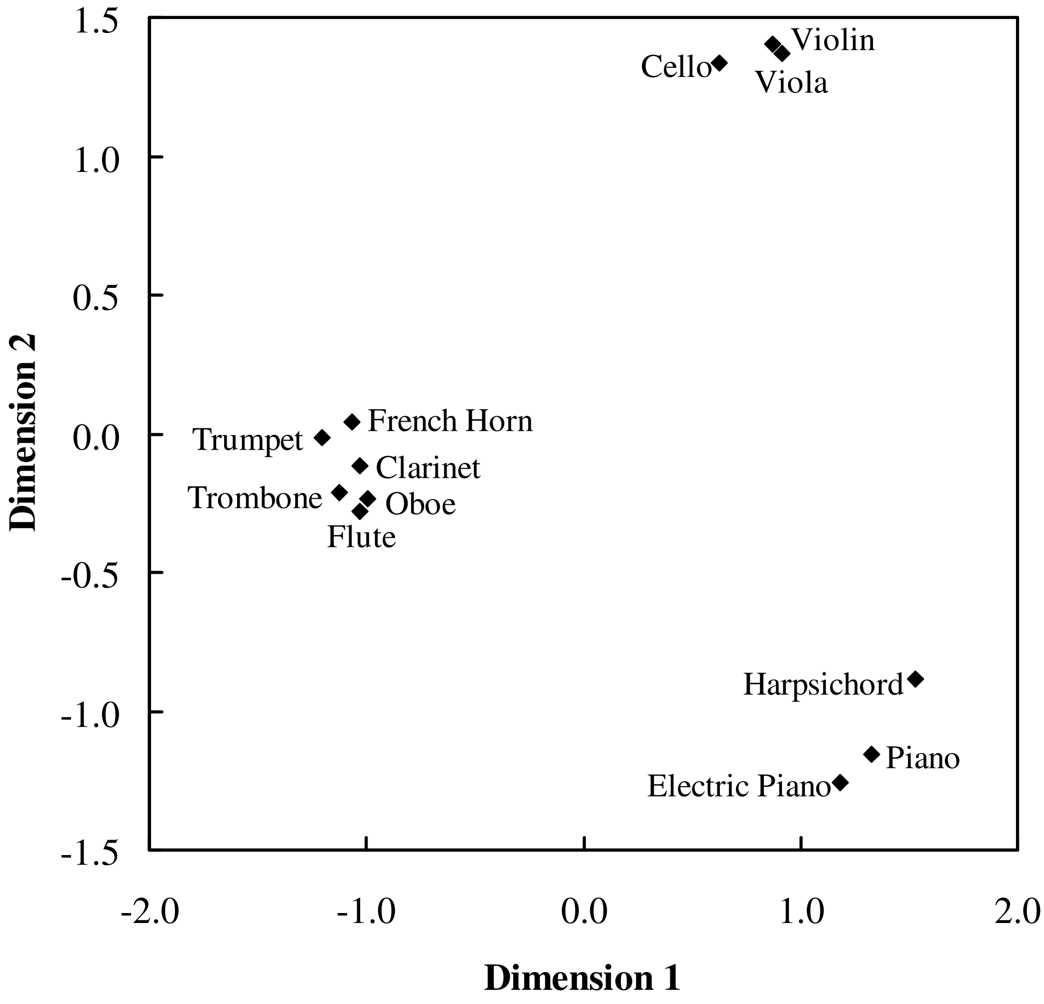


Figure 1. Two-dimensional MDS solution for 12 instruments.

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 values obtained are shown on Table 2. Inspection of the present values indicated that while there was a large increase in goodness-of-fit between the one- (Kruskal's stress = .295, R^2 = .72) and two-dimensional (stress = .095, R^2 = .97) solutions, the improvement for the three-dimensional solution (stress = .065, R^2 = .97) was not sufficient to justify this solution, implicating a two-dimensional solution as optimal.

Table 2

Kruskal's Stress and R^2 Values Obtained for Solutions with One through Three Dimensions

Solution	Kruskal's stress	R^2
1-dimensional	.295	.72
2-dimensional	.095	.97
3-dimensional	.065	.97

Dimension 1 was difficult to interpret but might be influenced by the presence or absence of attack (accent) in the initial part of the sound. For instance, the initial part of the sound produced by the violin or the piano tends to carry a more pronounced and "sharp" accent as compared with that produced by the flute or the horn. This interpretation is compatible with previous reports which suggest that the variation of the initial part of a sound can affect the perception of musical timbre (e.g., Berger, 1964; Clark, Robertson, & Luce, 1964; Grey & Moorer, 1977; Saldanha & Corso, 1964; Wedin & Goude, 1972). The second dimension appears to group the instruments by family, with the woodwind and brass families clustered together as two highly similar groups. Notwithstanding the interpretations offered, it should be noted that determining the nature of the two dimensions is peripheral to the experiments described in this project. The objective of deriving the MDS solution of timbre similarity was to provide a principled basis for determining suitable instruments that would eventually be used as stimuli in the experiments.

PRELIMINARY STUDY 2:
Melody Testing

Recent psychological research on music has been driven by cognitive psychology, which underscores the influence of knowledge on perception. This approach affords that a presented stimulus is interpreted by knowledge, sometimes called schemas, that is acquired through previous experience. In music, the schemas include typical rhythmic and pitch patterns. Rhythm and pitch are two primary dimensions of music, and are interesting psychologically because simple, well-defined units can combine to form highly complex and varied patterns. The elements of rhythm and pitch in music have been commonly defined in terms of specific musical aspects called *meter* and *tonality*, respectively (see Krumhansl, 2000).

Meter defines the underlying beat or pulse of a melody, based on the number of beats that are assigned to each bar of the melody. For instance, a melody written in a duple (e.g., 2/4) time meter consists of two equal beats in a bar, while a melody written in a triple (e.g., 3/4) time meter comprises three equal beats in a bar. Tonality refers to the underlying scale in which a melody is written, which in turn constrains the specific notes that will appear in a melody (see Boltz, 1991). For instance, a melody written in the key of C major would have its tonal intervals (notes) derived from the C major diatonic scale: C D E F G A and B.

For the stimulus database employed by the extant studies that examined timbre effects (e.g., Peretz *et al.*, 1998), it was observed that the melodies have not been systematically controlled for meter and tonality. In the present investigation, my intention was to create a new stimulus database comprising melodies that would control for these two technical aspects. Because these melodies were new and task difficulty was presumably a function of the number of melodies presented for study, a second preliminary study was essential to discover an appropriate number of these melodies to be used in the subsequent main experiments. By employing a suitable number of melodies, floor effects that could potentially obscure the traditional timbre effects found in melody recognition should not emerge. In other words, the melody discrimination task should not be too difficult (due to an overwhelming number of melodies to be studied) to the extent that the melodies become indiscriminable at test. This section describes the steps to establish this appropriate number.

Method

Participants

Twenty-four introductory psychology students from the National University of Singapore participated for course credit. None had participated in the first preliminary study.

Materials

The stimulus set comprised of 48 single-line (monophonic), newly composed melodies. Their meter and tonality properties are shown in Table 3. An equal number of four-bar melodies was composed in the tonality (key) of C major, C minor, G major, or G minor. The melodies began either on the tonic (i.e., the first interval or note of the home key), mediant (i.e., the third note of the key), or dominant (i.e., the fifth note of the key), but always ended with a single long note on the tonic of the key they were written in. For instance, a melody written in G minor opened either on G (tonic), B-flat (mediant), or D (dominant), and always closed on G. Each melody was written in simple triple (i.e., three equal beats in a bar) or simple quadruple (i.e., four equal beats in a bar) time, lasting approximately six seconds or 7.2 seconds respectively. The melodies were constructed using the *Finale 2009* software, and were recorded in .wav sound files. Musical notations of sample melodies are listed in Appendix A.

Table 3

Meter and Tonality Properties of the Present 48 Melodies

	Tonality			
Meter	C major	C minor	G major	G minor
Simple triple (3/4)	6	6	6	6
Simple quadruple (4/4)	6	6	6	6

Note. Numbers denote the quantity of melodies in each classification.

Based on the timbre scaling solution that was derived in the first preliminary study (see Figure 1), object coordinates in the space were used to estimate perceptual distances between all instruments. Estimates 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. The planar coordinates of the instruments as well as the Euclidean distance between each instrument pair are shown in Appendix B. Based on these Euclidean estimates, two timbre sets containing three instruments each were selected to present the melodies in this second preliminary study, as well as in the subsequent main experiments. Set A comprised of piano, flute, and violin, whereas Set B comprised of harpsichord, clarinet, and cello. Within each set, the three instruments were perceived as maximally (and as equally) distinct from each other.

Apparatus

The equipment was the same as that used in the first preliminary study, except that the PST Serial Response Box (Schneider, Eschman, & Zuccolotto, 2002), with the left- and right-most buttons of the button-box labeled *No* and *Yes* respectively, was used for data collection.

Design and Procedure

Forty-eight melodies were used. One list of 24 melodies was selected to be the old melodies, and a second list of the remaining 24 melodies was selected to be the new melodies. These were two equivalent melody groups, matched in terms of meter and tonality. List selection here was counterbalanced across participants.

Each participant was randomly allocated to listen to the melodies played by one of the six instruments. Participants were tested individually or in small groups of seven or fewer. The session consisted of two parts. The first part was the study phase and took approximately five minutes to complete. Participants were instructed to silently memorize each melody that was presented over the headphones. At the start of each trial, a ready prompt was displayed on the monitor for one second, after which it was erased. One second later, a single melody was played over the headphones; 800 ms following the first presentation, the melody was repeated. After two instances of each melody were presented, participants pressed the space key to proceed to the next melody. This sequence continued until all 24 melodies were presented. The melody presentation sequence was random across participants. Participants were informed of a forthcoming recognition test.

Following the study phase, a recognition test was given. 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. The primary purpose of this pilot study was to establish baseline discrimination performance based on an appropriate number of melodies presented for study, thus the melody remained in the same timbre from study to test. Participants were told to press the *Yes* button on the Serial Response Box if they thought they had heard the melody earlier. 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 minutes to complete all 48 randomly presented trials. Participants were debriefed at the end of the session.

Results and Discussion

d-prime (d'), a measure of discriminability commonly used in signal detection, was used to measure melody discrimination performance. Signal detection theory was initially developed to examine performance in perception experiments in which the participant's task was to detect the presence of a signal, for instance, a tone. When applied to the present case, the theory assumes that when a test melody is shown in the recognition test, there is a certain amount of evidence that the melody appeared in the study phase. The theory also assumes that this evidence is normally distributed and can be construed as frequency distributions, as illustrated in Figure 2.

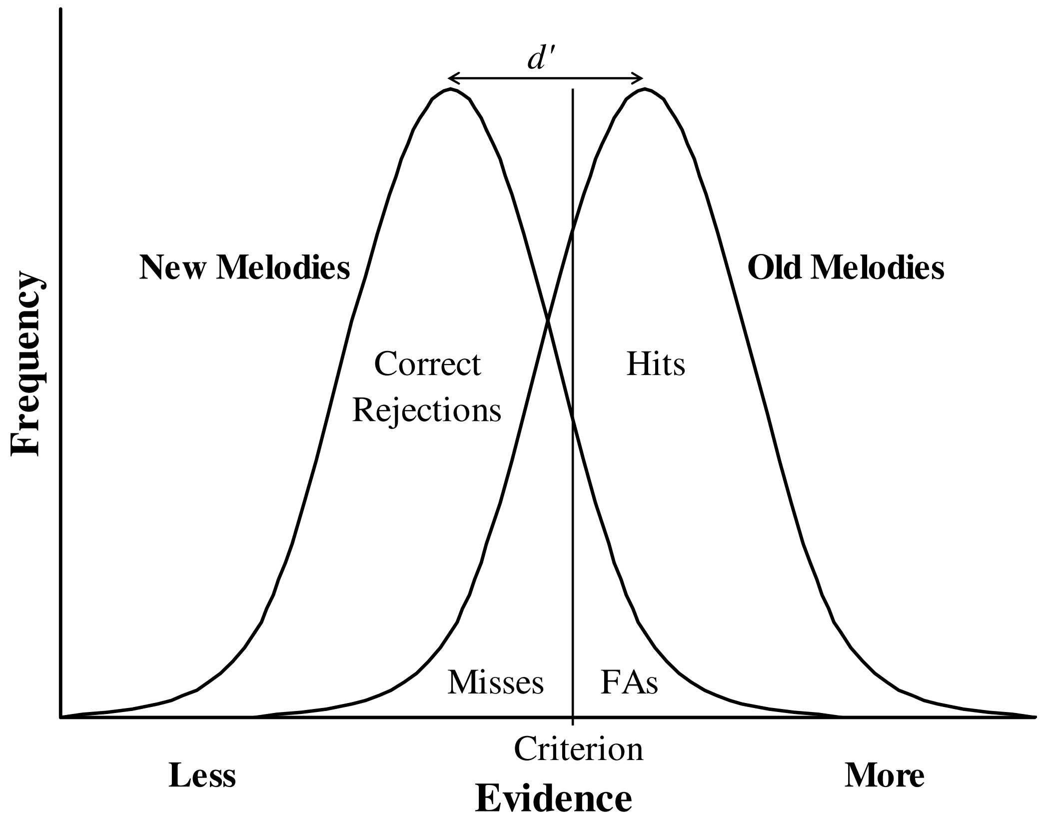


Figure 2. Graphical representation of criterion and d' in signal detection theory.

