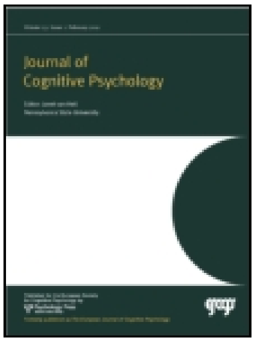

Journal of Cognitive Psychology

ISSN: 2044-5911 (Print) 2044-592X (Online) Journal homepage: http://www.tandfonline.com/loi/pecp21

Variability and recognition memory: Are there analogous indexical effects in music and speech?

Stephen Wee Hun Lim & Winston D. Goh

To cite this article: Stephen Wee Hun Lim & Winston D. Goh (2012) Variability and recognition memory: Are there analogous indexical effects in music and speech?, Journal of Cognitive Psychology, 24:5, 602-616, DOI: 10.1080/20445911.2012.674029

To link to this article: https://doi.org/10.1080/20445911.2012.674029

JOURNAL OF COGNITIVE PSYCHOLOGY, 2012, 24 (5), 602–616

Variability and recognition memory: Are there analogous indexical effects in music and speech?

Stephen Wee Hun Lim and Winston D. Goh

Department of Psychology, National University of Singapore, Singapore

Indexical effects refer to the influence of surface variability of the to-be-remembered items, such as different voices speaking the same words or different timbres (musical instruments) playing the same melodies, on recognition memory performance. The nature of timbre effects in melody recognition was investigated in two experiments. Experiment 1 showed that melodies that remained in the same timbre from study to test were discriminated better than melodies presented in a previously studied but different, or unstudied timbre at test. Timbre effects are attributed solely to instance-specific matching, rather than timbre-specific familiarity. In Experiment 2, when a previously unstudied timbre was similar to the original timbre and it played the melodies at test, performance was comparable to the condition when the exact same timbre was repeated at test. The use of a similar timbre at test enabled the listener to discriminate old from new melodies reliably. Overall, our data suggest that timbre-specific information is encoded and stored in long-term memory. Analogous indexical effects arising from timbre (nonmusical) and voice (nonlexical) attributes in music and speech processing respectively are implied and discussed.

Keywords: Melody recognition memory; Spoken word recognition; Talker-variability effects; Timbre-repetition effects.

By sheer appearance, music and language are grossly different. Beethoven's symphony cannot be mistaken for a political 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. According to McMullen and Saffran (2004), there are basic features, such as phonemes, that can be combined to form words, phrases, and sentences in spoken language. In music, such basic features may be musical notes that can be combined to create melodies. In both domains, the basic features are organised by regularities, such as prosodic structure in language and rhythmic structure in music. Many studies have argued for a commonality between music and language processing (see Patel, 2003, for a review). The present study investigates one specific aspect of this potential commonality—the extent to which indexical effects arising from voice (nonlexical) and timbre (nonmusical) attributes in spoken word and melody recognition respectively are analogous. Indexical effects refer to the influence of surface variability characteristics of the stimuli, such as variations in talkers or musical instruments presenting the words or melodies between study and test phases, on the ability to recognise the to-be-remembered words or melodies.

SPEECH PERCEPTION AND RESEARCH ON TALKER VARIABILITY

Traditionally, perception of the linguistic content of speech—the words, phrases, and sentences—has

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

http://www.psypress.com/ecp http://dx.doi.org/10.1080/20445911.2012.674029

been studied separately from perception of voice (talker) identity (Pisoni, 1997). Variation in the acoustic realisation of linguistic components due to individual talker differences is considered a source of noise that obscures the underlying abstract symbols. The proposed solution to this ''perceptual problem'' is a normalisation process where voice-specific acoustic-phonetic properties are evaluated in relation to prototypical representations of the linguistic constituents. The end product of perception consists of abstract, context-free linguistic units independent of the identification, recognition, and storage of nonlinguistic properties of speech, such as the talker's voice.

A contrasting approach proposes that representations include surface characteristics of speech (Goldinger, 1998; Pisoni, 1997). Talker information is not discarded through normalisation; instead, talker variation forms part of a rich and elaborate representation of speech. Indexical information such as voice attributes are retained in episodic memory along with lexical information, and both kinds of content contribute to speech perception. This view is compatible with exemplar-based models of long-term memory (LTM). When we encounter an item, a representation of it is stored in memory. This representation contains multiple aspects of the memory episode such as item, lexical, associative, and contextual information (Gillund & Shiffrin, 1984; Hintzman, 1988; Raaijmakers & Shiffrin, 1981; Shiffrin & Steyvers, 1997).

Evidence for this episodic approach is found in several learning and memory studies. Intelligibility of novel words and sentences produced by familiar talkers improved compared to new talkers (Nygaard & Pisoni, 1998; Nygaard, Sommers, & Pisoni, 1994), suggesting that acquisition of indexical knowledge increases perceptual sensitivity to linguistic information. Recognition for words or sentences repeated in the same voice surpassed recognition for those presented in a different voice (Goldinger, 1996; Pilotti, Bergman, Gallo, Sommers, & Roediger, 2000; Sheffert, 1998). Although some researchers did not observe this difference (e.g., Church & Schacter, 1994; Luce & Lyons, 1998), the general trend favours the position that indexical properties are retained in LTM during perception (see Goh, 2005, for a review).

