Empirical Musicology Review Vol. 2, No. 1, 2007

Timbre as an Elusive Component of Imagery for Music

FREYA BAILES
Sonic Communications Research Group
University of Canberra, Australia

ABSTRACT: Evidence of the ability to imagine timbre is either anecdotal, or applies to isolated instrument tones rather than timbre in real music. Experiments were conducted to infer the vividness of timbre in imagery for music. Music students were asked to judge whether the timbre of a sounded target note was the same or different from the original following a heard, imagined, or control musical context. A pilot experiment manipulated instrumentation, while the main experiment manipulated sound filters. The hypothesis that participants are able to internalise timbral aspects of music was supported by an ability to perform the timbre discrimination task, and by facilitated response when imaging the timbre context compared with non-imaging. However, while participants were able to mentally represent timbre, this was not always reported as being a conscious dimension of their musical image. This finding is discussed in relation to previous research suggesting that timbre may be a sound characteristic that is optionally present in imagery for music.

Submitted 2006 October 3; accepted 2006 October 9.

KEYWORDS: *musical imagery, timbre, perception*

INTRODUCTION

INTUITIVELY, it is possible to bring to mind (auralize) the timbral, and perhaps vocal, qualities of a particular piece or fragment of music. Composers and orchestrators claim to be able to auralize different timbres (see Buckley, 1912; Mountain, 2001; Saxton, 1998; Stravinsky & Craft, 1979): while some of their knowledge is declarative (for instance, knowing in the abstract that particular sounds blend well), the remainder is presumably based on inner hearing or 'musical imagery'. Yet evidence that musicians, or indeed non-musicians, are able to consciously 'hear' timbre when they imagine music is largely anecdotal. In the tradition of Western classical music, timbre has a cosmetic value, acting as a carrier for melody and harmony. The instrumentation of a piece of music may be altered without affecting its fundamental identity (Wolpert, 1990). That is not to say that timbre is not an important musical dimension, and one that may form part of a conscious inner hearing. For instance, Auhagen and Schoner (2001) investigated the mental representation of timbre, specifically its control by musicians through the communication and use of various verbal attributes. They found a certain agreement between the timbral properties produced by musicians in response to common metaphors used to qualify sound. However, it is not clear to what extent musical imagery, defined as a conscious mental image for music, was veridical (approximating the *qualia* of perception) for these musicians, as opposed to a subconscious schematic representation of the sounds.

Some empirical investigation of imagery for timbre has been undertaken. Pitt and Crowder (1992) found that imagery for timbre was stronger for spectral rather than time-varying properties of timbre. One limitation of this study with respect to music (see also Melara and Marks, 1990) is the focus on the timbre of isolated tones, and not on material consistent with real-life musical contexts. Similarly, recent work by Halpern and her colleagues (Halpern *et al.*, 2004) using functional MRI techniques employed single instrument tones to explore timbre perception and imagery. The authors had reasoned that when imaging individual timbres, the physical impossibility of vocalizing instrumental timbres would not implicate areas of the brain associated with vocal or subvocal rehearsal. To their surprise, they discovered that the supplementary motor area associated with subvocal rehearsal was activated during tasks that called for timbral processing. The finding is potentially crucial to a clearer understanding of timbral imagery, but cannot be generalised to situations of more complex *musical* imagery. Given the anecdotal evidence and presumed importance of timbral imagery to musicians, an exploration of the phenomenon as a musical process is overdue.

Data from an experience sampling study (Bailes, in press) suggest that music students rate timbre as a less veridical dimension than melody when describing their everyday, mostly unintentional experience of musical imagery as a 'tune on the brain'. A possible reason for this concerns certain characteristics of timbre: While the pitches and durations that form melody are structurally relative, attitudes toward the relatedness and therefore the structural potential of different timbres are mixed (McAdams, 1989). It is reasonable to suggest that structural cohesion has an impact on the facility to image timbre. Krumhansl and Iverson (1992) conducted an experiment exploring the perceptual interaction of timbre and pitch, both in single tones and musical sequences. One of their main findings was that timbre was perceived in relative terms only when pitch was held constant, and that it is consequently a more absolute than relative sound dimension. By way of contrast, other studies (for example Ehresman and Wessel, 1978) have indicated that timbres can be perceptually related to each other. Subjects showed agreement in their selection of appropriate timbres to fit the vector 'tone A is to B, as tone C is to '?''. Multidimensional scaling and timbre space models are the products of studies asking subjects to rate the dissimilarity of pairs of sounds. The ability to perform such tasks implies that timbre can be related and even transposed in a manner analogous to pitch. A geometrical space representing the perceptual proximity of different instrumental tones can be generated (Grey, 1977; Wessel, 1979). This has led theorists such as Lerdahl (1987) to postulate that timbre may be organised in a hierarchical fashion to provide musical structure, though this suggestion is contended (Cohen and Inbar, 2001).

Klangfarbenmelodie (a term coined by Schoenberg meaning 'tone colour melody') and electroacoustic musical forms work on the premise that timbre is a musical object rather than a melodic carrier, and consequently the focus of musical interest. Given the potential limitations of timbral relativity that have been raised, does it follow that a coherent mental representation of such music is difficult to achieve? Intons-Peterson (1980) suggests that certain musical properties including timbre may be optionally important to auditory perception and imagery. An understanding of the veridicality of imagery for the timbres of familiar music is necessary to resolve these issues. However, as with any dimension of musical imagery research, there are a number of methodological obstacles to be surmounted, with no well-established experimental paradigm.