The participant first evaluates the evidence supporting the idea that the melody is old or new. On the average, old melodies are assumed to appear more recognizable than new melodies, even though there will often be some overlap. The participant adopts a certain criterion, and melodies that appear more recognizable will be judged as “old” and therefore elicit a *yes* response, while melodies that appear less recognizable will be judged as “new” and elicit a *no* response. Melodies that fall to the right of the criterion but are from the new melody distribution constitute false alarms (FAs). Items that fall to the left of the criterion but are old from the old item distribution constitute misses (see Figure 2).

d' is the difference between the means of the two distributions divided by the standard deviation of the "new" distribution, or simply $z_{FA} - z_H$, where z_{FA} indicates the false alarm rate and z_H indicates the hit rate. A d' of 0 then represents the case where the distributions overlap and the participant cannot tell the difference. The greater the difference between the distributions, the larger d' is, and the more different the old and new melodies are. Values of d' between 1 and 2 usually represent good yes-no recognition performance (Neath & Surprenant, 2003, p. 202).

The average d' obtained, based on the hit rate (*M* = 68.00, *SD* = 13.60) and false-alarm rate (*M* = 23.50, *SD* = 14.34), was 1.31 (*SD* = 0.56), implicating good discrimination performance. *C*, a response bias measure, was 0.16 (*SD* = 0.37). Note that for *C*, a positive value indicates a conservative bias; the present value obtained indicates conservatism of participants' responses. The data outcome suggests that when a total of 48 melodies were used in the melody recognition task, participants could discriminate the melodies reasonably well at test; no floor effects were apparent.

CHAPTER 3

Are Music and Speech Similar? (Re-)Examining Timbre Effects in Melody Recognition

This chapter describes two experiments that were conducted to (re)examine the traditional timbre effects observed in melody recognition. Specifically, these experiments aimed to elucidate the processes that govern these timbre effects. Experiment 1 sought to reveal whether instance-specific matching or timbre-specific familiarity, or both of the processes, govern the timbre effects, while Experiment 2 aimed to show how timbre similarity contributes to these effects. Based on the data from both experiments, several similarities appear to exist between music and speech processing.

EXPERIMENT 1:

Instance-Specific Matching versus Timbre-Specific Familiarity

My primary goal in Experiment 1 was to examine the contributions of instance-specific matching and timbre-specific familiarity processes to the timbre effects observed in music recognition memory. As discussed earlier in Chapter 1, both instance-specific matching and timbre-specific familiarity can account for the advantage from using same-timbre repetitions. To recapitulate, in the standard procedure of assessing timbre effects (see Halpern & Müllensiefen, 2008; Peretz *et al.*, 1998; Radvansky *et al.*, 1995; Wolpert, 1990), the timbres used at test would have previously appeared at study, and were therefore likely to be equally familiar to participants. Thus, any timbre repetition effect obtained would logically be attributable to instance-specific matching processes *per se*, rather than to timbre-familiarity processes. On the other hand, in the alternative paradigm that compared performance between the same-timbre repetitions and new-timbre presentations (see Trainor *et al.*, 2004), any timbre repetition effect would be attributed to either instance-specific matching or a global timbre-specific familiarity with a previously studied timbre, or to both of these processes.

However, the limitation to both paradigms is that they are unable to elucidate the role of timbre-specific familiarity processes *per se* in melody recognition. The present study is designed to systematically assess the unique contributions of both types of processes to the timbre repetition effect, by examining discrimination performance in same-, different-, and new-timbre conditions simultaneously. The highlight is that the present new design allowed a novel comparison between the different- and new-timbre conditions, such that any observed differences can now be solely attributable to a timbre-specific familiarity effect due to the fact that there were no instance-specific matches in either condition.

While no study to date has investigated the timbre-repetition effect with same-, different-, and new-timbre conditions in a single experiment, Goh (2005) reported an analogous study using spoken words. The study examined voice context effects in recognition memory for words spoken by multiple talkers, by comparing performance when studied words were repeated with same, different, or new voices at test. The data revealed that discrimination performance improved only when the exact same voice was repeated at test, suggesting that the voice-specific attributes of individual talkers are preserved in LTM. The data also suggested that participants were the least conservative in responding towards words tested in the studied-same voice, followed by words in different-studied voices, and, last, words in completely new voices. To the extent that the surface (i.e., timbre) attributes of melodies are analogous to the voice attributes of spoken words (see Goh, 2005), the prediction is that discrimination performance will improve only when studied melodies are repeated in the exact same timbre during the recognition phase. In addition, participants will be the least conservative in responding towards melodies tested in the same timbre.

Method

Participants

Fifty-two introductory psychology students with varying music training experience from the National University of Singapore participated for course credit. None had participated in the preliminary studies.

Materials and Apparatus

The stimulus set comprised of the 48 single-line (monophonic) melodies that were used in the second preliminary study. The equipment used was the same as that in the second preliminary study.

Design

To ease exposition, the terms *old* and *new* are used to refer to the melodies, whereas the terms *studied* and *unstudied* are used to refer to the instruments (instead of “old” and “new” instruments). The timbre-context conditions were run within participants. For the old melodies, three levels of timbre context – studied-same, studied-different, and unstudied – were manipulated. For the new melodies, only two levels – studied and unstudied – were used. It is logically impossible to have studied-same or studied-different timbre conditions for new melodies during the test phase, because a new melody would not have been presented during the study phase.

The unequal number of timbre-context conditions for the old and new melodies made it impossible to ensure that there was an equal exposure of trials in each condition while also ensuring that there was equal exposure to each of the studied and unstudied instruments in the test phase. If there were an equal number of old-melody trials among the three timbre-context conditions, participants would be exposed to twice as many studied timbres as unstudied timbres (because two of the old-melody conditions, studied-same and studied-different, used studied timbres, whereas only one old-melody condition used unstudied timbres). Any bias effects observed using studied timbres might therefore be simply attributable to the amount of exposure of those timbres rather than timbre context *per se*. Hence, it was more critical to ensure that the number of exposures to each instrument in the test phase was equal. The following controls were used to achieve this.

One list of 24 melodies was selected to be the old melodies, and a second list of 24 melodies was selected to be the new melodies. These were two equivalent melody groups, matched in terms of meter and tonality. List selection here was counterbalanced across participants. Random selection was used in the assignment of melodies to instruments for all the procedures described below. In the study phase, the 24 old melodies were divided among two studied instruments (i.e., either piano and flute of Set A, or cello and harpsichord of Set B), so that each instrument presented 12 melodies. Therefore, both instruments were heard an equal number of times during the study phase.

In the test phase, the 24 new melodies were divided among three instruments so that each instrument presented eight melodies. For the 24 old melodies, one-third of them were switched to the unstudied instrument (i.e., either violin or clarinet). This manipulation constituted the unstudied timbre-context condition for old melodies. For the other 16 old melodies, eight were selected for the studied-same-timbre-context condition, while the remaining eight were assigned to the studied-different-timbre-context condition.

These control procedures ensured that the number of exposures to each instrument was equal during the test phase. For new melodies, all three instruments presented eight melodies each. For old melodies, the unstudied instrument presented eight melodies, whereas the two studied instruments also presented eight melodies each (see Table 4 for a summary). The fact that all three instruments presented an equal number of old and new melodies meant that participants would not be biased towards making old- or new-melody responses as a result of differing amounts of exposure to studied and unstudied timbres.

Table 4

Summary of the Design Used in Experiment 1

Memorization stage		Recognition stage					
Study melodies		Test melodies (Old)			Test melodies (New)		
		Timbre context					
		Studied-same	Studied-different	Unstudied	Studied		Unstudied
		Set A instruments					
Piano	Flute	Piano	Flute	Violin	Piano	Flute	Violin
12	12	8	8	8	8	8	8
		Set B instruments					
Cello	Harpsichord	Cello	Harpsichord	Clarinet	Cello	Harpsichord	Clarinet
12	12	8	8	8	8	8	8

Note. Numbers indicate the quantity of melodies played by the respective timbres in each classification.

Procedure

Half of the participants were randomly allocated to listen to melodies played by the Set A instruments, while the other half were randomly allocated to listen to melodies played by the Set B instruments. Participants were tested individually or in small groups of seven or fewer. The session consisted of two parts. Figure 3 shows the schematic of the sequence of a trial. The first part was the memorization phase and took approximately five minutes to complete. Participants were instructed to silently memorize each melody that was presented over the headphones. At the start of each trial, a ready prompt was displayed on the monitor for one second, after which it was erased. One second later, a single melody was played over the headphones; 800 ms following the first presentation, the melody was repeated. After two instances of each melody were presented, participants pressed the space key to proceed to the next melody. This sequence continued until all 24 melodies were presented. The melody presentation sequence was random across participants. Participants were informed of a forthcoming recognition test.

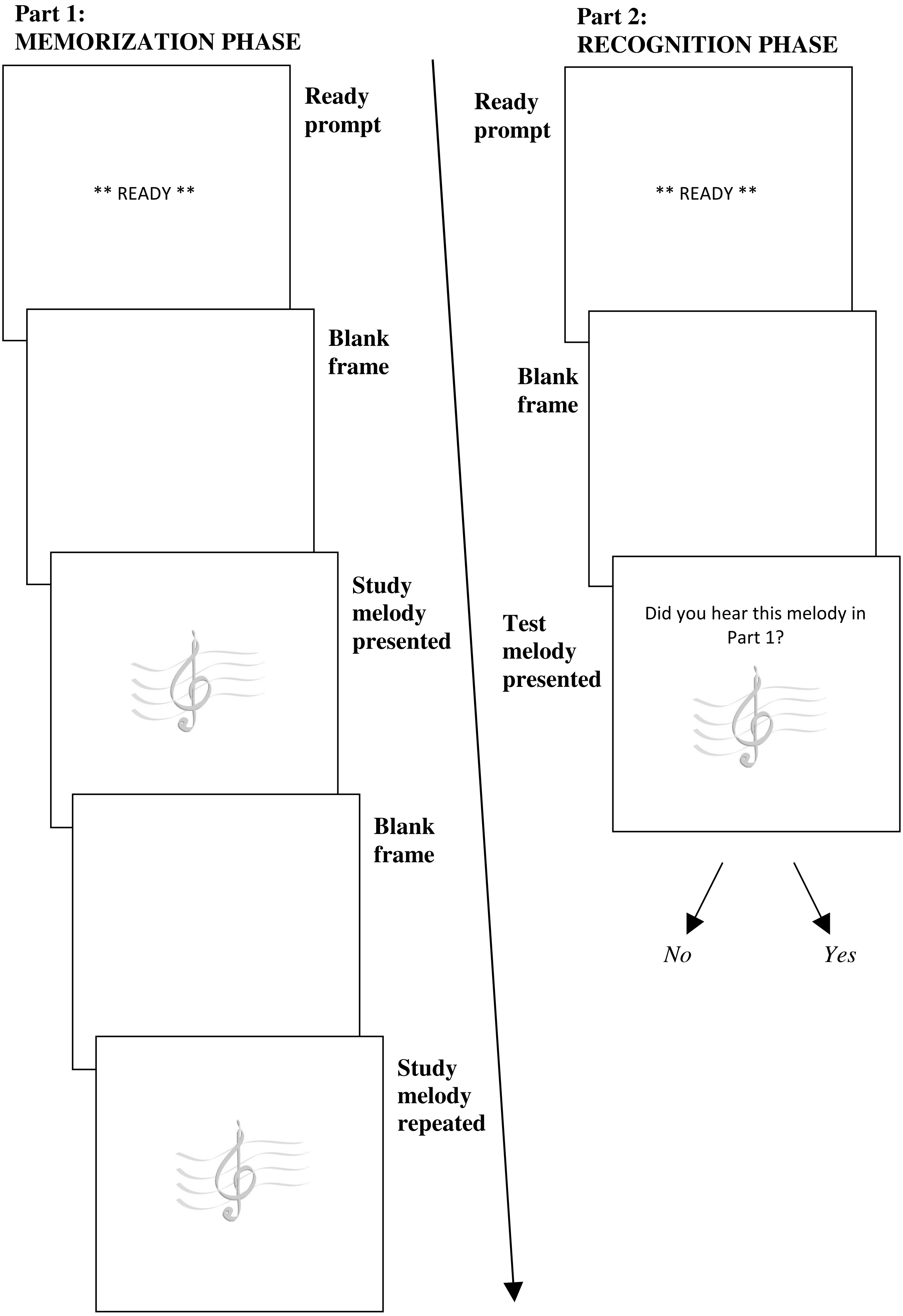


Figure 3. Schematic of the sequence of a trial in Experiment 1.