MUSIC PERCEPTION AND RESEARCH ON SURFACE FEATURE VARIABILITY

An analogous dichotomy can be found in music, namely abstract structure and surface characteristics (Trainor, Wu, & Tsang, 2004). Abstract structure consists of relative pitches and durations of tones in music, which refer to pitch relations between tones regardless of absolute pitch level, and ratios between durations, regardless of absolute length, respectively. A normalisation process must occur to capture this structural information. During this extraction, information about surface characteristics, such as absolute pitch, tempo, timbre, and prosodic rendering, is discarded.

Both abstract structure and surface characteristics are useful for music interpretation. Representation of the abstract structure enables recognition across different performances, and of musical variations of a motif within a musical composition (Large, Palmer, & Pollack, 1995). The song *Happy Birthday* is recognisable when presented at various pitches and tempos, or even when embellished and harmonised on various musical instruments. The song retains its identity if the relations (intervals) between tones are preserved, regardless of whether it is sung with a high or low voice, or played on a piano or flute (i.e., timbre is irrelevant). 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.

Whereas Raffman (1993) suggested that only the abstract structure is encoded into LTM, others argue that surface features are encoded as well (Radvansky, Fleming, & Simmons, 1995; Wolpert, 1990). Similar to speech recognition, melodies repeated in the same timbre were recognised more reliably than those presented in a different timbre (e.g., Peretz, Gaudreau, & Bonnel, 1998). Timbre attributes may be assumed to be computed during the perceptual analysis of the musical input. Some studies suggest that listeners' ability to perceive differences in timbre is remarkable. For instance, sequences composed of short tones played by a different timbre can be differentiated from comparison sequences containing the same tones played with other timbres (Warren, Gardner, Brubaker, & Bashford, 1991). Specific musical instruments can be identified in forced-choice tasks that involve tones presented for similarly short durations (Robinson & Patterson, 1995a).

INSTANCE-SPECIFIC MATCHING AND VOICE-SPECIFIC FAMILIARITY EFFECTS IN SPEECH PROCESSING

The present study is motivated by Goh's (2005) attempt to tease apart two possible processes that could account for the indexical word recognition effects. Goh (2005) examined the relative contribution of instance-specific matching and voice-specific familiarity processes to spoken word recognition. Previous studies adopted the standard procedure where words presented by multiple talkers were studied, and the task was to determine if test items were old, ignoring talker identity. Old items were either repeated by the same talker, or presented by a different studied talker, while new items were distributed among all studied talkers. An alternative procedure involved comparing same-voice repetitions and *new*-voice presentations using unstudied talkers (Lively, 1994). Any discrepancy between same and different or same and new conditions was evidence implicating encoding of indexical information.

In same versus new comparisons, the former is an exact match with the memory trace for the old item, making it more discriminable from competing traces. In contrast, the latter matches the trace only in terms of the word's lexical attributes rather than its indexical attributes. An advantage of same over new presentations can be obtained due to the specificity of the match between the episodic memory trace and the probe, or *instance-specific matching*, a term coined by Goh (2005). This is based on the encoding specificity principle (Tulving & Thompson, 1973), which posits that the effectiveness of a retrieval cue depends on its degree of relatedness to the initial encoding of an item.

However, studied voices are also more familiar compared to new voices. Greater overall familiarity with old voices, rather than a match in voice attributes between the traces and probe, confers an advantage to old over new voices, which Goh (2005) termed as the *voice-specific familiarity* advantage. Global memory frameworks (e.g., Gillund & Shiffrin, 1984; Hintzman, 1988) propose that all traces are probed concurrently, and the relative activation strengths of each trace depend on the degree of matching attributes with the probe. A previously studied (i.e., familiar) voice may induce heightened activation levels of all traces that contain attributes of the studied items; an unstudied (i.e., unfamiliar) voice will not constitute an effective cue because no memory trace will contain those indexical attributes. Hence, voice repetition effects can also be attributed to greater familiarity with the indexical properties of the studied items, rather than the extent of match between memory traces and probe per se.

Thus, both instance-specific matching and familiarity processes can account for the difference in same versus new comparisons. For same versus different comparisons, all voices used at test were previously studied and therefore equally familiar. Any repetition effect obtained would be attributable to instance-specific matching processes per se, rather than familiarity processes.

It is apparent that both of these designs are inadequate to elucidate the role of familiarity processes per se for variability effects in recognition. Goh (2005) assessed the unique contributions of instance-specific and voice-specific familiarity processes by examining discrimination in same-, different-, and new-voice conditions simultaneously in a single experiment and demonstrated that performance improved only when the same voice was repeated at test, implicating instance-specific matching processes for indexical and lexical information in discrimination.

ARE THERE ANALOGOUS INDEXICAL EFFECTS IN MUSIC PROCESSING?

The question of interest is whether analogous effects would be observed in melody discrimination. As discussed earlier, an analogue to talker variability as a surface characteristic of the stimuli would be timbre variability. The logic outlined in Goh's (2005) word recognition study can also apply to music recognition. A fundamental idea is that timbre properties, when studied, may become stored with the memory trace for a studied melody. For instance, in Experiment 1 of Crowder's (1989) study, subjects judged whether two consecutively presented pitches that could also vary in timbre were of the same, or of a different, diatonic pitch. The data revealed that correct same-pitch responses were significantly faster when the timbres of the two notes were the same than when they were different, implicating that timbre information was stored with the memory trace of the first pitch, at least until the second pitch arrived. In the present study, 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 by comparing same-timbre with different-timbre conditions (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 (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. This is because repeating an exact same token will obviously match a specific instance in memory, but will also share indexical attributes with old melodies that were played by that timbre.

