Journal of Experimental Psychology:
Human Perception and Performance
1992, Vol. 18, No. 3, 728-738

0096-1523/92/$3.00

The Role of Spectral and Dynamic Cues in Imagery for Musical Timbre

Mark A. Pitt
Ohio State University

Robert G. Crowder
Yale University

The musical quality of timbre is based on both spectral and dynamic acoustic cues. Four 2-part experiments examined whether these properties are represented in the mental image of a musical timbre. Experiment 1 established that imagery occurs for timbre variations within a single musical instrument, using plucked and bowed tones from a cello. Experiments 2 and 3 used synthetic stimuli that varied in either spectral or dynamic properties only, to investigate imagery with strict acoustic control over the stimuli. Experiment 4 explored whether the dimension of loudness is stored in an auditory image. Spectral properties appear to play a much larger role than dynamic properties in imagery for musical timbre.

Ludwig van Beethoven's Ninth Symphony is one example par excellence of the amazing reach of auditory imagery, for Beethoven was deaf when he composed this piece as well as most of his other famous works. His imagery capabilities are even more impressive in light of the artistic demands placed on any composer in the orchestral medium: Not only must the various melodic lines and accompaniment be planned, but they must also be realized by specific instruments and instrumental combinations that will produce a desired sound color.

Despite the importance that composers and musicians have confidently ascribed to musical imagery over the centuries (Seashore, 1938/1967), psychologists began to investigate this topic in the laboratory only in the last 10 years or so (see summaries in Crowder, 1989; Halpern, 1989). This work has examined imagery for pitch information (Farah & Smith, 1983; Segal & Fusella, 1970), chords (Hubbard & Stoeckig, 1988), and melodies (Halpern, 1988, 1989).

Recently, Crowder (1989) investigated imagery for another property of musical objects, timbre. *Timbre* can be defined as the characteristic quality of a sound—other than its pitch, loudness, or duration—that identifies it uniquely (as a flute, violin, piano, etc.). A primary aim of Crowder's study had been to demonstrate that auditory imagery is based on sensory (auditory) processing and not motor (articulatory) processing. Some of the studies that examined imagery for pitch (Farah & Smith, 1983; Segal & Fusella, 1970) can be explained by assuming that subjects produced the neural correlates of a to-be-imagined event by singing or humming it, either aloud or to themselves, rather than by a process of "hearing the event in the mind's ear." As a result, these studies do not unequivocally demonstrate the existence of genuinely auditory imagery. A similar argument applies to investigations of mental imagery for songs (Halpern, 1988), which can be assigned either to a process of mental hearing or alternatively to a process of singing to oneself.

Because the normal human vocal tract cannot faithfully reproduce the timbres of most musical instruments, examining imagery for musical timbre provided a straightforward means of showing that auditory imagery is sensory based rather than somehow motor based. In his demonstration, Crowder (1989) adapted an experimental technique that had been used to study visual imagery (Posner, Boies, Eichelman, & Taylor, 1969). Essentially, the technique is designed to assess whether imagery mirrors perception.

In Part A of Crowder's study (Experiment 1), subjects were required to judge whether two consecutively presented pitches that could also vary in timbre were of the same or of a different diatonic pitch. The results showed that correct same-pitch responses were significantly faster when the timbres of the two notes were the same than when they were different, documenting that timbre information was stored with the memory trace of the first pitch, at least until arrival of the second pitch. The same procedure was used in Part B of Crowder's study (Experiment 2), except that subjects were first presented with a pitch that was neutral in timbre (a sine-wave tone) and then asked to imagine this sine-wave pitch as it would *sound* when played by a particular instrument (e.g., flute, trumpet, or guitar). After subjects had indicated that an image had been formed, the second tone of the pair was presented for a same–different pitch judgment. Although reaction times (RTs) were slower overall in Part B than in Part A, the same pattern of results was found: Same-pitch responses were faster when the imagined timbre (of the first tone) was the same as that presented (the second tone) than when it was different. Crowder interpreted these findings as support for the existence of sensory-based auditory imagery.