Following the study phase, a recognition test was given. 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 instrument that played the melody. 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 minutes to complete all 48 randomly presented trials.

After the recognition test, information on the participants' music training experience was recorded. 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. Participants were debriefed at the end of the session.

Results and Discussion

Hits, false alarms, *d'*, and *C* were used to measure melody recognition performance. *d'* measured the accuracy of performance while *C* measured response bias. Specifically, participants discriminated between old and new melodies, regardless of the instrument that played them at test. A melody that was correctly identified at test as an old melody, thus leading to a correct *yes* response constitutes a "hit"; a new melody that was mistaken for an old melody, thus leading to a wrong *yes* response constitutes a "false alarm". There were three levels of timbre context for the old melodies – studied-same, studied-different, and unstudied, while there were only two levels of timbre context for the new melodies – studied and unstudied. As such, three hit rates were obtained, but there were only two false alarm rates. Consequently, *d'* and *C* for the studied-same and studied-different conditions were computed based on their respective hit rates but using a common false alarm rate for new melodies presented by a studied timbre; *d'* and *C* for the unstudied condition were calculated based on their respective hit and false alarm rates.

Musical training influences. The effects of musical training were not a primary goal of this project and there was no systematic attempt to control the number of participants with or without formal musical training. Nevertheless, the main analyses for all performance measures were always preceded by an examination of potential effects of musical training for the sample. This was done using a two-way mixed design Analysis of Variance (ANOVA), with musical training as the between-subjects factor (participants who had at least four years of formal music training, $n = 14$, versus those without or with less than four years of music training, $n = 38$) and timbre context as the within-subjects factor. The outcome of these analyses is reported before proceeding to the main findings involving the three timbre-context conditions.

No reliable interaction between musical training and timbre context was observed for hits, false alarms, *d'*, and *C*, all Fs < 1.60. This shows that musical training did not influence the timbre effects in all measures. A main effect of musical training was marginally significant for *d'* and false alarms, Fs$(1, 50) = 3.94$; 3.57, MSes $= 0.66$; 0.03, ps $= .053$; $.065$, respectively. Participants with at least four years of formal music training tended to have higher discrimination scores (*M* = 1.27, *SD* = 0.77) and fewer false alarms (*M* = 25.75, *SD* = 15.60) than those without or with less than four years of music training (*d'* *M* = 0.98, *SD* = 0.60; false alarms *M* = 32.93, *SD* = 15.36). There was no reliable main effect of music training for hits and *C*, all *F*s < 1.39. The present data were also reanalysed using musical training as a covariate, and the pattern of results did not change. Since musical training did not interact with timbre context, the main findings for the timbre-context conditions reported below can be generalized across all participants within the sample, and all subsequent tabulations of results are collapsed across musical training.

Timbre context effects. Table 5 presents the pattern of results for hit performance across the three timbre-context conditions. There was a main effect of timbre context, $F(2, 102) = 6.12$, $MSe = 0.02$, $p < .01$. Pairwise comparisons revealed that there were reliably more hits when melodies were repeated with the studied-same timbre than when melodies were presented with a studied-different timbre, $t(51) = 2.62$, $p < .05$, or an unstudied timbre, $t(51) = 3.73$, $p < .01$. The number of hits did not differ between the studied-different and unstudied timbre-context conditions, $t < 0.62$. This pattern of results indicates that the number of hits increased significantly only when melodies were repeated in their original timbre.

Table 5

Percentage of Hits Across Timbre-Context Conditions in Experiment 1

	Timbre context		
	Studied-same	Studied-different	Unstudied
M	72.86	65.38	63.46
SD	14.83	19.20	15.92

Table 6 shows the false alarm pattern. There was no evidence of a main effect of timbre context, $F < 1$, indicating that timbre context did not influence the false alarm rates.

Table 6

Percentage of False Alarms Across Timbre-Context Conditions in Experiment 1

	Timbre context	
	Studied	Unstudied
M	31.22	30.77
SD	13.74	17.48

Table 7 presents the pattern of results for *d′* performance across the three timbre-context conditions. A reliable main effect of timbre context was obtained, $F(2, 102) = 3.21$, $MSe = 0.31$, $p < .05$. Pairwise comparisons revealed that participants were significantly better at discriminating test melodies presented with the studied-same timbre than they were at discriminating melodies presented with a studied-different timbre, $t(51) = 2.33$, $p < .05$, or an unstudied timbre, $t(51) = 2.29$, $p < .05$. Discriminability did not differ between the studied-different- and unstudied-timbre-context conditions, $t < 0.58$. This pattern of results indicates that discriminability increased only when test melodies appeared in the studied-same-timbre context.

Table 7

Discrimination Performance (d') Across Timbre-Context Conditions in Experiment 1

	Timbre context		
	Studied-same	Studied-different	Unstudied
M	1.22	1.00	0.96
SD	0.62	0.67	0.69

Note. $d' = z_{FA} - z_H$, where H = hit rate and FA = false-alarm rate (Snodgrass & Corwin, 1988).

Table 8 shows the results for the bias measure *C* across the three timbre-context conditions. Recall that for *C*, a positive value indicates a conservative bias. There was a main effect of timbre context, $F(2, 102) = 4.24$, $MSe = 0.09$, $p < .05$. Pairwise comparisons revealed that participants were less conservative in responding to melodies presented with the studied-same timbre than they were in responding to melodies presented with a studied-different timbre, $t(51) = 2.30$, $p < .05$, or an unstudied timbre, $t(51) = 2.82$, $p < .01$. The bias difference between the studied-different and unstudied timbre conditions did not appear reliable, $t < 0.89$. This pattern of results suggests that participants were less conservative in responding towards melodies when they were presented in the same timbre.

Table 8

Bias (C) Across Timbre-Context Conditions in Experiment 1

	Timbre context		
	Studied-same	Studied-different	Unstudied
M	-0.07	0.04	0.10
SD	0.34	0.40	0.38

Note. $C = 0.5(z_H + z_{FA})$ (Snodgrass & Corwin, 1988).

Overall, the present finding appears compatible with Goh's (2005) finding on voice context effects in recognition memory for spoken words. As predicted, discrimination performance (d') improved substantively only when the studied melodies were repeated in the same timbre at test. The bias measure showed that participants were less conservative in responding towards melodies when they were presented in the studied-same timbre than when they appeared in a studied-different or an unstudied timbre, although the bias difference between the studied-different- and unstudied-timbre-context conditions was not reliable. The present data also replicated previous work that showed an advantage in melody recognition for same-timbre repetitions over different-timbre presentations (e.g., Halpern & Müllensiefen, 2008; Peretz *et al.*, 1998; Radvansky *et al.*, 1995; Wolpert, 1990), and over new-timbre presentations (see Trainor *et al.*, 2004). More important, d' and response bias between the different-timbre and new-timbre conditions did not differ reliably, implicating that instance-specific matching processes *per se* were at work in melody recognition.

One of the main goals of the present project was to specify the processes that govern the timbre effects found in recognition memory for melodies. Experiment 1 showed that the timbre-repetition effect in melody recognition studies is explained solely by the instance-specific matching process, rather than by any timbre-specific familiarity process. Both discrimination accuracy and liberal response tendencies appear to increase only when there was an exact match between the trace of the studied melody and test melody. The question I asked next was whether an *exact same* timbre is always necessary to induce an instance-specific match. Experiment 2 was designed to explore the alternative conditions under which such a match will obtain.

EXPERIMENT 2:

Can a Different (but Similar) Timbre Induce Matching?

Extant studies that examined timbre effects (e.g., Halpern & Müllensiefen, 2008; Peretz *et al.*, 1998; Radvansky *et al.*, 1995; Wolpert, 1990) have paid little attention to effects arising from varying magnitudes of intermediate perceptual differences in their test stimuli. As such, it remains unclear whether these details contributed to the disparate timbre effects observed in these studies. The present experiment was designed to assess the contribution of fine perceptual details to the timbre effects observed in the literature. Specifically, the effects of perceived similarity among the different timbres on recognition memory were explored, by including a similar-timbre condition that allowed a novel comparison between same-, similar-, and distinct-timbre conditions in a single experiment.

Recall that under the encoding specificity framework (Tulving & Thompson, 1973), the effectiveness of a retrieval cue depends on its degree of relatedness to the encoding of an item at first. Timbre information is initially encoded and stored in the memory traces of the melodies, and later used to retrieve or recover the melodies. Because a same-timbre repetition is really an exact match with the memory trace for the old melody, that trace becomes more salient compared to the other competing traces, enabling high discrimination performance at test. On the other hand, a melody presented by a *distinct* timbre, whether previously studied or unstudied, would match the memory trace for the melody only in terms of its structural properties, and not in terms of its timbre properties, resulting in low discrimination performance. Indeed, discrimination performance in the same-timbre condition was found to exceed performance in the distinct-timbre conditions (see both the different- and new-timbre conditions in Experiment 1 used maximally distinct instruments).

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 properties of the exemplar traces in memory and the target probe is an important aspect in exemplar models of memory and categorization (Gillund & Shiffrin, 1984; Hintzman, 1988). A different timbre at test, but which is similar to the original timbre that presented the melody at study, presumably shares many common properties with the original timbre. To the extent that the degree of similarity is high, both the melody presentations at study and test may become matched and coded as the same memory trace (i.e., instance-specific matching occurs). Consequently, discrimination performance in the similar-timbre condition ought to surpass performance in the distinct-timbre condition.

Goldinger (1996) reported an analogous study using spoken words that examined fine perceptual details in repetition effects found in speech. The finding was that perceptual similarity of study and test voices modulated the magnitude of different-voice repetition effects. Specifically, the more similar the test voice was to the study voice, the more probable the listener would classify the word as an old word. If the surface (i.e., timbre) attributes of melodies are analogous to the voice attributes of spoken words (see Goldinger, 1996), discrimination performance is expected to improve even when old melodies are not repeated with the exact same timbre, but with a similar timbre in its place instead, during the recognition stage.

Method

Participants

Forty-two introductory psychology students with differential music training experience from the National University of Singapore participated for course credit. None had participated in the preliminary studies and 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 design.

Based on the Euclidean estimates of the scaling solution in Figure 1, six different combinations of instruments were derived for melody presentation. For each combination, the instruments are listed in the order that constitutes the same-, similar-, and distinct-timbre-context conditions, respectively: (1) piano, harpsichord, violin, (2) harpsichord, piano, clarinet, (3) violin, cello, flute, (4) cello, violin, piano, (5) flute, clarinet, harpsichord, and (6) clarinet, flute, cello (see Table 9). Care was especially taken in selecting instruments for the similar-timbre condition, in order to preserve the integrity of the manipulation. For instance, when violin has been assigned to the same-timbre condition, viola would not be designated as violin's similar-timbre counterpart, because viola might be perceived as *virtually identical*, rather than *similar*, to violin due to its extreme similarity to violin. Set combination was counterbalanced across participants.

Table 9

Six Set Combinations of Instruments Derived for Melody Presentation at Test in Experiment 2

	Timbre context		
Set combination	Same	Similar	Distinct
1	Piano	Harpsichord	Violin
2	Harpsichord	Piano	Clarinet
3	Violin	Cello	Flute
4	Cello	Violin	Piano
5	Flute	Clarinet	Harpsichord
6	Clarinet	Flute	Cello

As in Experiment 1, there were two equivalent lists of 24 melodies each that were designated as the old melodies and new melodies, respectively. In the study phase, all the 24 old melodies were presented by a single instrument. In the test phase, the 24 new melodies were divided among three instruments so that the same instrument, a similar instrument, and a distinct instrument each presented eight melodies. For the 24 old melodies, eight were assigned to the same-timbre-context condition, eight to the similar-timbre-context condition, and the remaining eight to the distinct-timbre-context condition (see Table 10 for a summary).