It is important to examine the contributions of both instance-specific matching and timbre-specific familiarity processes because it will elucidate the underlying nature of timbre effects in melody discrimination. Specifically, how tightly integrated must surface (timbre) and abstract structural attributes in the melody's episodic trace be to serve as an effective retrieval cue? Is the timbre-repetition effect dependent on the coherence of both types of attributes of the melody as a single unit? Or does the exposure to, and storage of, other instances containing the same timbre attributes (but not the same structural attributes since they are different melodies) also contribute to the timbre-repetition advantage? We sought to answer these questions in Experiment 1.

EXPERIMENT 1

No one to date has investigated the timbre-repetition effect with same-, different-, and new-timbre conditions in a single experiment. Following Goh (2005), the highlight is that the present design allowed a critical comparison between the different- and new-timbre conditions. Suppose two timbres—piano and flute—were heard during the study phase. A melody in piano at study could reappear either in flute (different-timbre condition) or violin (new-timbre condition) at test. In either condition, the test melody does not constitute an instance-specific match with its studied trace because its test timbre (i.e., flute or violin) differs from its original timbre (i.e., piano). Notwithstanding, when the melody reappeared in flute, it should appear more familiar than if it reappeared in violin, because flute is a timbre that was studied while violin was not. Any difference in performance between the different-timbre and new-timbre conditions can now be solely attributable to a timbre-specific familiarity effect, due to the fact that there was no instance-specific match in either condition, and that a different but previously studied timbre is presumably more familiar than a different, previously unstudied timbre.

Method

Participants. Fifty-two psychology students with varying music experience from the National University of Singapore (NUS) participated for course credit.

Materials. The stimuli comprised 48 single-line (monophonic) melodies (see Figure 1 for samples) composed using the Finale 2009 software as .wav files. An equal number of four-bar melodies were composed in the tonality (key) of C major, C minor, G major, or G minor. The melodies began either on the tonic, mediant, or dominant, but always ended with a single long note on the tonic of the key they were written in. Each melody was written either in simple triple or simple quadruple time, lasting approximately 6 or 7.2 s respectively.

Figure 2 depicts the multidimensional timbre map showing the similarity relations between

Figure 1. Samples of the melodies used for the present study.

timbres. This map allowed a selection of specific timbres based on objective measures of the degree of perceived similarity among different instruments. The procedures to collect similarity ratings and to generate the map using multi-dimensional scaling (MDS) techniques (Kruskal & Wish, 1978) are described in the Appendix.

After deriving the scaling solution, object coordinates in the space were used to estimate perceptual distances between instruments, using 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 ($x_1y_1$) and ($x_2y_2$) are planar coordinates for Points 1 and 2 respectively. The further apart two instruments are positioned in space, the more perceptually distinct they are. For counterbalancing, two timbre sets containing three instruments each that were perceived as maximally distinct from each other (Set A: piano, flute, violin; Set B: harpsichord, clarinet, cello) were selected to present the melodies.

Apparatus. Computers equipped with 16-bit sound cards were used to present signals to participants via a pair of Beyerdynamic DT150 headphones at approximately 70 dB SPL. E-prime 1.2 and the PST Serial Response Box (Schneider, Eschman, & Zuccolott, 2002), with the left- and right-most buttons of the button-box labelled "No" and "Yes" respectively, were used for stimuli presentation and data collection.

Design. To ease exposition, the terms *old* and *new* refer to melodies, whereas *studied* and *unstudied* refer to instruments. The timbre-context conditions were run within participants. Old melodies had three levels of timbre context (studied-same, studied-different, unstudied), whereas new melodies had only two levels (studied, unstudied) because it is impossible to have studied-same or

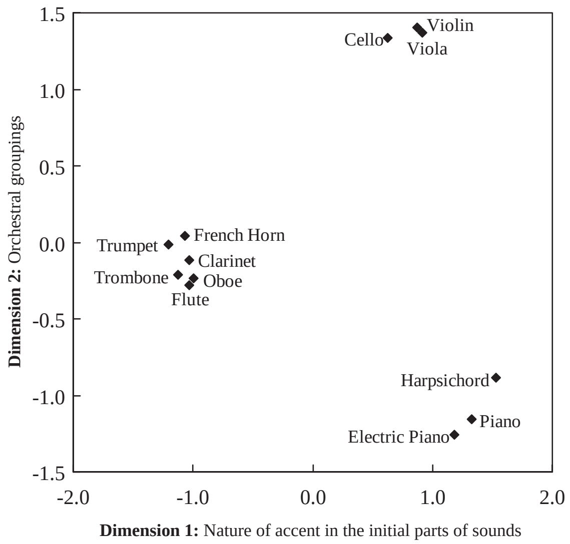

Figure 2. Two-dimensional MDS solution for the 12 instruments. Dimension 1 might be influenced by the presence or absence of accent in the initial part of the sound. Dimension 2 might relate to traditional orchestral families.

studied-different new melodies as they would not have been presented during the study phase.

The unequal number of timbre-context conditions for old and new melodies made it impossible to ensure equal number of trials in each condition while also ensuring equal exposure to studied and unstudied instruments at test. If there were equal numbers 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 old-melody conditions, studied-same and studied-different, used studied-timbres, whereas only one old-melody condition used unstudied-timbres), risking bias effects as a function of the amount of exposure to particular timbres. Hence, it was more critical to ensure that the number of exposures to each instrument at test was equal. The following controls achieved this (see Table 1 for a summary).