In the present study, we explored imagery for musical timbre further by investigating which acoustic attributes of a timbre are represented in its mental image. Two primary sources of acoustic information are known to contribute to the perception of timbre (Dowling & Harwood, 1986): spectral

This research was partially supported by National Science Foundation Grant GB 86 08344 to Robert G. Crowder.

We thank James Sawusch and three anonymous reviewers for many helpful comments on the article.

Correspondence concerning this article should be addressed to Mark A. Pitt, Department of Psychology, 1885 Neil Avenue Mall, Ohio State University, Columbus, Ohio 43210, or Robert G. Crowder, Department of Psychology, Box 11-A Yale Station, Yale University, New Haven, Connecticut 06520.

properties and dynamic properties. *Spectral properties* refer to differences found in the harmonic (overtone) structure of a pitch when played by different instruments. They are a result of the shape of the instrument, among other things. For example, the odd harmonics predominate in a clarinet tone because of its cylindrical wood bore. The conical bore of an oboe produces sounds of a different spectral configuration.

Dynamic properties of timbre refer to the rapid changes that typically occur at note onset and offset, such as attack rate (the rate of note onset), and changes in harmonic structure during note onset. Thus, the same instrument can have different timbres when excited in different ways, such as a piano string that is either struck by a hammer or plucked directly with a fingertip.

Four experiments, each with two parts, examined whether spectral and dynamic properties are represented in the image of a musical timbre. Experiment 1 investigated image formation with a single, real musical instrument. Experiments 2 and 3 used synthetic stimuli that varied in either spectral or dynamic properties in an attempt to investigate imagery with more control over acoustic factors than is possible with real instruments. The final experiment explored whether the dimension of loudness is stored in an auditory image. The main finding of these experiments was that spectral properties appear to play a much larger role than dynamic properties in imagery for musical timbre.

General Method

The experimental procedure was the same throughout the four experiments; only the stimuli varied. Parts A and B are described here, with the necessary additions detailed in each experiment.

Part A: Two-Tone Comparisons

Experimental design. The experiments used a within-subjects design with two orthogonal variables: the pitches of the two tones (same or different) and the timbres of the two tones (same or different). Thus, there were four conditions, each of which was presented equally often. The order of tone presentation was counterbalanced across trials. Five blocks of 24 trials (randomly permuted) were presented, the first being discarded as practice. All pitch–timbre conditions occurred with equal frequency in each block.

Materials. Although the timbres that were used varied across the four experiments, the pitches of the tones did not. Each tone was at one of three pitches: F4 (349 Hz), G4 (392 Hz), or A4 (440 Hz).

Apparatus. Subjects were tested individually in a sound-attenuated booth. Stimulus presentation and response collection were controlled by an IBM PC-AT computer that was interfaced with an audio amplifier. Stimuli were digitally recorded (10 kHz sampling rate, low-pass filtered with a corner frequency of 4.8 kHz) and were presented through loudspeakers at a comfortable listening level. The "z" and the "?" keys on the computer keyboard served as the same and different response buttons, respectively.

Procedure. Subjects sat in front of the computer and were instructed to make a speeded same–different judgment as to whether the second of two consecutively presented notes had the same pitch as the first. The instructions emphasized making the judgment on the basis of (diatonic) pitch information alone, ignoring changes in timbre. Example trials accompanied the instructions. A trial was initiated by pressing the space bar. After a pause of 1,500 ms, the first tone was presented (all tones were truncated, at the zero crossing, to be approximately 250 ms in duration), followed by an interstimulus interval (ISI) of 500 ms. The second tone was then played, at which point the speeded response was to be made. The experiment was self-paced and lasted about 15 min.