Table 10

Summary of the Design Used in Experiment 2

Memorization stage	Recognition stage					
Study melodies	Test melodies (Old)			Test melodies (New)		
	Timbre context					
	Same	Similar	Distinct	Same	Similar	Distinct
E.g., Set combination 1 instruments						
Piano	Piano	Harpsi-chord	Violin	Piano	Harpsi-chord	Violin
24	8	8	8	8	8	8
E.g., Set combination 5 instruments						
Flute	Flute	Clarinet	Harpsi-chord	Flute	Clarinet	Harpsi-chord
24	8	8	8	8	8	8

Note. Numbers indicate the quantity of melodies played by the respective timbres in each classification.

Results and Discussion

As in Experiment 1, hits, false alarms, *d′* and *C* were used to measure melody recognition performance. Here, there were three levels of timbre context for both the old and the new melodies – same, similar, and distinct. As such, three hit rates and three false alarm rates were obtained, and *d′* and *C* for each of the three conditions were calculated based on their individual hit and false alarm rates.

Musical training influences. Before proceeding to the main findings involving the three timbre-context conditions, the potential effects of musical training for the sample were first examined, with musical training as the between-subjects factors (participants who had at least four years of formal music training, $n = 10$, versus those without or with less than four years of music training, $n = 32$), and timbre context as the within-subjects factor.

No reliable interaction between musical training and timbre context for hits, false alarms, *d′*, and *C* emerged, all *F*s < 2.43. This shows that musical training did not influence the timbre effects in all measures. A main effect of musical training was significant for *d′*, false alarms, and *C*, *F*s(1, 40) = 8.17; 22.75; 10.85, *MSe*s = 0.36; 0.02; 0.16, all *p*s < .01, respectively. Participants with at least four years of formal music training had higher discrimination scores (*M* = 1.36, *SD* = 0.66), fewer false alarms (*M* = 22.97, *SD* = 12.44), and were more conservative in making their responses at test (*M* = 0.15, *SD* = 0.32) than those without or with less than four years of music training (*d′ M* = 0.93, *SD* = 0.66; false alarms *M* = 37.74, *SD* = 15.32; *C M* = -0.12, *SD* = 0.38). There was no main effect of musical training for hits, *F* < 1. The present data were also reanalysed using musical training as a covariate, and the pattern of results did not change. Since musical training did not interact with timbre context, the main findings for the timbre-context conditions reported below can be generalized across all participants within the sample, and all subsequent tabulations of results are collapsed across musical training.

Timbre context effects. Table 11 presents the pattern of results for hit performance across the three timbre-context conditions. There was a main effect of timbre context, $F(2, 80) = 10.35$, $MSe = 0.02$, $p < .01$. Pairwise comparisons revealed that there were reliably more hits when melodies were repeated with the same timbre than when melodies were presented with a distinct timbre, $t(41) = 4.00$, $p < .001$. Participants also performed better when melodies were repeated using a similar timbre than they did when melodies were repeated with a distinct timbre, $t(41) = 2.55$, $p < .05$. The number of hits did not differ reliably between the same- and similar-timbre-context conditions, $t < 1.54$. This pattern of results indicates that the number of hits increased significantly when melodies were repeated at least in a timbre that was similar to their original timbre at study.

Table 11

Percentage of Hits Across Timbre-Context Conditions in Experiment 2

	Timbre context		
	Same	Similar	Distinct
M	74.87	69.84	61.64
SD	15.87	17.79	15.12

Table 12 shows the false alarm pattern. There was no reliable evidence of a main effect of timbre context, $F < 2.89$, indicating that timbre context did not influence the false alarm rates.

Table 12

Percentage of False Alarms Across Timbre-Context Conditions in Experiment 2

	Timbre context		
	Same	Similar	Distinct
M	36.31	31.66	35.79
SD	16.23	16.24	15.77

Table 13 presents the pattern of results for d' performance across the three timbre-context conditions. There was a reliable main effect of timbre context, $F(2, 80) = 7.68$, $MSe = 0.40$, $p < .01$. Pairwise comparisons revealed that participants were significantly better at discriminating melodies that appeared in the same timbre than they were at discriminating melodies presented with a distinct timbre at test, $t(41) = 2.95$, $p < .01$; participants also performed better when melodies appeared in a similar timbre than they did when melodies appeared in a distinct timbre at test, $t(41) = 2.75$, $p < .01$. Discriminability did not differ reliably between the same and similar timbre-context conditions, $t < 0.79$. This pattern of results indicates that when melodies were tested at least in a similar timbre, discrimination performance improved significantly.

Table 13

Discrimination Performance (d') Across Timbre-Context Conditions in Experiment 2

	Timbre context		
	Same	Similar	Distinct
M	1.17	1.16	0.76
SD	0.75	0.65	0.68

Table 14 shows the results for the bias measure *C* across the three timbre-context conditions. There was a main effect of timbre context, $F(2, 80) = 3.49$, $MSe = 0.13$, $p < .01$. Pairwise comparisons revealed that participants tended to be less conservative in responding to melodies that appeared in the same timbre than they were in responding to melodies presented with a similar timbre, $t(41) = 1.90$, $p < .07$, or a distinct timbre, $t(41) = 3.38$, $p < .01$. The bias difference between the similar and distinct timbre conditions was not reliable, $t < 0.75$. This pattern of results suggests that participants were less conservative in responding towards melodies that were presented in the timbre that was heard during study.

Table 14

Bias (C) Across Timbre-Context Conditions in Experiment 2

	Timbre context		
	Same	Similar	Distinct
M	-0.20	-0.01	0.04
SD	0.41	0.44	0.33

The present experiment showed an advantage in melody discrimination for same-timbre repetitions over distinct-timbre presentations; participants were less conservative in responding towards melodies when they were presented in the same timbre than when they appeared in a different timbre *per se*, replicating Experiment 1's trend. The present finding appears compatible with Goldinger's (1996) finding on voice similarity effects in recognition memory for spoken words. As predicted, recognition in both the same-timbre and similar-timbre conditions were comparable; discrimination performance improved substantially even when melodies were not tested in the same-timbre context.

These data corroborate those from Experiment 1 to emphasize that timbre similarity primarily constitutes an integrated part of the matching and retrieval processes involved in melody recognition. More important, the present result extended the findings from Experiment 1 to suggest that the use of an *exact same* timbre between study and test was in fact not the only way to create an instance-specific match. Presenting the test melody in a different, but similar, timbre appears to be comparably effective in inducing matching.

Taken together, these two experiments demonstrated that music and speech converge in several aspects. An extended discussion on these similarities will be deferred to Chapter 6.

CHAPTER 4

Articulation Similarity Scaling

In Chapter 3, two experiments that examined the nature of timbre effects were reported. Recall that in the music domain, surface characteristics include, other than timbre, the exact pitch level and tempo (see Trainor *et al.*, 2004). While previous studies (see Trainor *et al.*, 2004) have examined the effects of these performance characteristics on melody recognition, the effects of a type of surface characteristics called *articulation* remain unexplored. 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. This chapter describes a third preliminary study that was conducted as part of the present project to first establish the degree of perceived similarity among different articulation formats. Chapter 5 will then be devoted to describe the main experiments that were conducted to investigate the effects of varying articulation format on melody recognition.

To examine the effects of articulation format on melody recognition, the melody was designed 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). A total of six mixed-articulation formats were planned for the present study. 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 4. 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 that lasts through the whole of the final bar.

My initial prediction was that among the eight different articulation formats, a melody presented fully in *legato* form would be perceived as most different from the same melody presented fully in *staccato* form. In other words, perceived articulation similarity was assumed 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.

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 4. Schematic of the eight different articulation format manipulations.

Prior to conducting the main experiments, this prediction was tested 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. This section describes the steps taken to collect similarity ratings and the generation of the "articulation map" using MDS techniques (Kruskal & Wish, 1978).

METHOD

Participants

Sixteen introductory psychology students from the National University of Singapore participated for course credit. None had participated in the previous preliminary studies or Experiments 1 and 2.

Materials and Apparatus

The stimulus set comprised of 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 quadruple time. The instruments used to present the melodies were violin and clarinet. The equipment used was identical to that in the first preliminary study.

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 minutes 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 which was not used during segment one earlier) was now used to present the melody throughout both part one (familiarization phase) and part two (rating phase), and an alternative melody that differed in meter and tonality was now used. The whole session lasted approximately 25 minutes. Participants were debriefed at the end of the session.

RESULTS AND DISCUSSION

MDS using the ALSCAL routine of SPSS version 16 was used to analyze these perceptual similarity data. Figure 5 shows the articulation format map from the ALSCAL solution derived by collapsing across all participants, the two test timbres and the four test melodies. Recall that 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). Therefore, a two-dimensional scaling solution was derived because there were eight articulation forms.

Recall that 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.

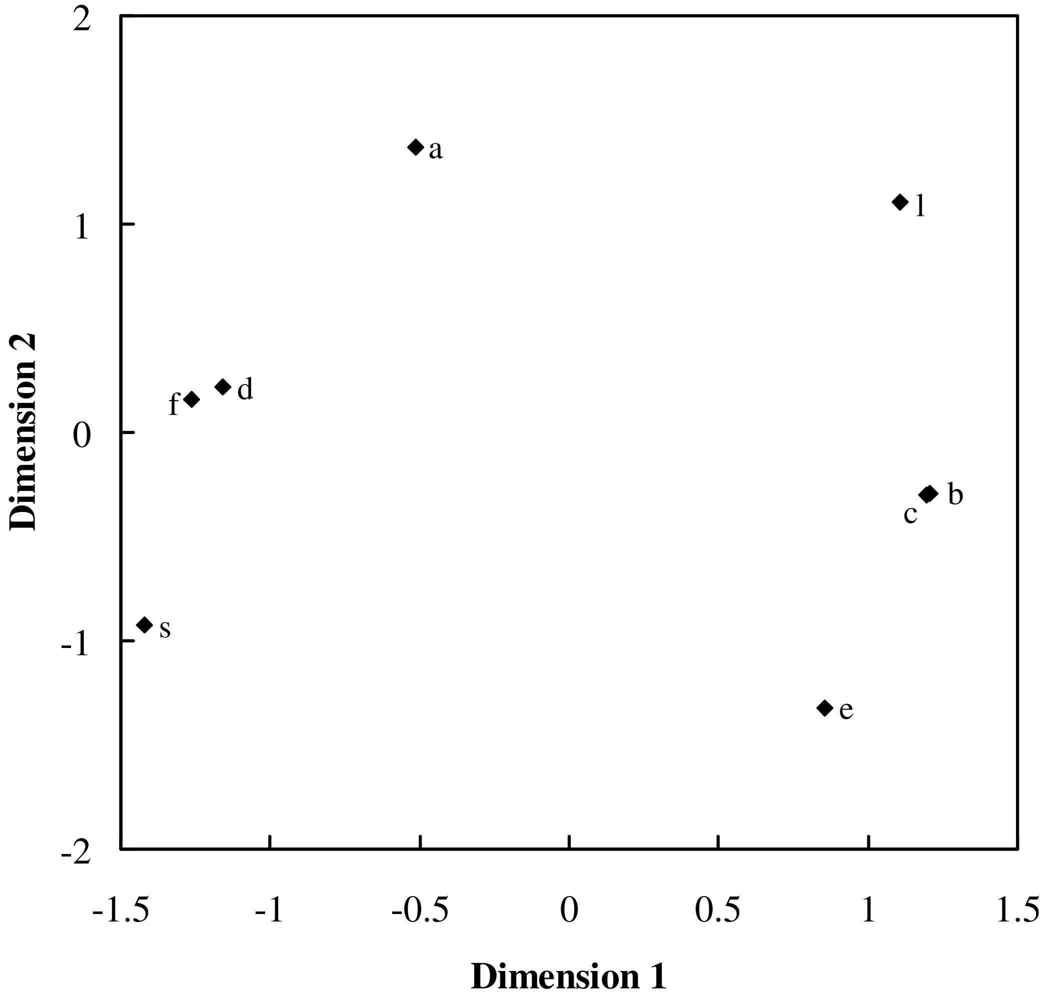


Figure 5. Two-dimensional MDS solution for eight articulation formats.