Two equivalent melody lists, matched in terms of meter and tonality, of 24 melodies each were created. One list was selected to be old melodies, and the other to be new melodies, counterbalanced across participants. Random selection was used in assigning melodies to instruments for all procedures described below.

At study, the old melodies were divided among two studied instruments (i.e., either piano and flute of Set A, or cello and harpsichord of Set B). Each instrument therefore presented 12 melodies; both instruments were heard an equal number of times.

At test, new melodies were divided among three instruments, where two of them were studied instruments (i.e., either piano and flute of Set A, or cello and harpsichord of Set B) and the third was an unstudied instrument (i.e., either violin or clarinet); each instrument presented eight melodies. For old melodies, one-third were switched to the unstudied instrument (i.e., either violin or clarinet), while the rest were equally divided into the studied-same and studied-different timbre-context conditions.

Procedure. Instrument set was counterbalanced across participants who were instructed to silently memorise each melody presented over headphones and were informed of a forthcoming recognition test. Each study trial comprised two consecutive presentations of the same melody, separated by 800 ms of silence. Participants

TABLE 1
Summary of the design used in Experiment 1

Memorisation stage		*Recognition stage*					
		Test melodies (old)			*Test melodies (new)*		
		Timbre context					
Study melodies		*Studied-same*	*Studied-different*	*Unstudied*	*Studied*	*Studied*	*Unstudied*
		Set A instruments					
Piano 12	Flute 12	Piano 8	Flute 8	Violin 8	Piano 8	Flute 8	Violin 8
		Set B instruments					
Cello 12	Harpsichord 12	Cello 8	Harpsichord 8	Clarinet 8	Cello 8	Harpsichord 8	Clarinet 8

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

pressed the spacebar to proceed to the next melody until all 24 melodies were presented. The presentation sequence was randomised across participants.

Following the study phase, examples of well-known melodies such as *London Bridge* were used to clarify the nature of the recognition task. Specifically, participants were told to respond "Yes" if they thought they heard *London Bridge* at study, regardless of whether the instrument playing *London Bridge* at test was the original instrument that played it. Participants were also told to respond as accurately as possible. Each randomly presented test trial comprised the question "Did you hear this melody in Part 1?" displayed visually, and a single melody played through headphones. A new trial was started 800 ms after a button response. No feedback was provided.

After the recognition test, participants' music training experience was recorded and followed by a debrief. 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 4 years of training, and pressed 2 on the keyboard if they had undergone at least 4 years of training.

Results and discussion

Discrimination was assessed by d'. Due to the unequal number of timbre-context levels for old and new melodies, three hit and two false alarm rates were obtained. Consequently, d' for the studied-same and studied-different conditions was derived from their respective hit rates but using a common false alarm rate for new melodies presented by a studied timbre; d' for the unstudied condition was calculated based on its respective hit and false alarm rates.

Average d' (*SD*s in parentheses) in the studied-same, studied-different, and unstudied timbre-context conditions were 1.22 (0.62), 1.00 (0.67), and 0.96 (0.69) respectively. Hit rates for the studied-same, studied-different, and unstudied conditions were 0.73, 0.65, and 0.63, respectively. False alarm rates for studied and unstudied timbres were both 0.31.

Timbre-context influenced discrimination, $F(2, 102) = 3.21$, $MSE = 0.31$, $p < .05$. Pairwise comparisons revealed that melodies repeated with the studied-same timbre were better discriminated than melodies presented with a studied-different timbre, $t(51) = 2.33$, $p < .05$, or an unstudied timbre, $t(51) = 2.29$, $p < .05$. No difference between studied-different and unstudied conditions was found, $t < 1$.[1]

This pattern of results suggests that discriminability increased significantly only when melodies were repeated in the same timbre, which is consistent with Goh's (2005) finding on voice-context effects in spoken word recognition. It also replicated previous work showing an advantage in melody recognition for same-timbre repetitions over different-timbre presentations (Peretz et al., 1998; Radvansky et al., 1995; Wolpert, 1990), and over new-timbre presentations (Trainor et al., 2004).

More important, discrimination performance between the different-timbre and new-timbre conditions did not differ reliably. When a melody reappeared at test in a different but previously studied timbre, it should appear more familiar than if it reappeared in a completely new timbre. As such, if discrimination performance were actually influenced by timbre-specific familiarity processes, the d' value observed in the different-timbre condition ought to be higher compared to the new-timbre condition. Discrimination performance did not differ across these two conditions, suggesting that timbre-specific familiarity processes did not prevail. In other words, the timbre-repetition effect in melody recognition is explained solely by the instance-specific matching process, rather than any timbre-specific familiarity process. Discrimination appears to be affected only when there was an exact match between the trace of the studied melody and test melody. The question we asked next was whether an *exact*

[1] The main analyses were preceded by an examination of potential effects of musical training for the sample, although there was no systematic attempt to control the number of participants with or without formal musical training as this was not a primary goal of the study. A two-way mixed design analysis of variance (ANOVA), with music training as the between-subjects factor (participants who had at least 4 years of formal training, $n = 14$, vs. those without or with less than 4 years of training, $n = 38$) and timbre context as the within-subjects factor, revealed no reliable interaction, $F < 1.60$. This shows that music training did not influence the timbre effects. A main effect of musical training was marginally significant, $F(1, 50) = 3.94$, $MSE = 0.66$, $p = .053$. Participants with at least 4 years of training ($M = 1.27$, $SD = 0.77$) tended to discriminate better than those without or with less than 4 years of training ($M = 0.98$, $SD = 0.60$). Since training did not interact with timbre context, the main findings for the timbre-context conditions in the main analyses can be generalised across all participants within the sample, and all tabulations of results are collapsed across music training.