A question that has only just begun to be addressed in the field of psychology is the extent to which musical imagery comprises properties distinct from the experience of perceived music. Experiments by the author to test the relationship of musical imagery to perception have been reported (Bailes, 2002). These novel experiments were modelled on work in visual imagery in which imagery has been conceptualised in terms of a level of equivalence with perception (Denis, 1989; Finke, 1989). The principle guiding the experiments is that the discovery of *quantitatively* different behaviour patterns on a discrimination task in the context of perceived music as compared with imagined music would constitute evidence that the two are situated at different points along a same experiential continuum: this would suggest a functional equivalence, with mental imagery for music a weak version of its perception. However, *qualitatively* different behaviour patterns might indicate that the relationship is more complex, allowing for the possibility that quite different processes are at work during musical imagery and musical perception. In Bailes (2002), differences in the facilitatory influence of imaged and perceived musical contexts on performance in timing and pitch discrimination tasks were found, though qualitative differences were less clear and require further investigation.

In the current study, a pilot experiment adapted the method reported in Bailes (2002) to test both the perception and imagery for timbre in music by music students. An aim was to determine whether participants would be able to internalise timbre sufficiently to perform a timbre discrimination task, and to compare experiment data with participants' reported ease or difficulty of imaging timbre, though post-experiment interviews.

PILOT EXPERIMENT

The overall purpose of the current experiment was to explore participants' ability to mentally represent musical timbre, by comparing timbral discrimination in imagery and perception tasks. Participants learned an instrumentally varied musical sequence, described below. Following this learning phase, participants hearing or imaging the sequence were asked to judge whether the instrumentation of a sounded target note was the same or different from the original (see Figure 1). Given the sensory immediacy of actual hearing, the perception of music is more veridical than its image [1]. In light of this, it was hypothesised that a perceived musical context should better facilitate the discrimination of a target note's timbre compared with the more 'virtual' context of imagined music. Faster reaction times and higher success rates for the perception task compared with the imagery task should reflect this facilitation. However, such facilitation should occur only if participants represent timbres in the test sequence in relation to each other, otherwise the physical presence (perception task) or absence (imagery task) of the context preceding the target note should have little bearing on the discrimination of its timbre.

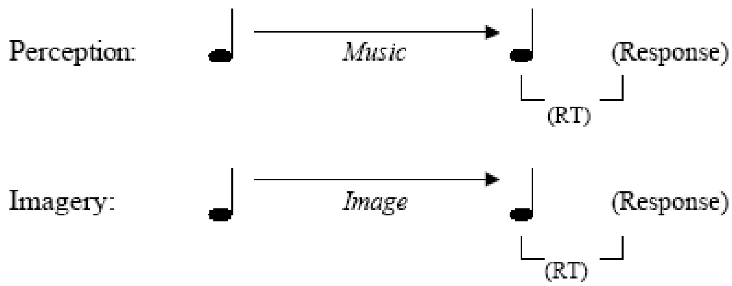

Fig. 1. A schematic representation of hearing and imagery tasks. The opening note and target note were sounded in both conditions, but in the imagery task the intervening music was 'silenced' to be filled by an auditory image of the sounds. Timbral discrimination of the target note was measured by key press, and response time (RT) from target onset to key press was measured in milliseconds.

A further prediction can be made based on the findings from previous research (e.g. Pitt & Crower, 1992; Halpern *et al.*, 2004) that imagery for timbre can be generated, and that timbre can be a perceptually important dimension of music. It was thus hypothesised that the music students participating in this experiment would be able to internalise and image timbral aspects of music.

Participants

Sixteen undergraduate music students were each paid three pounds to take part in the study [2]. Their degree course combined theoretical and practical training in music.

Material

An attempt was made to use stimuli similar to existing music composed with timbral change as a prominent feature, such as the first movement of Webern's *Symphony Op. 21* (1928), but with an appropriate degree of experimental control. In constructing the materials, the following considerations were made:

FACILITY OF SOUND SUBSITUTION

To create a same/different discrimination task, sampled orchestral instrument sounds were used, as they are simple to substitute at the appropriate target moment.

TIMBRAL CONTRAST

It was decided to contrast each target note with its immediate context, rather than to aim for transitions of overlapping timbre, since this would ensure distinct and memorable target sounds.

TIMBRAL TEXTURE

Paired instrument sounds were employed for each note. More sounds than this would have created an overly fused and homogenous texture, while using just one instrument sound per note was deemed too sparse and might have led to a ceiling effect due to the relative simplicity of learning one instrument at a time.

RATE OF SEQUENCE

A fairly slow speed was intuitively chosen on the basis that time would be required to fully register each new timbre pair.

A MIDI file was generated in *MAX (Opcode)*, consisting of the twelve pitches shown in Figure 2, with isochronous inter-onset intervals of 2 seconds. The sequence was sent to an *EMU (ES14000)* sampler, programmed to produce paired instrumental sounds for each pitch. Two independent listeners who were informed of the experimental aims confirmed that the choice of tempo and instrumentation was appropriate.

Fig. 2. Pitches used for the timbre sequence with numbers marking each note. The targets (notes 4, 7 and 10) are outlined by a box, while the dotted line marks the start of the imagery zone

Sounds for each note can be found in Table 1.