It should be noted that a definitive determination of the dimensions is not directly critical for the forthcoming 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 4 and Dimension 1 of Figure 5).

I therefore attempted to determine the possible dimensions that participants could be using when making the similarity judgments. 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 to formats *d* and *f*. The nature of the second dimension appeared to be less definitive.

CHAPTER 5

Establishing Articulation Effects in Melody Recognition

This chapter describes two experiments that were conducted to establish the influence of varying articulation format on melody recognition. Experiment 3 established the articulation effects in melody recognition, and allowed a comparison between the effects of articulation and timbre attributes in influencing melody recognition. Following which, Experiment 4 investigated the extent to which location of the articulation match *per se* is critical in inducing matching and determining discrimination performance in melody recognition.

EXPERIMENT 3:

Are Articulation and Timbre Attributes Functionally Similar?

In Experiments 1 and 2, the traditional timbre effects in melody recognition (e.g., Peretz *et al.*, 1998) were replicated. When the same timbre was used for the memory probe, an exact match between overlapping timbre attributes of the memory trace and probe was obtained. As a result, discrimination performance was enhanced. More important, in Experiment 2, when a different, but similar, timbre was used for the memory probe, discrimination performance was in fact comparable to that in the same-timbre condition. The data suggest that timbre similarity constitutes an integrated part of the matching and retrieval processes involved in melody recognition. In the similar-timbre condition, a close match between the overlapping timbre attributes of the memory trace and probe would have obtained, such that using a similar timbre to present the old melody at test appears to be as effective as using the same timbre in inducing matching. These data appear compatible with exemplar models of memory and categorization (Gillund & Shiffrin, 1984; Hintzman, 1988) that emphasize the importance of the degree of similarity among the properties of the exemplar traces in memory and the target probe in aiding effective stimulus retrieval.

As introduced earlier in Chapter 4, articulation format constitutes a type of surface characteristics of melodies. In the extant literature that examined the effects of surface characteristics on melody recognition performance, no study to date has manipulated articulation format. Experiment 3 was specifically designed to investigate the effects of manipulating articulation context on melody recognition. Based on the data from Experiments 1 and 2, timbre information is preserved in LTM, where timbre similarity constitutes an integrated part of the matching and retrieval processes involved in melody recognition. To the extent that the articulation attributes of melodies are similar to the timbre attributes in influencing melody recognition, discrimination performance ought to improve when old melodies are repeated in the same articulation format, as compared to when the melodies are repeated in a distinct articulation format during the recognition stage. Performance ought to improve even when old melodies are repeated with a different, but similar, articulation format, as compared to when the melodies are repeated with a distinct articulation format. In addition, participants ought to be less conservative in responding towards melodies when they were tested in the same articulation than when they appeared in a different articulation *per se*.

Method

Participants

Forty-seven introductory psychology students with varying music training experience from the National University of Singapore participated for course credit. None had participated in any of the preliminary studies or Experiments 1 and 2.

Materials and Apparatus

The stimulus set comprised of the 48 single-line (monophonic) melodies that were used in Experiments 1 and 2.

A multidimensional articulation map was created, showing the similarity relations between the individual articulation formats that were used as the stimulus materials. This map allowed the selection of specific articulation formats to be based on objective measures of the degree of perceived similarity among the different formats. The procedures to collect similarity ratings and to generate the articulation map using multidimensional scaling (MDS) techniques (Kruskal & Wish, 1978) are described in Chapter 4.

Based on the scaling solution, object coordinates in the space were used to estimate perceptual distances between all articulation formats. The planar coordinates of the articulation formats as well as the Euclidean distance between each articulation format pair are shown in Appendix C. Based on these Euclidean estimates, 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* (see Table 15), where formats *l* and *s*, based on their Euclidean distance, are maximally distinct from each other. Set combination was counterbalanced across participants. The equipment used was the same as that in Experiments 1 and 2.

Table 15

Two Set Combinations of Articulation Formats Derived for Melody Presentation at Test in Experiment 3

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

Design

As in Experiment 2, there were two equivalent lists of 24 melodies each that were designated as the old melodies and new melodies, respectively. In the study phase, 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 so that the same format, a similar format, and a distinct format each presented eight melodies. For the 24 old melodies, 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 16 for a summary).

Procedure

Half of the participants were randomly allocated to listen to melodies played by the clarinet, while the other half were randomly allocated to listen to melodies played by the violin. The present procedure was similar to that used in Experiment 1. Figure 6 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 first part was the study phase and took approximately five minutes to complete. Participants were instructed to silently memorize each melody that was presented over the headphones. At the start of each trial, a ready prompt was displayed on the monitor for one second, after which it was erased. One second later, a single melody was played over the headphones; 800 ms following the first presentation, the melody was repeated. After two instances of each melody were presented, participants pressed the space key to proceed to the next melody. This sequence continued until all 24 melodies were presented. The melody presentation sequence was random across participants. Participants were informed of a forthcoming recognition test.

Table 16

Summary of the Design Used in Experiment 3

Memorization stage	Recognition stage					
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

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

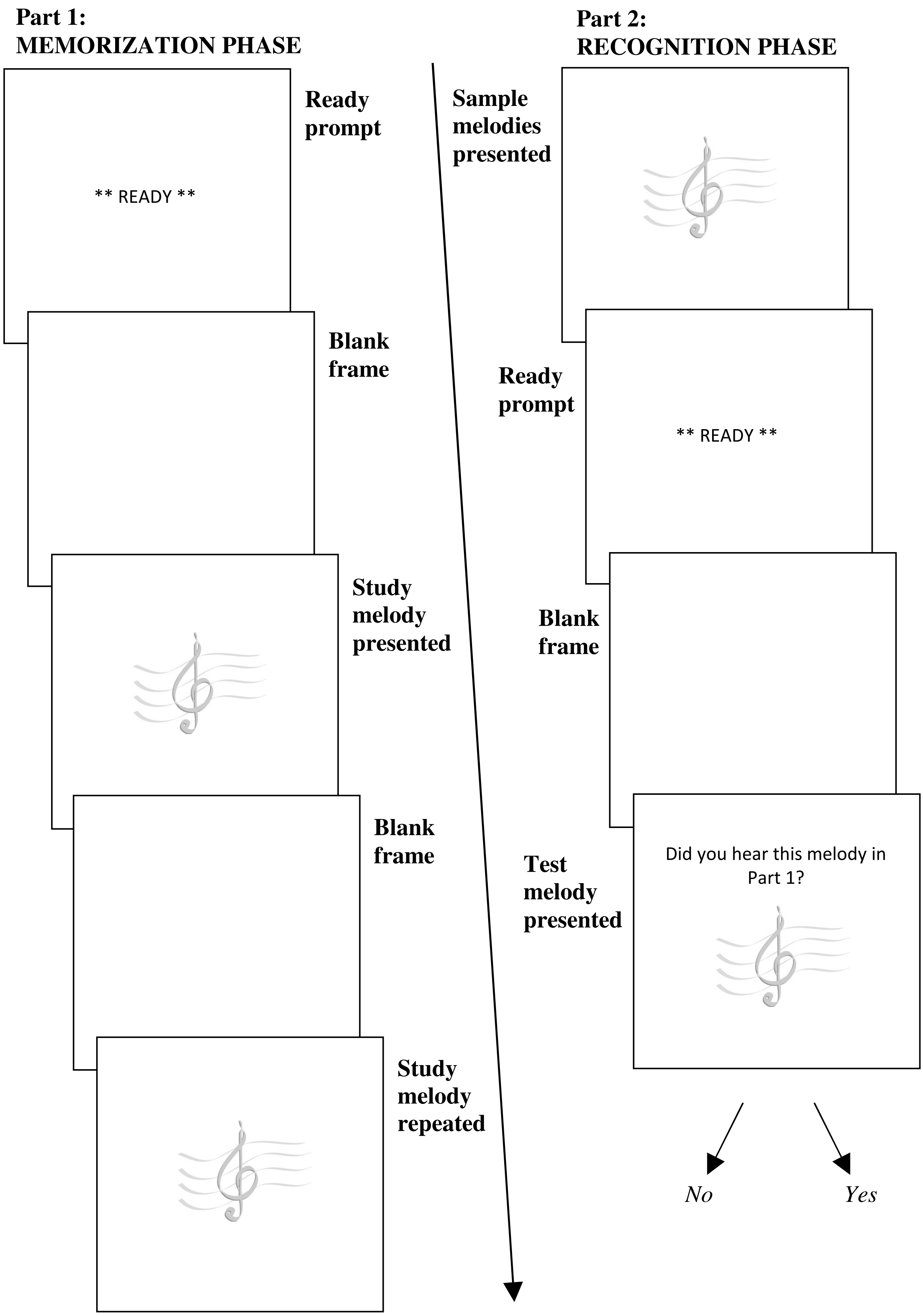


Figure 6. Schematic of the sequence of a trial in Experiment 3.

Following the study 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 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 minutes to complete all 48 randomly presented trials.

After the recognition test, information on the participants' music training experience was recorded. 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. Participants were debriefed at the end of the session.

Results and Discussion

As in Experiments 1 and 2, hits, false alarms, *d′*, and *C* were used to measure melody recognition performance. There were three levels of articulation context for both the old and the new melodies – same, similar, and distinct; three hit rates and three false alarm rates were obtained, and *d′* and *C* for each of the three conditions were calculated based on their individual hit and false alarm rates.

Musical training influences. Before proceeding to the main findings involving the three articulation-context conditions, the potential effects of musical training for the sample were first examined, with musical training as the between-subjects factor (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, $n = 32$) and articulation context as the within-subjects factor.

No reliable interaction between musical training and articulation context was observed for hits, false alarms, *d′*, and *C*, all Fs < 1. This shows that musical training did not influence the articulation effects in all measures. There was no reliable main effect of musical training for *d′*, hits, false alarms, and *C*, all Fs < 1. The present data were also reanalysed using musical training as a covariate, and the pattern of results did not change. Since musical training did not interact with articulation context, the main findings for the articulation-context conditions reported below can be generalized across all participants within the sample, and all subsequent tabulations of results are collapsed across musical training.

Articulation context effects. Table 17 presents the pattern of results for hit performance across the three articulation-context conditions. There was a main effect of articulation context, $F(2, 92) = 23.62$, $MSe = 0.03$, $p < .001$. Pairwise comparisons revealed that there were reliably less hits when melodies were repeated with a distinct articulation than when melodies were presented with the same articulation, $t(46) = 5.83$, $p < .001$, or a similar articulation, $t(46) = 4.99$, $p < .001$; the number of hits also differed between the same- and similar-articulation-context conditions, $t(46) = 2.32$, $p < .05$. This pattern of results suggests that the more similar the articulation format used at test was to the original format at study, the greater the number of hits was at test.

Table 17

Percentage of Hits Across Articulation-Context Conditions in Experiment 3

	Articulation context		
	Same	Similar	Distinct
M	70.57	63.71	48.58
SD	15.88	13.35	18.18

Table 18 shows the false alarm pattern. There was a main effect of articulation context, $F(2, 92) = 4.89$, $MSe = 0.03$, $p < .05$. Pairwise comparisons revealed that there were reliably more false alarms when melodies appeared in the same articulation format than when melodies were presented with a similar articulation, $t(46) = 2.13$, p

< .05, or a distinct articulation, *t*(46) = 2.67, $p < .05$; the number of false alarms did not differ between the similar- and distinct-articulation-context conditions, $t < 0.98$. Overall, this pattern of results indicates that participants made significantly more false alarms when melodies appeared in the same articulation format at test than when they appeared in a different format *per se*.

Table 18

Percentage of False Alarms Across Articulation-Context Conditions in Experiment 3

	Articulation context		
	Same	Similar	Distinct
M	37.71	32.03	27.07
SD	18.28	15.81	16.93

Table 19 presents the pattern of results for *d'* performance across the three articulation-context conditions. There was a reliable main effect of articulation context by participants, $F(2, 90) = 3.94$, $MSe = 0.36$, $p < .05$. Pairwise comparisons revealed that participants were significantly better at discriminating melodies that appeared in the same articulation format than they were at discriminating melodies presented with a distinct articulation format at test, $t(46) = 2.42$, $p < .05$; participants also performed better when melodies appeared in a similar articulation format than they did when melodies appeared in a distinct format at test, $t(46) = 2.03$, $p < .05$. Discriminability did not differ between the same- and similar-articulation-context conditions, $t < 1.05$. Overall, this pattern of results indicates that discriminability increased significantly so long as melodies were tested in at least a similar articulation format.