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

Extant studies that examined timbre effects have used test stimuli that were denoted as either 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. The present experiment was designed to assess the contribution of fine perceptual details to the timbre effects observed in previous work. 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 original encoding of an item. Timbre information is initially encoded and stored in the memory traces of melodies, and later used to retrieve or recover the melodies. Because a same-timbre repetition is really an exact match with the trace for the old melody, that trace becomes more salient compared to other competing traces, enabling high discrimination. On the other hand, a melody presented by a *distinct* timbre, whether different or new, would match the trace for the melody only in terms of its structural properties, and not in terms of its timbre properties, resulting in low discrimination. Indeed, in Experiment 1, discrimination in the same-timbre condition was found to exceed performance in the other two conditions that used maximally distinct timbres.

The comparison of shared properties between the memory trace and probe implies that item similarity per se constitutes an integral part of the retrieval process. An assumption by exemplar theorists is that memory for a stimulus is essentially memory for features contained in that stimulus. The global matching approach (see Clark & Gronlund, 1996) suggests that features in a test item, when matched with the features that have earlier been stored in memory, evoke a familiarity signal. Specifically, the greater the degree of match is, the stronger the signal will be. In the present case, overlapping features between the (timbres of the) two melody instances from study to test presumably contribute to the familiarity signal evoked by resemblance to the studied melody (see Cleary, 2004). Suppose a melody was presented in violin at study, but reappears in cello at test. Cello, a different but similar timbre, presumably shares many common indexical (timbre) features with violin. These overlapping features from study to test lead to a high degree of similarity between the two melody instances which invokes a strong feeling of familiarity, such that the two instances become matched and coded as the same memory trace (i.e., instance-specific matching occurs). On the other hand, if the melody reappeared at test in a timbre that is distinct (e.g., flute) from its original timbre, there are few overlapping features between the two timbres and, thus, the melody instances. As such, the familiarity signal is presumably weaker, and melody discrimination is hampered. The prediction is that 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 demonstrating 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 old. If the surface (i.e., timbre) attributes of melodies are analogous to the voice attributes of spoken words, we expect discrimination to improve even when old melodies are not repeated with the exact same timbre, but with a similar timbre in its place instead.

Method

Participants. Forty-two psychology students from NUS who had not taken part in Experiment 1 participated for course credit.

Materials, apparatus, design, and procedure. The materials and procedures were the same as those of Experiment 1, with a slight modification in design.

Six different combinations of instruments were derived for melody presentation as listed in Table 2. The selection of instruments and their assignment to the timbre conditions was based on the scaling solution in Figure 1, in which the instruments appear in three perceptual groups. Any two instruments within each group are taken as similar since they appear close in proximity to each other on the timbre map. When violin, for instance, has been assigned to the same-timbre condition, an instrument that is perceptually similar to (i.e., within the same perceptual group as) violin, such as cello, is assigned to the similar-timbre condition; an instrument that is perceptually dissimilar from (i.e., belongs to a different perceptual group from) violin, such as flute, is assigned to the distinct-timbre condition. In addition, care was especially taken to avoid assigning certain instruments to the similar-timbre condition, in order to preserve the integrity of the manipulation of timbre similarity. For instance, when violin has been assigned to the same-timbre condition, viola, which almost superimposed on violin on the timbre map, would not be designated as violin's similar-timbre counterpart, because the danger is that viola could be perceived as *virtually identical*, rather than *similar*, to violin due to its extreme similarity to violin. Set combination was counterbalanced across participants.

As in Experiment 1, there were two equivalent lists of 24 melodies each that were designated as the old and new melodies, respectively. At study, all old melodies were presented by a single instrument. At test, old and new melodies were divided so that the same instrument, a similar instrument, and a distinct instrument each presented an equal number of melodies (see Table 3 for a summary).

TABLE 2
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

Results and discussion

Discrimination was assessed by d'. Unlike Experiment 1, there were three levels of timbre context (same, similar, distinct) for both old and new melodies. As such, three hit and false alarm rates were obtained, and d' for each condition was calculated based on individual hit and false alarm rates.

Average d' (*SD*s in parentheses) in the same, similar, and distinct timbre-context conditions were 1.17 (0.75), 1.16 (0.65), and 0.76 (0.68), respectively. For the same, similar, and distinct conditions, hit rates were 0.75, 0.70, and 0.62, and false alarm rates were 0.36, 0.32, and 0.36, respectively.

Timbre-context influenced discrimination, $F(2, 80) = 7.68$, $MSE = 0.40$, $p < .01$. Discrimination of melodies repeated with the same-timbre was better than those with a distinct timbre, $t(41) = 2.95$, $p < .01$; discrimination was also better when melodies were presented with a similar timbre compared to a distinct timbre, $t(41) = 2.75$, $p < .01$. Discriminability did not differ between the same and similar timbre-context conditions, $t < 1$. This pattern of results indicates that when melodies were repeated either in the same or similar-timbre at test, discrimination improved.[2]

An advantage for same-timbre repetitions and similar-timbre presentations over distinct-timbre presentations emerged. The results were consistent with Goldinger's (1996) finding on voice similarity effects in spoken word recognition. As predicted, discrimination in both the same-timbre and similar-timbre conditions were comparable; performance improved even when the old melodies were not repeated with the exact same timbre at test.