Table 1. Instrumentation for each note of the stimulus

Note 1 - Bassoon & oboe	Note 5 - Fr. horn & oboe	Note 9 - Clarinet x2 (octave)
Note 2 - Trombone & trumpet	Note 6 - Trumpet & flute	Note 10 - Fr. horn x2 (octave)
Note 3 - Clarinet & flute	Note 7 - Clarinet & oboe	Note 11 - Cello & clarinet
Note 4 - Cello & violin	Note 8 - Bassoon & Fr. horn	Note 12 - Bassoon & oboe

An initial attempt was made to juxtapose timbres according to their spatial proximity in multi-dimensional scaling (MDS) mappings of timbral relations (see for example Grey, 1977; Halpern *et al.*, 2004; and Wessel, 1979). This was to maximise the perceptual logic of the sequence through assuring some form of timbral continuity from event to event. Yet the choice of instrument sound was ultimately based on the perceptual criteria of the investigator, as the particular samples available did not inter-relate well according to MDS models. An effort was made to maintain octave equivalence in the transition from one pair of instruments to the next. However, the normal instrument range of each sampled sound was such that this was not always possible. For example, the high flute of note 3 (written F4, sounding F5) did not lead to an equivalent octave for either the ‘cello or violin of note 4.

Once created, this sequence was recorded onto audiocassette, so that participants could use this to familiarise themselves with the experiment material. A score of the music was distributed at the same time, with each note numbered (see Figure 2). Participants were instructed to consult this score while listening, to ensure that the music was thoroughly attended to.

Notes 4, 7 and 10 were selected as target notes for timbral discrimination because of their distance into the sequence and distance from each other in time. Moreover, they represented distinct pitches (at least two semitones distant from each other) to ensure minimal confusion between the targets based on pitch similarity. Experiment trials presented target notes either with the original timbre, or in alternative form with one note of the paired instruments being replaced by a different instrumental sound. Table 2 shows 'same' and 'different' instrumentation for the pre-selected target notes.

Table 2. Target notes, their position within the sequence, their original instrumentation and their substitutions

Target	Note number	Same instrumentation	Different instrumentation
1	4	Cello & violin	Cello & trumpet
2	7	Clarinet & oboe	Cello & oboe
3	10	French horn x 2 (octave)	Flute & French horn

Counterpart sequences for the imagery task borrowed those created for the perception task and 'silenced' all but the target note and opening note of the sequence (the start of the imagery zone is represented by a dotted line in Figure 2). For both perception and imagery tasks, the music was discontinued after the onset of the target note. A flashing light on a computer monitor provided a visual anchor to the task, blinking in time to the onset of each perceived or imaged note.

Procedure

Participants first took an instrument identification test lasting ten minutes, to establish a basic competence in instrument identification. This asked them to select from a list the instrumental sound and pairs of sounds they believed to be playing isolated tones. These tones were the same sampled sounds to be used in the main experiment, transferred from computer to audiocassette. Participants listened individually over headphones.

Following the test, cassettes of the experiment sequence were distributed to participants with instructions to pay particular attention to the timbre of the music. Internalising the experiment sequence demanded listening to the cassette a minimum of fifteen times, over a minimum of five days.

The experiment session could then take place, during which participants were asked to attend to a score of the music that was placed below the computer monitor. Instructions were given for task requirements, and the keys to press in response to target notes ('S' for same timbre, 'D' for different timbre). It was requested that subjects respond as quickly, but as accurately, as possible. The order of performing each condition was reversed for half of the listeners:

Group A - (n=8) pre-test, perception, imagery
Group B - (n=8) pre-test, imagery, perception

Testing was grouped by condition (perception/imagery) and test block. Each block comprised the six trials (three target notes x two variations) presented in random order. Each block was presented three times per condition, and interspersed with a replay of the original music, followed by a 35 seconds gap. Distributing a questionnaire that asked for basic information about musical background separated perception and imagery halves of the experiment. Following the final experimental block, each participant was briefly interviewed to find out more qualitative information about the experience.

Response time (RT) and accuracy rates were measured. RT for correct response were analysed using a multifactorial ANOVA, with order as the between-subjects factor, and three within-subjects factors of condition, target number, and target type (i.e. ‘same’ or ‘different’ instrumentation). Accuracy rates were analysed using a generalised chi-square. Post-experiment interviews were analysed thematically, and are discussed both in relation to quantitative findings, and as a source of new issues arising during the course of the study.

Results

Participants scored highly in the pre-test identification of single instrument sounds (mean score of 8 out of 9), but less well on the paired sound test (mean score of 5.5 out of 8). The difficulty of identifying the paired instruments was noted but not considered important, as discrimination between sounds rather than their identification formed the basis of the main experiment.

Before the experiment session began, participants were asked how well they felt they knew the music they had been asked to listen to at home. Responses were largely negative and uncertain: participants generally doubted the extent to which they had internalised the music. After the experiment the following question was asked: 'How true to the real piece do you feel your image of the music to have been?' Respondents were variously confident as to the clarity of their mental re-presentation. The most common qualification was that while the pitches and the general timbre of a few key notes were in place, there existed no detailed replica of the music in their minds.

The anticipated main effect of condition (imagery/perception) was significant for accuracy ($X^2[1] = 7.31$, $p < 0.005$) with 87.5% correct response in the perception task and 79.1% in the imagery task. This finding supports the hypothesis that participants perform better in a timbral discrimination task under perception rather than imagery conditions.

Group differences in accuracy emerged (see Figure 3). Group A (those who did the perception task before the imagery task) was on average 91% accurate, while Group B (those who did the imagery task before the perception task) was 75.6% accurate. The difference is significant ($X^2[1] = 24.38$, $p < 0.001$). A closer examination of the factor 'condition' reveals that the difference in accuracy is attributable to the performance of Group B ($X^2[1] = 17.28$, $p < 0.0005$). Because their performance was significantly more accurate in the perception condition (86.1%) than the imagery condition (63.9%), which is to say the second of their two tasks, it seems likely that practice effects contributed to an improved performance at this later stage of the experiment.