At the end of the testing session, subjects filled out a questionnaire inquiring about their musical background. The questionnaire focused on three aspects of musical training: (a) the number of years of formal training on an instrument, (b) how recent this training had been, and (c) what academic courses in music had been taken (e.g., theory, composition).

Part B: Imagery

The experimental procedure in the imagery session was modified from the two-tone comparison session in the following ways: On each trial, subjects first heard a sine-wave tone (250-ms duration) at the fundamental frequency of one of the three pitches (349, 392, or 440 Hz). The sine wave was chosen because it was thought to be relatively neutral with respect to timbre. Simultaneously with the presentation of the sine wave, the name of one of the timbres was printed in the center of the computer screen. Subjects were instructed to imagine the specified timbre at the pitch of the sine wave. The image-formation process was self-paced. Subjects indicated that the image had been formed by pressing the return key. None indicated an inability to generate these mental images. The second tone was presented after a pause of 1,500 ms, which was set at this duration to prevent any clicking sounds produced by the return key from being simultaneous with the second tone. Subjects had to make a speeded response as to whether the pitch of the second tone was the same as or different from the imagined tone. (Note that this task can in principle be performed accurately without generating the image of the timbre. A similar pattern of data found here as in Part A, however, suggests that subjects indeed complied with the imagery instructions.)

A lengthy instruction phase was conducted prior to the testing session to provide subjects with practice in imagining timbres. First, subjects were introduced to the vocabulary of stimuli, hearing each timbre at all three pitches; the name of the timbre being played was printed on the computer screen. Next, subjects practiced imagining the timbres at different pitches. A sine-wave tone was presented along with the name of a timbre. Subjects had to imagine the specified timbre at the pitch of the sine wave. When the image was formed, subjects pressed the space bar and received feedback in the form of the specified timbre playing the same pitch as the sine wave. At least nine practice trials were given, and subjects were free to continue this process until they felt confident in their ability to imagine the timbres. Six blocks of 32 trials followed, with the first discarded as practice. The testing session lasted approximately 40 min.

Experiment 1a: Perception of Timbre Differences in Natural Cello Tokens

In Crowder's (1989) experiment, three different instruments (flute, guitar, and trumpet) constituted the timbres that were investigated. These instruments not only sound distinct from one another, but they also are quite varied in other respects as well, such as the way they look and feel, and their names. The first of these, visual imagery, is far too sluggish to have been the agency for faster same-pitch RTs with matching timbres than with mismatching timbres (Paivio, 1979). But strictly on logical grounds, *any* imagined property of the target instrument could have been responsible for the matching

Table 1
Means and Standard Deviations (in Milliseconds and Percentage Correct) for the Perception (Experiment 1a) and Imagery (Experiment 1b) Sessions With the Two Cello Timbres (Pluck and Bow)

	Perception				Imagery			
	Same pitch		Different pitch		Same pitch		Different pitch	
Dependent measure	Same timbre	Different timbre	Same timbre	Different timbre	"Same timbre"	"Different timbre"	"Same timbre"	"Different timbre"
Reaction time	805	1,053	930	1,017	1,231	1,257	1,339	1,332
SD	146	232	198	257	608	595	743	655
Accuracy (in %)	90	42	66	86	76	60	69	79
SD	25	41	41	21	9	21	19	17

Note. Quotation marks indicate that the timbre of the first note was imagined.

effect. For this reason, we wanted to keep the source instrument constant while varying the timbre of the tones. This was accomplished by using the plucking and bowing sounds of a violincello (Cutting, Rosner, & Foard, 1976; see also Rosen & Howell, 1981).

The purpose of Experiment 1a was to determine whether variations in timbre from a single instrument would affect pitch-comparison judgments. The prediction was that same responses should be faster when the timbres are the same than when they are different. Additionally, we thought that plucked and bowed notes on the same instrument would form a good first approximation to studying dynamic cues to timbre while holding spectral cues constant (however, see the *Results and Discussion* section of Experiment 1b).