These data corroborate Experiment 1's data to emphasise that timbre similarity primarily constitutes an integrated part of the matching and retrieval processes involved in melody recognition. More important, the present result extended

[2] No reliable interaction between music training (participants who had at least 4 years of formal music training, $n = 10$, vs. those without or with less than 4 years of music training, $n = 32$) and timbre context was observed, $F < 2.43$. Again, this shows that music training did not influence timbre effects. A main effect of musical training was significant, $F(1, 40) = 8.17$, $MSE = 0.36$, $p < .01$. Participants with at least 4 years of training ($M = 1.36$, $SD = 0.66$) discriminated better than those without or with less than 4 years of training ($M = 0.93$, $SD = 0.66$). As in Experiment 1, the main findings for the timbre-context conditions can be generalised across all participants.

TABLE 3
Summary of the design used in Experiment 2

Memorisation stage	*Recognition stage*					
	Test melodies (old)			*Test melodies (new)*		
	Timbre context					
Study melodies	*Same*	*Similar*	*Distinct*	*Same*	*Similar*	*Distinct*
	E.g., Set combination 1 instruments					
Piano	Piano	Harpsichord	Violin	Piano	Harpsichord	Violin
24	8	8	8	8	8	8
	E.g., Set combination 5 instruments					
Flute	Flute	Clarinet	Harpsichord	Flute	Clarinet	Harpsichord
24	8	8	8	8	8	8

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

Experiment 1's findings 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.

GENERAL DISCUSSION

Analogous indexical effects in music and speech processing

This study demonstrated that music and speech converge in recognition memory on the basis of their analogous indexical effects in two specific aspects. First, some form of instance-specific matching must occur for effective melody discrimination to take place. In Experiment 1, melodies that remained in the same timbre from study to test were discriminated better than melodies presented in a studied but different, or unstudied timbre. In addition, melodies presented in a different timbre were not discriminated better than melodies presented in a new timbre at test. The present data are consistent with an analogous study using spoken words to examine voice-context effects in recognition memory (Goh, 2005). Surface (e.g., timbre) and 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 match exactly, discrimination is enhanced, but performance is hampered so long as there is a clear mismatch of the melody's surface attributes between study and test.

Second, the *exact same* timbre was not necessary to induce an instance-specific match in melody discrimination; a different, but similar, timbre sufficed to induce matching. In Experiment 2, melodies that remained in the same timbre from study to test were discriminated better than melodies presented in a distinct timbre. But, when a timbre that was different from, but similar to, the original timbre played the melodies at test, discrimination was comparable to the same timbre playing them. This result is generally consistent with the pattern in an analogous study using spoken words which found that perceptual similarity of study and test voices modulated the magnitude of different-voice repetition effects (Goldinger, 1996). When the test voice was similar to the study voice, it was probable that the listener would classify the word as an old word. It appears that the surface (timbre) attributes of melodies are analogous to the voice (nonlinguistic) attributes of spoken words.

Melody recognition as a form of familiarity-based recognition

Our data 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 recognise a previously heard melody. The fact that song recognition can occur reliably when surface features, in their case the tempo of the music, 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 our case, similar timbre 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 number of studies had obtained empirical support for the idea that melody recognition can reflect familiarity-based recognition (see Kostic & Cleary, 2009), and our findings are compatible with this idea and extend it by suggesting the nature of the familiarity processes involved.

It should be noted that merely being familiar *with a studied timbre* (Experiment 1) is futile in enhancing melody recognition at test, because merely hearing (and becoming familiarised 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 we 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 (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 when violin was not previously studied (i.e., an unfamiliar timbre), 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.

In line with dual-process theories of recognition, it should be acknowledged that timbre similarity may also have facilitated explicit recollection of the specific episodic trace of the studied melodies. Data on participants' subjective ratings of the basis of their responses, such as *remember/know* judgements (Mandler, 1980), in a future study may help to elucidate this.

Relation to memory for general music

The present findings are compatible with previous work that examined long-term memory in a wider context of music (e.g., songs). Pitch and rhythmic structure primarily specify a song's identity, and have become a focus of psychological research on music (see Krumhansl, 1990). A song's identity is determined from abstracted information about relations between tones, rather than from the tones' absolute characteristics per se. For instance, *Happy Birthday* retains its identity if the relations between tones are maintained, regardless of the pitch level of its initial tone. Regardless of whether a song is sung using a high or low voice, it retains its identity so long as the intervallic structure is preserved. In the same vein, so long as the differences in duration between consecutive tones stay at the same ratios, the song retains its identity regardless of whether it is sung fast or slow. Timbre is also irrelevant to a song's identity. We recognise *Happy Birthday* no matter whether it is played on a piano or a flute.

Yet, absolute surface attributes, such as absolute pitch, tempo, and timbre, serve an important role in memory for popular recordings, even though they are irrelevant to a song's identity. Respondents, when asked to sing short passages from well-known recordings, did so at a pitch (Levitin, 1994) and tempo (Levitin & Cook, 1996) that closely approximate the pitch and tempo of the original recordings. More recent work by Schellenberg and Trehub (2003) suggests that good pitch memory for popular selections is common even among adults with no formal musical training. Respondents were presented with excerpts of familiar soundtracks played either at the original, or at an adjusted, pitch level, and reliably identified the original pitch levels. These data suggest that listeners actually retain information about pitch level over extended periods of time. Schellenberg, Iverson, and McKinnon (1999) extended the findings on pitch and tempo by showing that memory representations for popular recordings contain absolute information about timbre. Specifically, when recordings were presented only very briefly, recognition of the recordings was in fact observed to be a function of timbre, rather than of pitch or tempo. The data are consistent with findings that involve music and speech, which showed that respondents rely more heavily on timbre—a specific musical instrument or vowel—than pitch in identifying stimuli that are presently only very briefly (Robinson & Patterson, 1995a, 1995b).