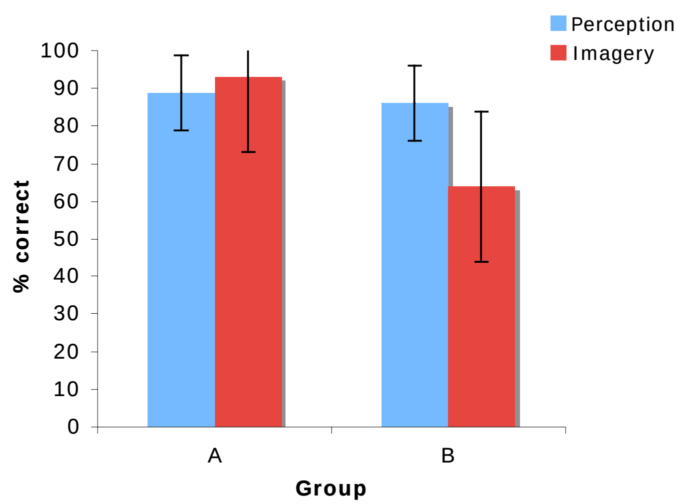

Fig. 3. Percentage of correct response for Group A (perception then imagery task order) and Group B (imagery then perception task order).

Accuracy data for Group A did not indicate that performance on the perception task (88.9%) was any better than on the imagery task (93.1%). There may be a number of reasons for this outcome. First, it may be that as discussed above, perception for individual timbres was not sufficiently relative for a perceived context to facilitate the discrimination of a target note more than an imaged context. Alternatively, timbres might have been perceived and represented in relation to each other, but imagery for the opening context might have been sufficiently veridical as to facilitate the task as well as actually hearing the context. This would be surprising in view of post-experiment interviews, which indicate that far from generating a veridical timbral image during the imagery task, many participants lacked a conscious timbral awareness in their mental representation of the musical sequence. Finally, the most plausible explanation that would account for group differences is that of a practice effect in which completing the perception task before the imagery task was an advantage to both overall accuracy and performance on the imagery task.

Target note was significant for the accuracy of Group B but not for Group A, with a mean of 78.1% for target 1, 66.7% for target 2, and 80.2% for target 3 ($X^2[2] = 9.21$, $p < 0.01$).

There were no effects of condition or of group for response time (RT). However, RT was significantly shorter the later the target note occurs in the sequence [$F_{2,\ 16} = 7.42$, $p < 0.05$], with a mean of 1327 ms for target 1, 1295 ms for target 2, and 1252 ms for target 3.

PARTICIPANT FEEDBACK

It is interesting to analyse how easy or difficult participants reportedly found performing the perception and imagery tasks. On the whole, the perception task was viewed as being moderately difficult to easy, though two respondents framed their experiences as wholly problematic. For both, the difficulty was an effect of a sensory overload when perceiving the full timbral sequence. One participant reasoned that in the perception task he had ‘more to deal with’ than in the imagery task.

Overall, the evidence from the post-experiment interviews is that imagery for timbre, defined as a conscious internal hearing of timbre, was not veridical for many participants in the study. Yet the hypothesis that participants can internalise timbral aspects of music is at least partly borne out by the ability to perform the discrimination task. Mean accuracy for imagery tasks (78.5%) is above chance, suggesting that an implicit knowledge of the timbre facilitated response.

Discussion

The ability to perform the discrimination task is evidence of the ability to internalise timbre. This timbral knowledge appears to be largely tacit, given the few reports of any ability to inwardly hear it. As predicted, participants were better able to perform the discrimination task under perceptual rather than imagery conditions (though this result is reflected only in accuracy, and not in response times). Two alternative accounts can be suggested. First, it could be inferred that the veridicality of the imagined context was insufficient to match the facilitation of the perceived timbral context. Facilitation in the discrimination task by perceiving the immediately preceding context is consistent with relative perception from the timbre of one note to another, rather than an absolute perception for each note in isolation. Alternatively, better timbre discrimination following perception than imagery could indicate that imaging a context is effortful, and thus diminishes the cognitive resources available for response to the target note.

RT was long (M = 1291 ms, SD = 544 ms), suggesting that the discrimination task was challenging, and in such instances accuracy data provide a more meaningful measure of response. For Group A, perception and imagery tasks elicited undifferentiated RT and accuracy. Could it be that imagery for this group was particularly accurate, providing as stable a context for the discrimination task as a perceptual context? Participant feedback is suggestive, indicating that perceptual cues were not particularly helpful to the timbre discrimination task, perhaps because the timbre events were not perceptually related so as to form a coherent musical context. Contrary to the hypothesis that timbre discrimination would be facilitated by a perceived timbral context, a couple of respondents felt they required silence in order to focus on hearing and judging the timbral target. It is far from clear from the varied testimonies whether timbre is perceived and imagined in an absolute or a relative way. Important to an investigation of imagery is the finding that many participants focused attention exclusively on the target notes without consciously drawing on the context to help.

In learning a traditional piece of music, the melody is primarily attended to. For most people, relative rather than absolute pitch is the useful skill in doing so: there is no need to identify the exact pitches heard, as long as they can be related to an overall contour. Repeatedly however, participants in this study expressed a desire to identify the sounds they heard, stressing that this would have helped them to learn the music. Recognition of the instrument was deemed important to image formation. It was mainly for this reason that the majority of respondents (nine) clearly stated that they would have found the use of acoustic rather than sampled sounds an advantage. The recognition and identification of the stimulus sounds were particularly important for three participants who explained that they would tend to visualise an instrument or instrumentalist when imaging a timbre. The current experiment used sampled sounds, believed to relate fairly well perceptually to acoustic instrumental sound, but the majority of respondents thought that the task might have been helped by the use of actual acoustic musical instruments.