Method

Subjects. Twenty students from an introductory psychology course participated in this experiment in exchange for course credit.

Materials. A cellist played the six notes (three bowed at the standard pitches and three plucked). The notes were recorded onto audiotape in a sound-attenuated booth. Measuring from note onset, the notes were truncated to 250 ms. The plucked tones reached peak amplitude an average of 22 ms after note onset and then decayed slowly until note offset. The bowed notes reached the steady-state portion of the waveform 220 ms after note onset. The plucked and bowed notes were equated for loudness during the digitization process.

Results and Discussion

We calculated mean RT (correct responses only) and percentage correct for each subject in each of the pitch–timbre conditions. These data were then collapsed across subjects and are shown on the left side of Table 1 (perception).

The primary variable of interest in this experiment was whether same–different judgments would be influenced by variations in timbre when the pitches of the notes were the same. Such an outcome would suggest that different cello timbres are represented in memory long enough to affect the response to the second tone. The tabulated results show that this was the case for both measures. Reaction times were an average of 147 ms faster in the same-pitch–same-timbre condition than in the same-pitch–different-timbre condition (805 vs. 1,053 ms). This difference was statistically reliable in a test of simple effects, $F(1, 13) = 17.70$, $p < .001$; 13 of 14 subjects showed the effect in the same direction.[1] Accuracy was 48 percentage points higher in the same-pitch–same-timbre condition than in the same-pitch–different-timbre condition, $F(1, 19) = 24.31$, $p < .0001$; 17 of 20 subjects showed the effect.

Reliable effects were also found in the omnibus analyses of variance (ANOVAs) on RT and accuracy. In the RT analyses, a main effect of timbre was significant, $F(1, 13) = 28.60$, $p < .0001$, with RTs being faster in the same-timbre than in the different-timbre condition. With different overt responses being compared across the two pitch conditions, this result is difficult to interpret. In the accuracy analyses, both main effects and the interaction were reliable: pitch, $F(1, 19) = 6.30$, $p < .02$; timbre, $F(1, 19) = 11.52$, $p < .003$; Pitch × Timbre interaction, $F(1, 19) = 16.57$, $p < .0007$. Again, the data are uninterpretable because different responses were made in the two pitch conditions.

The data of Experiment 1a are clear: Timbre differences can influence pitch judgments, even when the timbres are produced by the same instrument. This result replicates Crowder (1989) and provides a mandate for proceeding with the imagery part of the experiment.

Experiment 1b: Imagery for Natural Cello Tokens

The data from Experiment 1a suggest that different ways of producing a tone on the same instrument can be registered, at least briefly, in auditory memory. If these cues form part of the enduring memory trace of a generated timbre, then a similar pattern of data should be found here. The absence of such an effect would suggest that these cues are not part of the representation of an imagined timbre.

Method

Subjects. Thirteen students participated in this experiment. They were drawn from the same pool as those in Experiment 1a. Recall that in the imagery condition, subjects had to imagine a specified

[1] Six of the 20 subjects never responded correctly in the same-pitch–different-timbre condition (accuracy scores of 0). We removed their data from the RT analyses because there were no correct-response RTs from which to calculate a mean RT.

timbre (sine-wave tone + timbre name) and then to judge whether its pitch was the same as that of an actual timbre.

Results and Discussion

The data were collapsed and analyzed by using the same procedure that was used in the previous experiment. Mean RT and percentage accuracy are shown on the right side of Table 1 (imagery).

Inspection of the RT data shows responses to be slower than those in Experiment 1a. Crowder (1989) found a similar result, which can be attributed to the difficulty of performing the same–different task with an internally generated image rather than an explicit comparison. The RT data provide little support for the proposal that single-instrument cues to timbre can be imagined. Although RTs were in the expected direction, 26 ms faster in the same-pitch–"same-timbre" condition than in the same-pitch–"different-timbre" condition, the difference was not statistically significant, $F(1, 12) = 1.36$, $p = .27$. No other effect in the omnibus analysis was reliable.