Table 19

Discrimination Performance (d') Across Articulation-Context Conditions in Experiment 3

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

Table 20 shows the results for the bias measure *C* across the three articulation-context conditions. There was a main effect of articulation context, $F(2, 92) = 18.99$, $MSe = 0.16$, $p < .001$. Pairwise comparisons revealed that participants were less conservative in responding to melodies that appeared in the same articulation format than they were to melodies presented with a similar articulation format, $t(46) = 2.89$, $p < .01$, or a distinct format, $t(46) = 5.12$, $p < .001$; the bias difference between the similar- and distinct-timbre conditions was also reliable, $t(46) = 3.96$, $p < .005$. This pattern of results suggests a bias effect towards increasing conservatism as articulation format mismatches increase from the same- to similar- to distinct-articulation contexts.

Table 20

Bias (C) Across Articulation-Context Conditions in Experiment 3

	Articulation context		
	Same	Similar	Distinct
M	-0.12	0.08	0.38
SD	0.41	0.34	0.45

The present data revealed an advantage in melody discrimination for same-articulation repetitions over distinct-articulation presentations. There was also an advantage in melody recognition for similar-articulation presentations over distinct-articulation presentations. In addition, participants were the least conservative in responding towards melodies that appeared in the same format at test. Overall, this pattern of results is compatible with the pattern observed in the earlier experiments of this project, although a bias difference was now observed between the similar-articulation- and distinct-articulation-context conditions (cf. Experiment 2). When the melody was presented in the same format at test, a repetition advantage effect obtained. Furthermore, matching between the two instances of the melody from study to test occurred regardless whether the test melody appeared in the same, or a different but similar, format, enhancing discrimination accuracy. The present data revealed that articulation attributes of melodies resemble timbre attributes in their capacity to influence melody recognition accuracy.

EXPERIMENT 4:

Does *Perception* Always Determine *Performance*?

An examination of the articulation similarity scaling solution reported in Chapter 4 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 first demonstrated that instance-specific matching, rather than timbre-specific familiarity, processes contribute to melody recognition. Subsequently, Experiment 3 showed that when two instances of a melody were perceptually similar in terms of their match in articulation format from study to test, instance-specific matching presumably 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. 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.

The question I raised next was whether this perceptual dissimilarity between two instances of the melody (e.g., in formats *d* and *e*) due to the location of the (mis)match would 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 data thus far suggest that to the extent that two instances of a melody are perceptually similar to each other (see Experiments 2 and 3), matching obtains. While the absolute amount of match in timbre or articulation format appears to determine this perceptual similarity and, consequently, discrimination performance, the goal of Experiment 4 was to foreclose 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 from the National University of Singapore volunteered to participate in the experiment. None had participated in any of the preliminary studies or Experiments 1, 2, and 3.

Materials, Apparatus, Design, and Procedure

The materials and procedures were essentially the same as those of Experiment 3, with a slight modification in materials.

Based on the Euclidean estimates of the articulation similarity scaling solution in Figure 5, 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* (see Table 21). Set combination was counterbalanced across participants. Table 22 shows a summary of the present design.

Table 21

Four Set Combinations of Articulation Formats Derived for Melody Presentation at Test in Experiment 4

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

Table 22

Summary of the Design Used in Experiment 4

Memorization stage	Recognition stage					
Study melodies	Test melodies (Old)			Test melodies (New)		
	Articulation format context					
	Same	Similar	Distinct	Same	Similar	Distinct
	E.g., Set combination 1 articulation formats					
s	s	d	e	s	d	e
24	8	8	8	8	8	8
	E.g., 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.

Results and Discussion

Hits, false alarms, *d'*, and *C* were used to measure melody recognition performance. As in Experiment 3, there were three levels of articulation context for both the old and the new melodies – same, similar, and distinct; three hit rates and three false alarm rates were obtained, and *d'* and *C* for each of the three conditions were calculated based on their individual hit and false alarm rates.

Musical training influences. Before proceeding to the main findings involving the three articulation-context conditions, the potential effects of musical training for the sample were first examined, with music training as the between-subjects factor (participants who had at least four years of formal music training, $n = 20$, versus those without or with less than four years of music training, $n = 44$) and articulation context as the within-subjects factor.

No reliable interaction between musical training and articulation context were observed for hits, false alarms, *d'*, and *C*, all *F*s < 2.50. This shows that musical training did not influence the articulation effects in all measures. A main effect of musical training was significant for *d'* and hits, $Fs(1, 62) = 4.68$; 5.55, *MSe*s = 0.63; 0.03, *p*s < .05, respectively. Participants with at least four years of formal music training had higher discrimination scores ($M = 1.24$, $SD = 0.81$) and more hits ($M = 69.25$, $SD = 14.35$) than those without or with less than four years of music training (*d'* $M = 0.97$, $SD = 0.66$; hits $M = 62.88$, $SD = 17.00$). There was no reliable main effect of musical training for false alarms and *C*, *F*s < 1. The present data were also reanalyzed using musical training as a covariate, and the pattern of results did not change. Since musical training did not interact with articulation context, the main findings for the articulation-context conditions reported below can be generalized across all participants within the sample, and all subsequent tabulations of results are collapsed across musical training.

Articulation context effects. Table 23 presents the pattern of results for hit performance across the three articulation-context conditions. There was a main effect of articulation context $F(2, 126) = 14.92$, $MSe = 0.03$, $p < .001$. Pairwise comparisons revealed that there were reliably more hits when melodies were repeated with the same articulation format than when melodies were presented with a similar format, $t(63) = 5.21$, $p < .001$, or with a distinct format, $t(63) = 3.94$, $p < .001$. The number of hits did not differ reliably between the similar- and distinct-articulation-context conditions, $t < 1.46$. This pattern of results indicates that the number of hits increased significantly when melodies were repeated in their original format at test.

Table 23

Percentage of Hits Across Articulation-Context Conditions in Experiment 4

	Articulation context		
	Same	Similar	Distinct
M	73.44	58.51	62.67
SD	13.00	19.77	16.55

Table 24 shows the false alarms pattern. There was a main effect of articulation context, $F(2, 126) = 6.39$, $MSe = 0.02$, $p < .005$. Pairwise comparisons revealed that there were reliably more false alarms when melodies appeared in the same articulation format than when melodies were presented with a similar format, $t(63) = 3.14$, $p < .005$, or with a distinct format, $t(63) = 3.00$, $p < .005$. The number of hits did not differ reliably between the similar- and distinct-articulation-context conditions, $t < 0.07$. This pattern of results indicates that the number of false alarms increased significantly when melodies were presented in the articulation format that was used during study.

Table 24

Percentage of False Alarms Across Articulation-Context Conditions in Experiment 4

	Articulation context		
	Same	Similar	Distinct
M	34.55	26.39	26.22
SD	19.26	13.93	15.70

Table 25 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 25

Discrimination Performance (d') Across Articulation-Context Conditions in Experiment 4

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

Table 26 shows the results for the bias measure *C* across the three articulation-context conditions. There was a main effect of articulation context, $F(2, 126) = 17.21$, $MSe = 0.13$, $p < .001$. Pairwise comparisons revealed that participants were less conservative in responding to melodies that appeared in the same articulation format than they were in responding to melodies presented with a similar articulation format, $t(63) = 5.72$, $p < .001$, or a distinct format, $t(63) = 4.23$, $p < .001$. The bias difference between the similar and distinct articulation conditions was not reliable, $t < 0.71$. This pattern of results suggests that participants were less conservative in responding towards melodies that were presented in the articulation format that was heard during study.

Table 26

Bias (C) Across Articulation-Context Conditions in Experiment 4

	Articulation context		
Measures	Same	Similar	Distinct
M	-0.12	0.23	0.19
SD	0.39	0.37	0.38

The present data revealed that while participants were less conservative in responding towards melodies that were tested in the same-articulation context, discrimination performance was comparable across the same-, similar-, and distinct-articulation conditions. The important implication is that perceptual dissimilarity, as a function of the location of (mis)match in articulation format, did not appear to modulate melody recognition accuracy. For instance, even though formats *d* and *e* were *perceived* as different from each other due to the fact that they differ 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 contain 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*.

CHAPTER 6

General Discussion and Conclusions

The general goal of this project was to examine the underlying nature of the representational units used in melody recognition. The overarching question I asked was: To what extent are the surface features of melodies represented and utilized in memory? Four main experiments were conducted to determine if information about surface characteristics, specifically timbre and articulation attributes, is encoded and stored in LTM and how these performance attributes influence discrimination performance during melody recognition. Three specific research issues were explored. First, are timbre-specific familiarity processes or instance-specific matching processes, or both types of processes, responsible for the traditional timbre effects found in melody recognition? Second, what is the function of timbre similarity in these timbre effects? Third, do articulation format attributes influence melody recognition memory?

In the sections below, I summarize the main results from the experiments and discuss the implications of these findings. Following which, the general implications of the present findings for the nature of music processing will be discussed. Finally, I will draw general conclusions based on these findings, and offer suggestions for future work.

SUMMARY AND IMPLICATIONS OF MAJOR FINDINGS

Instance-Specific Matching Effects in Melody Recognition

The nature of timbre effects in recognition memory for melodies played by multiple instruments was investigated in Experiment 1 of this project, by comparing performance when studied melodies were repeated in the same, in different, or in new timbres at test. Melodies that remained in the same timbre from study to test were recognized better than were melodies that were presented in a previously studied but different, or previously unstudied (new) timbre at test; recognition for melodies that were presented in a different timbre at test did not differ reliably from recognition for melodies in a new timbre at test. In addition, participants were more liberal in responding towards melodies when they were tested in the studied-same timbre than when they appeared in a studied-different or an unstudied timbre.

Recall that while differences between same- and new-timbre contexts can be attributed to either encoding specificity (instance-specific matching) or a global timbre-specific familiarity with a previously studied timbre, or to both of these processes, differences between same- and different-timbre contexts are attributed to only the specificity of match between the memory trace and the recognition probe. Timbre-specific familiarity *per se* is implicated when differences are observed between the different- and new-timbre conditions because there are no instance-specific matches in either condition. Experiment 1 included and compared these two conditions to enable an investigation of the extent to which the timbre-specific familiarity process *per se* contributes to discrimination performance in recognition memory for melodies. The data revealed no difference in discrimination performance between the different- and new-timbre contexts. Additionally, the bias measure revealed no response bias difference between these two conditions, implicating that instance-specific matching processes *per se* govern melody recognition.

An interpretation based on the encoding specificity principle (Tulving & Thompson, 1973) offers a good fit with these data. 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. The present view is that surface (e.g., timbre) and structural attributes of a melody are stored together in the LTM trace. It appears that recognition performance would be enhanced only when both the trace of the studied melody and the test melody match exactly; performance would be hampered so long as there is some clear mismatch of the melody's surface attributes between study and test. In other words, discrimination performance in melody recognition improves only when instance-specific matching obtains.

Timbre Similarity Effects in Melody Recognition

Was the *exact same* timbre always necessary to induce an instance-specific match in melody recognition? More specifically, would a different, but similar, timbre suffice to induce matching? To directly address these questions, the effects of timbre similarity in recognition memory for melodies played by multiple instruments were investigated in Experiment 2 of this project, by comparing recognition performance when studied melodies were repeated in the same, in similar, or in distinct timbres at test. Melodies that remained in the same timbre from study to test were recognized better than were melodies that were presented in a distinct timbre at test. But when a timbre that was different from, but similar to, the original timbre played the melodies at test, recognition was comparable to that when the same timbre played them.

An interpretation based on exemplar models of memory and categorization (Gillund & Shiffrin, 1984; Hintzman, 1988) offers a good account of these data. These exemplar models assume that the degree of similarity among the properties of the exemplar traces in memory and the target probe is an important aspect. When a similar timbre was used for the memory probe, it presumably shared many common timbre properties with the original timbre during study. As a result, a close match between the overlapping timbre attributes of the memory trace and probe occurred. The present data suggest that the use of a similar timbre to present the old melody at test constitutes an alternative way that was comparable to using the exact same timbre of creating an instance-specific match between the study melody and the test melody.

Similarities Between Music and Speech Processing

Overall, Experiments 1 and 2 demonstrated that music and speech converge in several aspects. The data from Experiment 1 suggest that discrimination performance improved substantively only when the old melodies were repeated with the same timbre at test. The novel finding that was added to the extant literature examining timbre effects was that discrimination performance between the different-timbre and new-timbre conditions did not differ reliably, implicating that instance-specific matching processes *per se*, rather than any timbre-specific familiarity processes, underlie melody recognition.