Two participants pointed out that the sequence was not easy to internalise due to its atonal construction. This characteristic led to problems for some in correctly pitching the sequence during the imagery task: ‘The notes weren’t necessarily the pitch I expected to hear, which then threw me because I wasn’t expecting that sound’ [3]. While a sense of tonality was deliberately avoided to focus attention on timbre rather than a pitch pattern, it is possible that a monotone or tonal sequence might have produced different results (see Krumhansl and Iverson, 1992). In this pilot study, the equal presentation of original and altered target trials meant that the altered version was reinforced, potentially leading to confusion as to the correct version. Moreover, as each trial was presented, the sequence cut off after the onset of the target note. This meant that target one was heard or imagined in all trials, and target three in only a third of trials. It might be expected that the greater familiarity with target one than later targets would enhance performance, which appears to have translated to reduced RT. Additionally, an insufficient familiarity with the stimulus was commonly reported, suggesting that more hearings of the music prior to the main experiment session could have been demanded. Defined as a conscious inner hearing, imagery for timbre was not described as veridical by participants in the pilot experiment. By encouraging greater familiarity with ecologically valid musical material, a further experiment should show to what extent this phenomenon was a methodological artefact, or an intrinsic characteristic of timbre. In the remainder of this paper a more refined experiment is outlined to further explore imagery and perception for timbre in familiar music.

MAIN EXPERIMENT

A different facet of timbre was investigated in the main experiment: rather than focusing on the discrimination of individual instruments, participants were tested for discrimination between target moments of music that had been artificially filtered, and those that had been left unfiltered (the original) in familiar pop music. This approach encouraged a more relative than absolute mental representation of timbre: the juxtaposition of a filtered target to the original timbre might entail more of a relating between events than an absolute encoding of isolated timbres.

In order to ensure that participants do generate a musical image through the ‘gap’, a control condition was added, in which participants would hear a fragment of an unknown pop song, silence, then hear more of the same song (either filtered or unfiltered). The task was to decide whether the second extract of song seemed to have a filtered sound or not. This condition was closely equivalent to the imagery condition, but without instructions to image through the ‘silence’, and before the music had been learned by the participants.

Hypotheses remain from the previous experiment. First, participants can internalise and image timbral aspects of music. Second, participants will discriminate timbre better under perception than imagery conditions. In addition, it was hypothesized that participants would discriminate timbre better under imagery than control conditions. Any outcome to the contrary would be an indication that the cognitive load involved in imaging a context negatively impacts on target discrimination as compared with a ‘silent’ context.

Participants

Seventeen university music students (aged 18-37 years) volunteered to take part in the study. Some of these had previously participated in musical imagery experiments, but their prior experience was not judged to affect the study. All reported normal hearing.

Material

Pulp *This is Hardcore* was chosen as music comprising clear blocks of instrumentally varied music. Björk *Scatterheart* was chosen as music comprising sounds varied in filtered quality. Each track was passed through a pre-determined filter from the target moment onwards. Either bandstop or bandpass filters of different bandwidths around a selected centre frequency provided a contrast to the original timbre. These were set to output after a delay appropriate to the relevant target moment. ‘Bandpass’ filters out partials at the extremes of the sound spectrum, while ‘bandstop’ has the opposite effect. It was important to ensure that altered target moments (filtered) were relatively subtle to avoid a ceiling effect. Consequently, perceptual effectiveness rather than parametric equivalence was the guiding principle in choosing filters. Three independent listeners, who were aware of the experiment aims, listened to the sequences and confirmed the appropriateness of the degree of filtering applied.

Sections of each track rather than the entire piece were presented via a Macintosh G3 computer. Examples of score sections that were used are shown in Figure 4. The arrows pointing upwards indicate the onset of the target moment.

Fig. 4. Score samples from each of the pieces used (Pulp target one, Björk target three). The silent or imaged section lies between the dotted lines of each score. Arrows indicate the onset of the target music.

Max MSP software (*Opcode*) enabled the manipulation of the music and the creation of a separate music file for each target and its variations.

Initially, sequences for the perception condition of the task were created. Three target moments in each track were selected for timbral discrimination. These were allocated three variations of timbre: the original timbral quality, bandpass filtered sound, and bandstop filtered sound. Counterpart sequences for the imagery and control tasks made use of the materials created for the perception task and 'silenced' all but the opening few seconds of each sequence and the target music (the imagery zone is marked between dotted lines in Figure 4). This experiment differs from the pilot as the music continued after the onset of the target moment. For altered targets this meant that all of the music from the target moment onwards was filtered.

During the imagery condition a visual metronome (a blinking spot) on the computer screen served to keep participants in time when imaging music.

Procedure

A control condition was introduced in the place of the pre-test of the pilot study. In this, participants heard the first few seconds of music, followed by a period of silence, then more music (the target). At the onset of this target music they were instructed to decide if it seemed to be filtered or unfiltered, indicating their choice by computer key press (key 'S' labelled unfiltered and 'K' labelled filtered) as quickly but as accurately as possible. Testing in this control session was grouped by piece. Each piece comprised nine trials (three target notes x three variations) that were presented once and in random order. Participants listened individually over headphones. A number of practice attempts were allowed, to learn what was understood by a filtered sound.