Towards an episodic theory of music perception

Overall, the present findings have implications for the role of abstract structure and surface characteristics in music processing and interpretation. Specifically, they lend support to the view that the surface feature of a melody, such as timbre, does get encoded, along with structural information, into LTM (Peretz et al., 1998; Radvansky et al., 1995). Because of similar indexical effects that appear to subsume music and speech, it bears to explore here a theoretical framework that might potentially accommodate both speech and music perception.

The framework that we propose is used in exemplar approaches in speech perception (e.g., Goldinger, 1998; Pisoni, 1997) that assume representations of spoken words in memory to contain both lexical and indexical information, such that talker information is encoded and used in lexical access and retrieval. These approaches were developed primarily to solve a fundamental problem in speech—how perceptual constancy is achieved despite the lack of acoustic–phonetic invariance in the speech signal. A basic premise of these exemplar theories is that episodes are not decomposed into abstract and surface features. By holistically storing all instances of an encountered word in memory, Goldinger's episodic lexicon model obviates the need for normalisation, by proposing that the perception of the abstract word essentially emerges from the statistical aggregation of all these multiple episodes without any explicit representation of the abstract form (see McClelland & Rumelhart, 1985). The abstract form can be thought of as the average of the multiple instances. Hence, the inherent variability found in multiple instances actually contributes to the perceptual process.

The problems of perceptual constancy and invariance are also clearly relevant in, and can be generalised to, music perception. We can recognise different musical pieces as instances of the same song or tune regardless of the tempo and key in which they are played, the instrument used to play them, or the singer. The basic principles of the exemplar approaches proposed to solve analogous problems in speech perception—holistic encoding of instances without decomposition, no explicit representation of the abstract form, and statistical composition of multiple instances—may prove fruitful in the development of a similar episodic theory of music perception.

Exemplar theories of speech perception are not without criticism. Although the design of the present study cannot shed light on these criticisms, it is worthwhile to highlight some of the latter for future tests of potential episodic approaches to music perception. Magnuson and Nusbaum (2007) argued that storing episodes as unanalysed wholes begs the question of what the basic units are, given that the extant episodic models assume segmentation into units such as words. The passive exemplar-matching mechanism inherent in the statistical compositing process also begs the question of how categories would arise. That is, how would the system know that two acoustically different instances are in the same category? Magnuson and Nusbaum argue that, rather than adopting a completely nonanalytic approach, some form of active analysis must occur during the perceptual process. One motivation for this was their demonstration of talker variability processing costs when participants expected two different talkers, compared to when participants expected only one talker, even though the speech stimuli were the same. The implication is that there was active, attentional processing based on expectancies.

Segmentation into different units such as phonemes, syllables, or words in speech, and the related problem of category formation, also has analogues in music, such as musical notes, motifs, and melodies. We agree that allowing the idea of segmentation appears to preclude the notion of *holistic* encoding of instances, but it should be noted that this does not undermine the fundamental proposal that surface and abstract features are tightly integrated as the results of the present study suggest. A future model of music perception may require the form of active control mechanisms proposed by Magnuson and Nusbaum (2007), where simultaneous constraints from multiple levels and sources of information may be used to interpret the acoustic signal. However, it is plausible that each level of hierarchical structural units, such as notes, motifs, and melodies, may still be encoded with all the surface characteristics found in the multiple instances that are relevant to the units at those levels. Specifically, it is possible that there are musical surface characteristics encoded at each of these levels that may be more salient at a particular level (e.g., in a melody context, tempo might be more salient; absolute pitch, on the other hand, might be more relevant in an isolated note context).

Future directions

Whereas some studies found that listeners have remarkable abilities to perceive differences across timbres (e.g., Warren et al., 1991; see earlier), Experiment 2 of our study seems to offer some contrasting evidence, where listeners in fact demonstrated comparable discrimination performances across the same-timbre and similar-timbre conditions. The implication is that our listeners did not distinguish a (highly) similar-sounding timbre (e.g., cello) from the actual timbre (violin). Future work could investigate the specific boundaries beyond which two timbres become just (in)discriminable from each other. In addition, of particular interest would be the effects of specialised musical training. A possibility is that although a pianist might not discriminate (nor be influenced by a switch) between violin and cello timbres, a (highly trained) violinist ought to recognise even very subtle differences to differentiate between these two timbres, since one of the test timbres belonged to his/her own instrument and he/she would presumably have developed a particular perception of the violin timbre.

CONCLUSION

This study extends previous work that examined the effects of timbre-specific information on melody recognition in several novel directions. The comparison of same-, different-, and new-timbre conditions simultaneously in the first experiment offered new insights into the nature of timbre effects observed in the extant literature—instance-specific matching, rather than timbre-specific familiarity, processes govern these effects. The comparison of same-, similar-, and distinct-timbre conditions simultaneously in the second experiment elucidated the contribution of timbre similarity to these effects, and extended the finding from Experiment 1 by showing that the use of an exact same timbre was in fact not the only way to effectively induce matching. Overall, our data suggest that there are analogous indexical effects in music and speech processing.