The present data are compatible with Goh's (2005) finding in an analogous study using spoken words to examine voice context effects in recognition memory. Goh (2005) reported that discrimination performance improved only when the exact same voice was repeated at test, suggesting that the voice-specific attributes of individual talkers are preserved in LTM. It appears that the surface (timbre) attributes of melodies are analogous to the nonlinguistic (voice) attributes of spoken words. The present interpretation is that timbre and abstract structural attributes of a melody, analogous to the nonlinguistic and linguistic properties in speech, are stored together in the LTM trace. When both the trace of the studied melody and the test melody coincide, discrimination performance improves. This implies that some form of instance-specific matching must occur before one can accurately discriminate melodies at test in a melody recognition task, or words in a spoken word recognition task (see Goh 2005).

Experiment 2 was designed to examine the function of perceived similarity in the traditional timbre effects found in melody recognition, by including a third similar-timbre condition that allowed a novel comparison between same-, similar-, and distinct-timbre conditions in a single experiment. The data revealed that the recognition advantage observed in both the same-timbre and similar-timbre conditions were comparable. Discrimination performance improved substantially even when the studied melodies were now repeated with a different, but similar, timbre at test.

The present data are consistent with Goldinger's (1996) finding in an analogous study using spoken words to examine voice similarity effects in recognition memory found in speech. It was reported that perceptual similarity of study and test voices modulated the magnitude of different-voice repetition effects. That is, the more similar the test voice was to the study voice, the more probable the listener would classify the word as an old word. The present data suggest that the surface (timbre) attributes of melodies are analogous to the voice (nonlinguistic) attributes of spoken words. When a different, but similar, timbre was used for the memory probe, a reliable match between the overlapping timbre attributes of the memory trace and probe occurred, enhancing melody recognition performance. This observation is consistent with that in the spoken word recognition literature (see Goldinger, 1996).

Similarities Between Articulation and Timbre Effects in Melody Recognition

Experiment 3 was the first in the literature examining effects of surface features of melodies to manipulate articulation format and explore its effects on melody recognition. In this experiment, recognition performance was compared when studied melodies were repeated in the same, in similar, or in distinct articulation formats at test. Two important findings emerged. First, there was an advantage in melody recognition in the same-articulation condition as compared to the distinct-articulation condition. Analogous to timbre information (see Experiments 1 and 2), articulation information is initially encoded and stored in the memory traces of the melodies, and subsequently used to recover the melodies. A same-articulation repetition constitutes an exact match, in terms of both its structural and performance feature attributes, with the memory trace for the old melody. That trace then becomes more pronounced than the other competing traces, resulting in high discrimination performance at test. But when a melody was presented in a distinct articulation format, it matched the memory trace for the melody only in terms of its structural attributes. Consequently, discrimination performance was weakened.

Second, there was an advantage in melody recognition in the similar-articulation condition as compared to the distinct-articulation condition. Analogous to timbre similarity effects (see Experiment 2), articulation format similarity is integrated in the matching and retrieval processes involved in melody recognition. The use of a similar articulation format for the memory probe resulted in a close match between the overlapping articulation attributes of the memory trace and probe. Thus, matching occurred regardless of whether the melody reappeared in the same or in a similar articulation format during the recognition stage, enhancing discrimination performance. The present data suggest that articulation attributes are homologous to timbre attributes in their capacity to modulate melody discrimination performance.

The Nature of the Instance-Specific Matching Process in Melody Recognition

Experiment 1 demonstrated that instance-specific matching, rather than timbre-specific familiarity, processes underlie melody recognition. Experiments 2 and 3 established that surface feature information of melodies is first encoded and stored in the memory traces of the melodies, and later used to retrieve the melodies. 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. By the same argument, a distinct-feature presentation would not match with the trace for the old melody, thus performance is hampered. The interpretation is that the more similar the retrieval cue is to the initial encoding of the melody in terms of its articulation format, the more effective the cue would be in allowing the melody to be recovered at test.

However, Experiment 4 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 was good, average d' = 1.09.[5] Thus, the logical inference is that whether instance-specific matching would obtain is likely to be contingent on the *absolute 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 instance-specific matching observed in melody recognition under which matching would (or would not) occur.

[5] Recall that 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, I conducted three planned comparisons. The first and second comparisons established that the data sets between Experiments 3 and 4 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 3's distinct-articulation condition as baseline, and revealed that performance in Experiment 4's distinct-articulation condition reliably exceeded performance in this baseline condition, $t(109) = 3.44$, $p < .01$, implicating good discrimination performance in this case.

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 timbre perception and music processing at large. The present finding from Experiment 4 can be 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, I 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 7). 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 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.

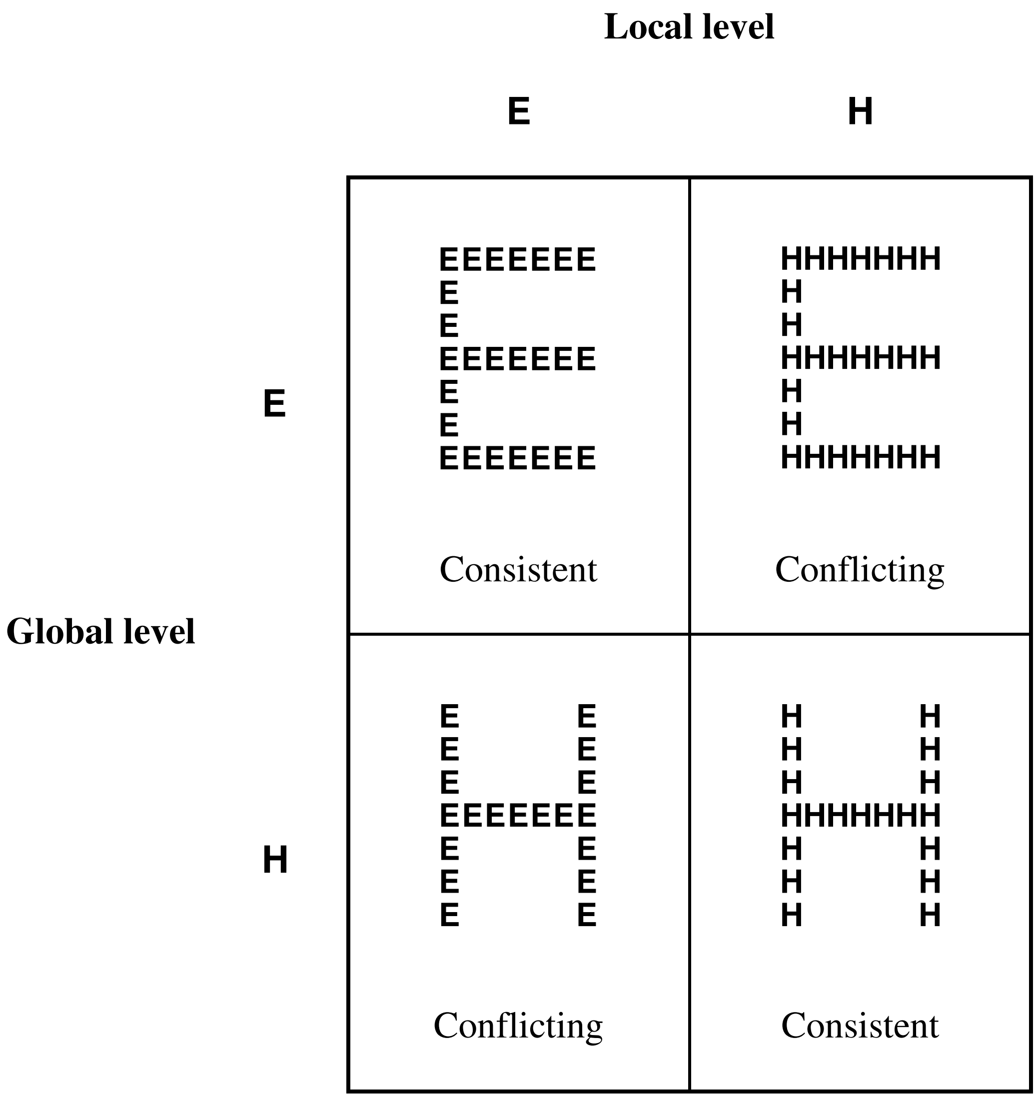


Figure 7. An example of Navon's (1977) type hierarchical stimuli. Large Es and Hs are composed using small Es and Hs.

The present interpretation is that an analogous global advantage mechanism operates in the instance-specific 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 instance-specific 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 obtains, melody discrimination performance is enhanced.

IMPLICATIONS FOR THE NATURE OF MELODY RECOGNITION AND REPRESENTATION

Experiments 2 and 3 of the present study yielded data that are consistent with recent work by Kostic and Cleary (2009) who demonstrated that the exact tempo does not need to be reinstated at test in order for listeners to recognize a previously heard melody. The fact that song recognition can occur reliably when surface features, in their case the tempo of the rhythm, had been changed (i.e., made faster or slower) implies that such recognition can be based on memory for relative tempo (timing) information and, in this case, similar timbre or articulation format information.

Dual-process theories of recognition memory (see Yonelinas, 2002) posit that recognition of a test item can emerge either through recollecting the earlier episode in which the item was previously presented, or through a mere feeling of familiarity with the test item. A substantial number of studies had obtained empirical support for the idea that melody recognition reflects familiarity-based recognition (e.g., Cleary, 2004; Kostic and Cleary, 2009), and the present findings are compatible with this idea and extends it by suggesting the nature of the familiarity processes involved.

It should be noted that familiarity merely *with a studied timbre* (see Experiment 1) is futile in enhancing melody recognition at test, because merely hearing (and becoming familiarized with) a timbre *per se* at study elicits no sense of familiarity for the melody at test later. For instance, two timbres – cello and piano – were studied, and let us suppose that a melody was presented in cello at study. When it reappeared in piano at test, this same melody presumably would not appear familiar to the listener because piano, albeit a familiar timbre *per se* because it was studied, is primarily a perceptually distinct timbre from cello. The interpretation is that piano did not contribute to any sense of familiarity towards that melody because dissimilarity between the two timbres prevented the test instance of the melody from mapping to its original instance in the memory trace. When mapping fails, melody recognition is hampered.

On the other hand, if the melody were repeated in a similar timbre, regardless of whether this timbre was previously studied or completely new, it invokes a feeling of familiarity towards a melody (see Experiment 2). As a result, familiarity-based melody recognition is enhanced. Suppose a melody was heard in cello at study but reappeared in violin at test. Even where violin was not previously studied (i.e., an unfamiliar timbre *per se*), it shares many common timbre properties with the original timbre. As such, the test instance of the melody could be mapped successfully with its original instance in the memory trace. Reliable mapping of overlapping features in the two timbres leads to a heightened sense of familiarity for the studied melody which in turn enhances melody recognition. The same argument extends to the case where overlapping features in two articulation formats between study and test propagate familiarity for the studied melody, leading to reliable recognition performance (see Experiment 3).

Overall, the findings of this project have implications for the role of abstract structure and surface characteristics in music processing and interpretation. Specifically, they support the view that the surface features of a melody actually get encoded, along with structural information, into LTM (e.g., Halpern & Müllensiefen, 2008; Peretz *et al.*, 1998; Radvansky *et al.*, 1995; Wolpert, 1990). This view is compatible with the exemplar models in speech perception (see Pisoni, 1997) which assume that representations of spoken words in memory contain both lexical and indexical information, such that talker information is encoded and used in lexical access and retrieval.

In a similar vein, the representations of melodies in memory are assumed to be very detailed configurations that contain both abstract structural as well as feature information. Information about a melody's performance attributes, such as timbre and articulation format, is encoded and stored in LTM, and utilized in melody access and retrieval later. The retention of such detailed, fine-grained surface feature information in music, analogous to phonetic information in speech, could potentially enhance music perception, because the encoding of peripheral information in musical inputs would reflect how robust music perception is under a wide variety of listening conditions (see Pisoni, 1997).