Having completed the control condition, a CD of each original track was distributed with instructions to pay particular attention to the timbre of the music. Participants were provided with skeleton scores of the three target moments of each track. They were also given song lyric sheets, and told that they could look at these while familiarising themselves with the music. A minimum of ten, but preferably at least twenty listenings over an average of seven days was required. After the 3rd, 10th and final listening participants were asked to note how well they liked the music on a scale of 1 to 7. The purpose of this was to focus the listeners' attention on the music via a specific task.

After the familiarisation phase, participants performed the main experiment session: hearing or imaging the musical context they were asked to judge whether a sounded target moment was filtered or unfiltered (original). The score relevant to the trial (the same skeleton scores provided for the familiarisation phase) was positioned below the computer monitor. As in the control condition, key press indicated response. Filling out a questionnaire asking for background information separated the imagery and perception blocks of this main session. Participants took part in a short post-experiment interview.

Participants performed the perception and imagery tasks in a counterbalanced order:

Control, perception, imagery (n=8)
Control, imagery, perception (n=9)

The main experiment consisted of two conditions (imagery and perception), which in turn comprised separate blocks for each of the two pieces, made up of nine trials each (3 x 3 targets per piece). Trials within each block were presented once and in random order.

As in the pilot experiment, RTs for correct response were analysed using a 5-factor ANOVA (factors 'group', 'piece order', 'condition', 'target' and 'filter'). Accuracy was analysed using a generalised chi-square. Post-experiment interviews were again analysed thematically.

Results

When data from Pulp and Björk targets are pooled, piece becomes a highly significant factor. RT was longer for Björk (mean of 4104 ms compared with 3732 ms for Pulp) [$F_{1,16} = 9.75$, $p < 0.01$], and accuracy was higher for Pulp (mean of 91.5% compared with 76% for Björk) ($X^2[1] = 17.1$, $p < 0.0001$). As neither of the between-subjects variables ('group' and 'piece order') was significant in RT or accuracy, they have been collapsed to focus on the within-subjects factor of condition. Table 3 provides a summary, including mean RT and accuracy (%) for each piece.

Table 3. A summary of the main effects of condition, separated per piece

	Condition	Mean RT	Mean accuracy
Pulp	Control	5482 ms	88.9%
	Imagery	3131 ms	90.0%
	Perception	2365 ms	96.0%
		$F_{2,32} = 11.1$, $p = 0.001$	Not significant
Björk	Control	6138 ms	61.4%
	Imagery	3662 ms	75.8%
	Perception	2587 ms	92.2%
		$F_{2,32} = 10.8$, $p = 0.001$	$X^2[2] = 40.2$, $p < 0.001$

The anticipated effect of condition (control/imagery/perception) was highly significant for RT [Pulp $F_{2,32} = 11.1$, $p = 0.001$, Björk $F_{2,32} = 10.8$, $p = 0.001$]. Planned comparisons reveal that imagery tasks took significantly less time to perform than control tasks ($p < 0.005$ for both pieces). Perception tasks produced the shortest RT, taking significantly less time than the imagery task ($p < 0.005$ for both pieces). Figure 5 illustrates this result collapsed across both pieces. This finding supports the hypothesis that participants are facilitated in their judgement of timbre under imagery more than control conditions. Of course, it could be argued that as the imagery condition was performed after the control condition, the shorter RT in the former arose from practice in the timbre discrimination task. However, we would expect to see such a practice effect in the form of a correlation of accuracy with trial number. This was not the case for the control block [r = -0.37; d.f. = 16; p = n.s.]. Nor was a correlation between trial number and accuracy found across the two main portions of the experiment (imagery and perception combined) [r = -0.07; d.f. = 34; p = n.s.].

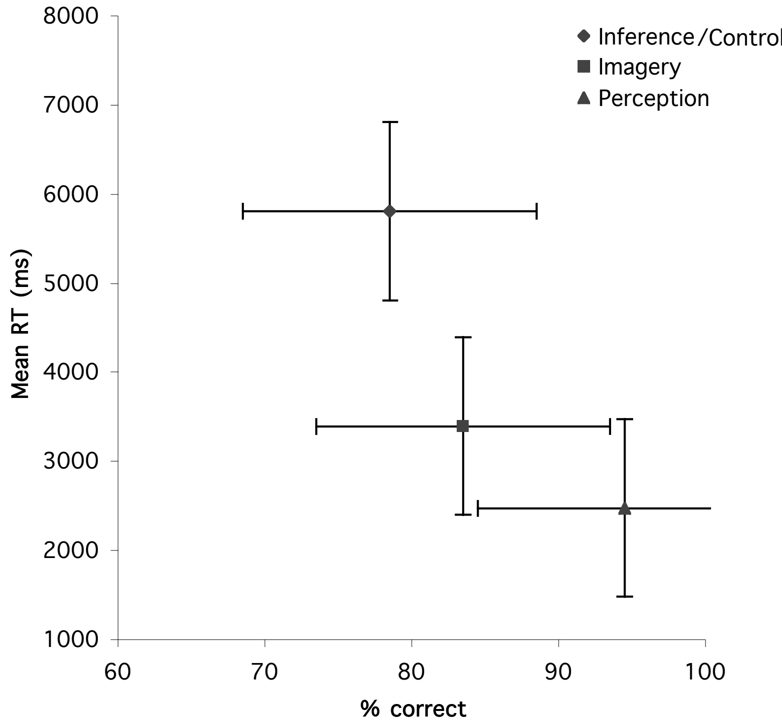

Fig. 5. Overall mean RT (ms) and accuracy (%) for control, imagery and perception conditions.