Original manuscript received May 2011
Revised manuscript received February 2012
First published online May 2012

REFERENCES

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

Church, B., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Memory for voice 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.

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

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

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 472–478.

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

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

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

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

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

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. (1990). *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.
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.
Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics*, *56*, 414–423.
Levitin, D. J., & Cook, P. R. (1996). Memory for musical tempo: Additional evidence that auditory memory is absolute. *Perception and Psychophysics*, *58*, 927–935.
Lively, S. E. (1994). *Preserving the perceptual record: Retention of talker-specific information in long-term memory*. Unpublished doctoral dissertation, Indiana University, Bloomington, IN.
Luce, P. A., & Lyons, E. A. (1998). Specificity of memory representations for spoken words. *Memory and Cognition*, *26*, 708–715.
Magnuson, J. S., & Nusbaum, H. C. (2007). Acoustic differences, listener expectations, and the perceptual accommodation of talker variability. *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 391–409.
Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review*, *87*, 252–271.
McClelland, J. L., & Rumelhart, D. E. (1985). Distributed memory and the representation of general and specific information. *Journal of Experimental Psychology: General*, *114*, 159–188.
McMullen, E., & Saffran, J. R. (2004). Music and language: A developmental comparison. *Music Perception*, *21*, 289–311.
Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception and Psychophysics*, *60*, 355–376.
Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1994). Speech perception as a talker-contingent process. *Psychological Science*, *5*, 42–46.
Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*, 674–681.
Peretz, I., Gaudreau, D., & Bonnel, A. (1998). Exposure effects on music preference and recognition. *Memory and 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 and Review*, *7*, 347–353.
Pisoni, D. B. (1997). Some thoughts on ''normalization'' in speech perception. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego, CA: Academic Press.
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.
Robinson, K., & Patterson, R. D. (1995a). The duration required to identify an instrument, the octave, or the pitch chroma of a musical note. *Music Perception*, *13*, 1–15.
Robinson, K., & Patterson, R. D. (1995b). The stimulus duration required to identify vowels, their octave, and their pitch chroma. *Journal of the Acoustical Society of America*, *98*, 1858–1865.
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.
Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin and Review*, *6*, 641–646.
Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*, 262–266.
Schneider, W., Eschman, A., & Zuccolott, A. (2002). *E-prime user's guide*. Pittsburgh, PA: Psychology Software Tool Inc.
Sheffert, S. M. (1998). Contributions of surface and conceptual information to recognition memory. *Perception and Psychophysics*, *60*, 1141–1152.
Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM—retrieving effectively from memory. *Psychonomic Bulletin and Review*, *4*, 145–166.
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.
Warren, R. M., Gardner, D. A., Brubaker, B. S., & Bashford, J. A. (1991). Melodic and nonmelodic sequences of tones: Effects of duration on perception. *Music Perception*, *8*, 277–290.
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.

APPENDIX: TIMBRE SIMILARITY SCALING

Method

Participants. Twenty introductory psychology students from NUS participated for course credit.

Materials. The stimulus set comprised of monophonic C-major arpeggios for the familiarisation phase and C-major diatonic scales for the similarity rating phase, all of which were played by 12 different instruments. Each arpeggio and scale lasted approximately 5 s and 8 s, respectively. These tunes were constructed in .wav files using the Finale 2009 software.

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

Design and procedure. The session consisted of two parts. In the first part, participants listened to a random order of 12 instruments playing the same arpeggio pattern. No ratings were collected during this phase; participants were told to simply listen to the instruments in order to familiarise themselves with the timbres they would be subsequently rating for similarity. On each trial, a single arpeggio was played by a particular instrument over headphones, after which participants pressed the spacebar to proceed to the next arpeggio until all 12 timbres were presented. The timbre presentation sequence was randomised across participants.

In the second part, each trial comprised the question "How similar are the two instruments?" displayed on the monitor, followed by two different instruments playing the same scale, with an interval of 500 ms between the two instances. After participants keyed their similarity rating, a new trial began. The software controlling the experiment was written to ensure that responses made before the onset of the second instrument of each pair were not admissible. Presentation of the pairwise comparisons was randomised, and the instrument presentation order within each pair was counterbalanced across participants.

Results and discussion

MDS using the ALSCAL routine of SPSS version 16 was used to analyse the data. 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 12 timbres were scaled, solutions with one to three dimensions were obtained, and the amount of variance accounted for and Kruskal's stress values, a goodness-of-fit statistic ranging from 1.0 to 0.0 with smaller values indicating a good fit of the derived solution to the data, were examined for each solution.

Inspection of the present values indicated that although there was a large increase in goodness-of-fit between the one- (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.

Dimension 1 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 piano tends to carry a more pronounced and "sharp" accent compared with that produced by the flute or horn. This interpretation is compatible with previous reports suggesting that variation of the initial part of a sound can affect the perception of musical timbre (Berger, 1964; Clark, Robertson, & Luce, 1964; Grey & Moorer, 1977; Saldanha & Corso, 1964; Wedin & Goude, 1972). The second dimension appears to group the instruments by traditional orchestral families (i.e., woodwind, brass, strings, and keyboard), with the woodwind and brass instruments clustered together as two highly similar groups (one reason for the lack of perceptual differentiation between woodwind and brass instruments here could be that the timbre quality of brass instruments [e.g., trumpet] offered by the present software tends to be mellow and woodwind-like, which departs from the quality produced on an authentic brass instrument that typically gives a distinctively stronger sound.)

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 to be used as stimuli in the main experiments.