CONCLUSIONS AND FUTURE DIRECTIONS

This dissertation extended previous work that examined the effects of surface feature information on memory recognition in several novel directions. Experiment 1 offered new insights into the nature of the traditional timbre effects observed in the extant literature – instance-specific matching, rather than timbre-specific familiarity, processes govern these effects. Experiment 2 discovered the contribution of timbre similarity to these effects, demonstrating that the use of a timbre that is perceptually different from, but similar to, the original timbre to present the melody at test provides an alternative way to induce matching effectively. These observations appear compatible with those in the spoken word recognition literature, elucidating several similarities between music and speech. Experiment 3 demonstrated the potency of articulation information, comparable with that of timbre information, to influence the recovery of melodies at the recognition stage. Experiment 4 revealed a new boundary condition of the instance-specific matching process found in melody recognition: That this process will be successful depends on global matching, rather than a localized match, between two instances of the melody.

The present global matching advantage hypothesis can be tested further in a future study that manipulates the overall (global) and local matches in 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 other aspects of music articulation such as 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 disjunct 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, along with the idealized canonical structure of music, serves an indispensable function in music perception and processing.

References

Aslin, R., Saffran, J., & Newport, E. (1992). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, **9**, 321–324.

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

Boltz, M. (1991). Some structural determinants of melody recall. *Memory & Cognition*, **19**, 239–251.

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

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.

Deutsch, D. (2002). The puzzle of absolute pitch. *Current Directions in Psychological Science*, **11**, 200–204.

Fodor, J. A. (1983). *The modularity of mind*. Cambridge: MIT Press/Bradford Books.

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

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

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.

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.

Houston, D., & Jusczyk, P. (2000). The role of talker-specific information in word segmentation by infants. *Journal of Experimental Psychology: Human Perception and Performance*, **26**, 1570–1582.

Ilari, B., & Polka, L. (2002). *Memory for music in infancy: The role of style and complexity.* Paper presented at the International Conference on Infant Studies, Toronto.

Jusczyk, P., & Hohne, E. (1997). Infants' memory for spoken words. *Science*, **277**, 1984–1986.

Kolers, P. A. (1973). Remembering operations. *Memory & Cognition*, **1**, 347–355.

Kostic, B., & Cleary, A. M. (2009). Song recognition without identification: When people cannot "name that tune" but can recognize it as familiar. *Journal of Experimental Psychology: General*, **138**, 146–159.

Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychological Bulletin*, **126**, 159–179.

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.

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: 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.

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

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

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 and J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, **88**, 93–134.

Radvansky, G., Fleming, K., & Simmons, J. (1995). Timbre reliance in nonmusicians’ and musicians’ memory for melodies. *Music Perception*, **13**, 127–140.

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

Saffran, J. R. (2003a). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, **12**, 110–114.

Saffran, J. R. (2003b). Absolute pitch in infancy and adulthood: The role of tonal structure. *Developmental Science*, **6**, 35–43.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, **274**, 1926–1928.

Saffran, J. R., & Griepentrog, G. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology*, **37**, 74–85.

Saffran, J. R., Johnson, E., Aslin, R. N., & Newport, E. L. (1999). Statistical learning of tone sequences by human infants and adults. *Cognition*, **70**, 27–52.

Saffran, J. R., Loman, M., & Robertson, R. (2000). Infant memory for musical experiences. *Cognition*, **77**, B15–B23.

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.

Samson, S., Zatorre, R. J., & Ramsay, J. O. (1997). Multidimensional Scaling of synthetic musical timbre: Perception of spectral and temporal characteristics. *Canadian Journal of Experimental Psychology*, **51**, 307-315.

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

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

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin & Review*, **4**, 145–166.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, **117**, 34–50.

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.

Wolpert, R. (1990). Recognition of melody, harmonic accompaniment, and instrumentation: Musicians vs. nonmusicians. *Music Perception*, **8**, 95–106.

Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language*, **46**, 441–517.

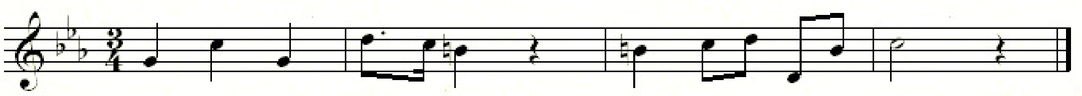
Appendices

Appendix A

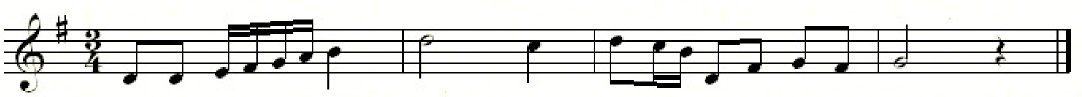
Musical Notations of Sample Melodies Used in the Present Study

Key: C Major
Meter: Simple quadruple

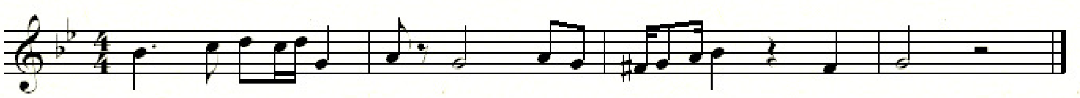
Key: C Minor
Meter: Simple triple



Key: G Major
Meter: Simple triple



Key: G Minor
Meter: Simple quadruple



Appendix B

Planar Coordinates of Instruments and Euclidean Distances Between Pairs of Instruments

Instrument pair	Planar coordinates of Instrument 1		Planar coordinates of Instrument 2		Euclidean distance
	x_1	y_1	x_2	y_2	
P_H	1.33	-1.16	1.53	-0.88	0.34
P_EP	1.33	-1.16	1.19	-1.25	0.17
P_Vn	1.33	-1.16	0.87	1.41	2.60
P_Va	1.33	-1.16	0.91	1.37	2.56
P_Ce	1.33	-1.16	0.63	1.34	2.59
P_Ft	1.33	-1.16	-1.03	-0.28	2.52
P_Ob	1.33	-1.16	-0.99	-0.24	2.49
P_Ct	1.33	-1.16	-1.03	-0.12	2.58
P_Tp	1.33	-1.16	-1.21	-0.01	2.78
P_FH	1.33	-1.16	-1.07	0.04	2.68
P_Tb	1.33	-1.16	-1.12	-0.21	2.63
H_EP	1.53	-0.88	1.19	-1.25	0.51
H_Vn	1.53	-0.88	0.87	1.41	2.38
H_Va	1.53	-0.88	0.91	1.37	2.33
H_Ce	1.53	-0.88	0.63	1.34	2.40
H_Ft	1.53	-0.88	-1.03	-0.28	2.63

H_Ob	1.53	-0.88	-0.99	-0.24	2.61
H_Ct	1.53	-0.88	-1.03	-0.12	2.68
H_Tp	1.53	-0.88	-1.21	-0.01	2.88
H_FH	1.53	-0.88	-1.07	0.04	2.76
H_Tb	1.53	-0.88	-1.12	-0.21	2.74
EP_Vn	1.19	-1.25	0.87	1.41	2.68
EP_Va	1.19	-1.25	0.91	1.37	2.64
EP_Ce	1.19	-1.25	0.63	1.34	2.65
EP_Ft	1.19	-1.25	-1.03	-0.28	2.42
EP_Ob	1.19	-1.25	-0.99	-0.24	2.40
EP_Ct	1.19	-1.25	-1.03	-0.12	2.49
EP_Tp	1.19	-1.25	-1.21	-0.01	2.70
EP_FH	1.19	-1.25	-1.07	0.04	2.60
EP_Tb	1.19	-1.25	-1.12	-0.21	2.53
Vn_Va	0.87	1.41	0.91	1.37	0.06
Vn_Ce	0.87	1.41	0.63	1.34	0.26
Vn_Ft	0.87	1.41	-1.03	-0.28	2.54
Vn_Ob	0.87	1.41	-0.99	-0.24	2.48
Vn_Ct	0.87	1.41	-1.03	-0.12	2.44
Vn_Tp	0.87	1.41	-1.21	-0.01	2.52
Vn_FH	0.87	1.41	-1.07	0.04	2.38
Vn_Tb	0.87	1.41	-1.12	-0.21	2.57
Va_Ce	0.91	1.37	0.63	1.34	0.29
Va_Ft	0.91	1.37	-1.03	-0.28	2.55
Va_Ob	0.91	1.37	-0.99	-0.24	2.49

Va_Ct	0.91	1.37	-1.03	-0.12	2.45
Va_Tp	0.91	1.37	-1.21	-0.01	2.53
Va_FH	0.91	1.37	-1.07	0.04	2.39
Va_Tb	0.91	1.37	-1.12	-0.21	2.58
Ce_Ft	0.63	1.34	-1.03	-0.28	2.32
Ce_Ob	0.63	1.34	-0.99	-0.24	2.26
Ce_Ct	0.63	1.34	-1.03	-0.12	2.21
Ce_Tp	0.63	1.34	-1.21	-0.01	2.28
Ce_FH	0.63	1.34	-1.07	0.04	2.14
Ce_Tb	0.63	1.34	-1.12	-0.21	2.34
Ft_Ob	-1.03	-0.28	-0.99	-0.24	0.06
Ft_Ct	-1.03	-0.28	-1.03	-0.12	0.16
Ft_Tp	-1.03	-0.28	-1.21	-0.01	0.32
Ft_FH	-1.03	-0.28	-1.07	0.04	0.32
Ft_Tb	-1.03	-0.28	-1.12	-0.21	0.11
Ob_Ct	-0.99	-0.24	-1.03	-0.12	0.12
Ob_Tp	-0.99	-0.24	-1.21	-0.01	0.31
Ob_FH	-0.99	-0.24	-1.07	0.04	0.29
Ob_Tb	-0.99	-0.24	-1.12	-0.21	0.13
Ct_Tp	-1.03	-0.12	-1.21	-0.01	0.21
Ct_FH	-1.03	-0.12	-1.07	0.04	0.16
Ct_Tb	-1.03	-0.12	-1.12	-0.21	0.13
Tp_FH	-1.21	-0.01	-1.07	0.04	0.15
Tp_Tb	-1.21	-0.01	-1.12	-0.21	0.22
FH_Tb	-1.07	0.04	-1.12	-0.21	0.26

Note. The abbreviations P, H, EP, Vn, Va, Ce, Ft, Ob, Ct, Tp, FH, and Tb represent piano, harpsichord, electric piano, violin, viola, cello, flute, oboe, clarinet, trumpet, french horn, and trombone, respectively. *x* and *y* represent values in Dimensions 1 and 2 of the MDS map in Figure 1, respectively.

Appendix C

Planar Coordinates of Articulation Formats and

Euclidean Distances Between Pairs of Articulation Formats

Articulation format pair	Planar coordinates of Articulation format 1		Planar coordinates of Articulation format 2		Euclidean distance
	x_1	y_1	x_2	y_2	
l_s	1.10	1.10	-1.42	-0.93	3.24
l_a	1.10	1.10	-0.51	1.37	1.64
l_b	1.10	1.10	1.21	-0.29	1.40
l_c	1.10	1.10	1.19	-0.30	1.41
l_d	1.10	1.10	-1.16	0.22	2.43
l_e	1.10	1.10	0.86	-1.32	2.44
l_f	1.10	1.10	-1.27	0.16	2.55
s_a	-1.42	-0.93	-0.51	1.37	2.47
s_b	-1.42	-0.93	1.21	-0.29	2.70
s_c	-1.42	-0.93	1.19	-0.30	2.69
s_d	-1.42	-0.93	-1.16	0.22	1.17
s_e	-1.42	-0.93	0.86	-1.32	2.31
s_f	-1.42	-0.93	-1.27	0.16	1.09
a_b	-0.51	1.37	1.21	-0.29	2.39
a_c	-0.51	1.37	1.19	-0.30	2.39
a_d	-0.51	1.37	-1.16	0.22	1.32

a_e	-0.51	1.37	0.86	-1.32	3.02
a_f	-0.51	1.37	-1.27	0.16	1.42
b_c	1.21	-0.29	1.19	-0.30	0.01
b_d	1.21	-0.29	-1.16	0.22	2.42
b_e	1.21	-0.29	0.86	-1.32	1.09
b_f	1.21	-0.29	-1.27	0.16	2.51
c_d	1.19	-0.30	-1.16	0.22	2.41
c_e	1.19	-0.30	0.86	-1.32	1.07
c_f	1.19	-0.30	-1.27	0.16	2.50
d_e	-1.16	0.22	0.86	-1.32	2.54
d_f	-1.16	0.22	-1.27	0.16	0.12
e_f	0.86	-1.32	-1.27	0.16	2.59

Note. The abbreviations l, s, a, b, c, d, e, and f represent eight different articulation formats. x and y represent values in Dimensions 1 and 2 of the MDS map in Figure 5, respectively.