Condition was significant in accuracy for Björk ($X^2[2] = 40.2$, $p < 0.001$) but not for Pulp. Perception tasks led to more accurate performance than imagery tasks (Pulp $X^2[1] = 4.9$, $p < 0.05$; Björk $X^2[1] = 15.2$, $p < 0.001$ – see Table 3). Figure 5 also illustrates that imagery trials were more accurate than control trials ($X^2[1] = 7.3$, $p < 0.01$ for Björk, not significant for Pulp). Overall, accuracy and RT support the experimental hypotheses. This constitutes evidence that timbre is a musical dimension that may be mentally represented, and that imaging music in this particular experiment provided a facilitatory context for the task discrimination as compared to the music-free control context.

While various significant differences emerge between targets, their interest is limited to a general indication of the ease or difficulty of each target. As such they will not be presented here. There was no main effect of filter type for either piece.

PARTICIPANT FEEDBACK

All participants reported finding the perception task easier than the imagery task. The purpose of this experiment was to focus on imagery for timbre. Nevertheless, many participants say they were unable to image sound quality, or that timbre only occasionally featured in a patchy fashion. Only two respondents spoke of their experience of the control test. Both claimed that it was more difficult than the two tasks of the main experiment session. One explicitly volunteered that having an image of the music helped her to complete the tasks. As in the pilot study, participants were asked how well they felt they knew the music they had been asked to listen to before the main session began. This time responses were fairly positive, contrasting with the more insecure reaction of participants in the pilot. A commonly shared view about the

Björk was that this was difficult to get to know. In light of this impression, poorer performance on the Björk than the Pulp track fits with a general feeling of greater familiarity with the latter. In the previous experiment the issue of electronic versus acoustic instruments was raised. This time a similar subject was introduced by participants who expressed a difficulty with the electronic sounds in the Björk track.

Discussion

The hypotheses tested by this study have been supported by both the experiment and the interview data collected. Significant global response time patterns for both pieces indicate that the imagery condition was sufficient to facilitate task performance as compared to the control condition. With no evidence of a practice effect within experimental blocks, it is suggested that a mental representation of the music facilitated response in the imagery condition as compared to the control in which no mental representation of the music was possible. Differences in accuracy for the variable 'condition' were significant for Björk but not Pulp. Here though, perception and imagery responses were statistically different showing that imagery was less effective than perception as a context.

Examining perception, imagery and control data separately, some patterns emerge which are consistent with previous experiment findings. Interview data suggest that participants approach imagery and perception tasks differently, tackling the first as an active comparison of musical events, with the latter as more of an instantaneous decision. Imagery tasks involve recalling music against which to match the heard stimulus, while perception tasks call for a more instinctive and immediate comparison of adjacent musical moments. For instance, one participant found the perception task easier to perform for this reason, saying that it was "easier because you had the two things next to each other." In imagery tasks the lack of sensory input denies the participant the auditory feedback and checking mechanism that might lend them confidence to perform the task (regardless of the true veridicality of their imagery).

The differences found between the Björk and Pulp results are largely attributable to the perceived complexity of each piece. The Björk track proved more challenging than the Pulp track. Interview data show that this music was felt to change a lot, and to contain a lot of detail. The increased demands that extra detail in the sound place on memory will affect the conscious recollection, or imagery, of that music. *Scatterheart* makes use of filter effects as an integral part of the musical interest. It would not be surprising to find that participants judged additionally filtered targets to be part of the original musical fabric, thus reducing their accuracy.

No group or order effects emerged in this experiment, and familiarity with the experiment stimuli seemed to be fairly assured. The practice effects of the pilot study were not replicated, indicating that the choice of music and increased number of suggested hearings for stimuli produced the desired improvement in experimental design. The length of the songs may have necessitated a broad appreciation of the music as a whole rather than a detailed analysis. This is in keeping with the sort of listening exposure and attention that might naturally be paid to music absorbed as a mental image.

GENERAL DISCUSSION AND FURTHER WORK

These studies demonstrate that music students can internalise timbre sufficiently to aid performance on a forced-choice timbre discrimination task. The quantitative differences between the imagery and perception performance combine to suggest at least a functional equivalence between the two. In other words, the imagery context to the discrimination task acted as weaker, but similar, facilitation than the heard musical context. In the main experiment, the control condition allowed a comparison between behaviour with and without a mental image of the music. Differences between imagery and control data show that performance with knowledge of the music was significantly superior to that without. Thus the imagery task evoked a more developed mental representation than the implicit musical knowledge adequate to the control task. However, while the music students participating in the study were able to mentally represent timbre, this was not always reported as being a conscious dimension of their image. This is unlike the imagery for pitch and timing reported in Bailes (2002), and would fit with the findings of Intons-Peterson (1980), who proposes that timbre may be a sound characteristic that is optionally represented. As discussed in the introduction, there is a suggestion that while the human body may easily reproduce pitch and timing, the capacity for timbral production is limited. Thus timbre may more properly be associated with the imagery of the 'inner ear' than the 'inner voice' (Kalakoski, 2001).

The two experiments reported in this paper progressed from separate and distinct timbre events towards the more inter-related timbral arrangement of 'real' music. While participants struggled to listen, learn and recall the timbre sequence in the pilot study, results in the main experiment are the result of greater familiarity with the musical stimuli, providing more consequential data. This is reflected in the different importance of response time over accuracy measures for each experiment. While the main hypothesis was supported by accuracy and not response times in the pilot study, both forms of data are useful in the main experiment, suggesting that response time was facilitated by greater conscious awareness of the music employed. In the pilot study there was a greater dependence on tacit timbral knowledge rather than explicit timbral imagery. Results were noticeably poor even under perception conditions. By contrast, response in the main experiment was still dependent on tacit knowledge, but to a lesser extent, with greater familiarity and image veridicality being reported. The more meaningful nature of the music seems to have led to greater knowledge of the timbral content on implicit and explicit levels of cognition.

To summarise, the experiments reported in this paper are an important preliminary investigation of the veridicality of timbral imagery in music. It seems that timbre can be internalised by music students, and that imaging a timbrally rich context may facilitate performance on a timbral discrimination task. However, important questions remain as to the nature of any equivalence between imagery and perception for timbre, and these may be more comprehensively addressed in future research by manipulating the association of the discrimination task parameter with the musical context. In other words, a clearer understanding of the relative or absolute perception of timbre will help in devising a discrimination task that is highly dependent on either a mental or an actual hearing of the preceding, or priming, context. Until such a systematic control is undertaken, it is uncertain to what extent the relatively subconscious nature of timbral re-presentation revealed by this research is indeed an integral feature of intentionally creating imagery for timbre, and whether this characteristic is linked to its experience as the product of an 'inner ear' rather than an 'inner voice' [4].

NOTES

[1] Note that the auditory imagery experienced during hallucination, which is sufficiently veridical as to be confused with actual perception, is not under consideration in the present study.

[2] Four of these participants had previously volunteered and participated in a musical imagery experiment, but as this was quite dissimilar it was not anticipated that this would affect their performance.

[3] Verbatim evidence in this article has been edited for the sake of brevity.

[4] Research reported in this paper was supported by *Arts and Humanities Research Board* funding for doctoral work undertaken at the University of Sheffield. I would like to thank Emmanuel Bigand, Eric Clarke, and Roger Dean for helpful comments on previous versions of this manuscript.

REFERENCES

Auhagen, W., & Schoner, V. (2001). Control of Timbre by Musicians – A Preliminary Report. In: R.I. Godøy & H. Jørgensen (Eds.), *Musical Imagery.* Lisse: Swets & Zeitlinger, pp. 201-217.

Bailes, F.A. (2002). *Musical Imagery: Hearing and Imagining Music.* PhD thesis, University of Sheffield.

Bailes, F. (in press). The prevalence and nature of imagined music in the everyday lives of music students. *Psychology of Music.*

Björk (2000). 'Scatterheart' *Dancer in the Dark.* Björk Overseas Ltd./One Little Indian Ltd. TPLP151CD. Audio CD.

Buckley, R.J. (1912). *Sir Edward Elgar.* 2nd edition. London: John Lane.

Cohen, D., & Inbar, E. (2001). Imagery of Emotional Expression in Music and Speech. In: R.I. Godøy & H. Jørgensen (Eds.), *Musical Imagery.* Lisse: Swets & Zeitlinger. pp. 137-159.

Denis, M. (1989). *Image and Cognition.* Trans. By M. Denis & C. Greenbaum. Harvester Wheatsheaf.

Ehresman, D., & Wessel, D. (1978). Perception of timbral analogies. *Rapports IRCAM, 13/78.* Paris.

Finke, R.A. (1989). *Principles of Mental Imagery.* Cambridge, Massachusetts: The MIT Press.

Grey, J.M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, Vol. 61, pp. 1270-1277.

Halpern, A.R., Zatorre, R.J., Bouffard, M., & Johnson, J.A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, Vol. 42, pp. 1281-1292.

Intons-Peterson, M.J. (1980). The role of loudness in auditory imagery. *Memory & Cognition*, Vol. 8, No. 5, pp. 385-393.

Kalakoski, V. (2001). Musical Imagery and Working Memory. In: R.I. Godøy & H. Jørgensen (Eds.), *Musical Imagery.* Lisse: Swets & Zeitlinger. pp. 43-55.

Krumhansl, C.L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 18, No. 3, pp. 739-751.

Lerdahl, F. (1987). Timbre hierarchies. *Contemporary Music Review*, Vol. 2, No. 1, pp. 25-44.

McAdams, S. (1989). Psychological constraints on form-bearing dimensions in music. *Contemporary Music Review*, Vol. 4, pp. 181-198.

Melara, R.D., & Marks, L.E. (1990). Interaction among auditory dimensions: Timbre, pitch, and loudness. *Perception & Psychophysics*, Vol. 48, pp. 169-178.

Mountain, R. (2001). Composers and imagery: Myths and realities. In: R.I. Godøy & H. Jørgensen (Eds.), *Musical Imagery.* Lisse: Swets & Zeitlinger. pp. 271-288.

Pitt, M.A., & Crowder, R. G. (1992). The role of spectral and dynamic cues in imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance*, Vol. 18, pp. 728-738.

Pulp (1998). 'This is Hardcore' *This is Hardcore.* Island. 524 486-2. Audio CD.

Saxton, R. (1998). The Process of Composition from Detection to Confection. In: W. Thomas (Ed.), *Composition, Performance, Reception: Studies in the Creative Process in Music.* Ashgate: Aldershot, pp. 1-16.

Stravinsky, I., & Craft, R. (1979). *Conversations with Igor Stravinsky*. London: Faber Music Ltd.

Wessel, D.L. (1979). Timbre space as a musical control structure. *Computer Music Journal*, Vol. 3, No. 2, pp. 45-52. Reprinted in C. Roads & J. Strawn (Eds.), (1985) *Foundations of Computer Music.* Cambridge, Massachusetts: MIT Press.

Wolpert, R.S. (1990). Recognition of melody, harmonic accompaniment, and instrumentation: Musicians vs. nonmusicians. *Music Perception*, Vol. 8, No. 1, pp. 95-106.