On the other hand, the accuracy data suggest that single-instrument cues can be represented in the memory trace of a timbre. Subjects were significantly more accurate at responding *same* when the imagined and the presented timbres were the same (75.8%) than when they were different (59.6%), $F(1, 12) = 9.24$, $p < .01$. A modest main effect of pitch was also obtained, with accuracy being lower when the pitches were the same than when they were different, $F(1, 12) = 6.55$, $p < .03$. This result, however, is difficult to interpret because (a) the accuracy data move in the opposite direction of the RT data, which indicates a speed–accuracy trade-off, and (b) different overt responses were produced in the two conditions.

Because we obtained a similar pattern of results in Experiments 1a and 1b, the results suggest that different timbres stemming from the same instrument can be imagined. This finding rebuts criticisms that the data of Crowder (1989) were the result of subjects' use of nonauditory imagery strategies (such as imaging the shapes or names of the instruments). Such strategies would have been ineffective under the current circumstances, and had they been used, should have resulted in chance performance or much slower RTs.

Although the results of Experiment 1 replicate the findings of Crowder (1989) in demonstrating imagery for timbre, they cannot be taken to show that dynamic cues form part of the mental image of a timbre, as one might have supposed from our earlier discussion indicating that different timbres can be produced by exciting the same instrument in different ways. We had originally supposed that demonstrating imagery with the pluck and bow timbres would imply that dynamic cues to timbre were imagined. With natural instrumental timbres, however, it is possible that spectral cues covary with dynamic ones. Spectral analysis of the plucking and bowing sounds confirmed our suspicions of covariation: The plucking sounds consisted primarily of low-frequency harmonics, whereas a wide band of harmonics was present in the bowing sounds. Consequently, our findings could have resulted from either spectral or dynamic variation among the sounds used.

Only with synthesized tones can one be certain in advance of the acoustic properties of the stimuli. Therefore, to have this control, we decided to replicate Experiment 1 with synthesized timbres. The experiment was conducted twice. In Experiment 2, spectral cues were varied while dynamic cues were held constant. The reverse was carried out in Experiment 3.

Experiment 2a: Perception of Timbres Varying in Spectral Properties

Two timbres were synthesized that differed only in harmonic structure. One of these consisted of the fundamental frequency of tone + Harmonics 1–3. The other included the fundamental + Harmonics 4–6. We decided on these two timbres because of their spectral simplicity and because subjectively they were perceptually distinct. If spectral properties are represented in the memory trace of a timbre, then Experiment 2a and 2b should yield similar results.

Method

Subjects. Fifty Yale University undergraduates participated in this experiment in exchange for $5 or credit in an introductory psychology course. (The large number of people tested in this experiment, and their recruitment, is rationalized later.)

Materials. Two timbres varying only in harmonic structure were synthesized on a Yamaha DX100 digital synthesizer. One timbre consisted of the fundamental + Harmonics 1–3; this timbre sounded similar to that of a car horn. The other timbre comprised the fundamental + Harmonics 4–6; it had a timbre similar to that of an organ. All partials were of equal amplitude. The three pitches that were used in Experiment 1 (F4, G4, and A4) were synthesized on each timbre. The waveform envelope consisted of an abrupt onset and offset (recall that all notes were truncated to 250 ms) and a steady-state portion that maintained a constant amplitude.

Results and Discussion

Mean RT and percentage accuracy (collapsed over subjects) are shown in the top half of Table 2 as a function of the usual pitch and timbre conditions. Reaction times in the two same-pitch conditions yielded an unexpected pattern of results. Not only were RTs in the same-pitch–different-timbre condition not slower than those in the same-pitch–same-timbre conditions, they were significantly faster, $F(1, 49) = 43.38$, $p < .001$.

The accuracy data reveal that performance in the same-pitch–different-timbre condition was extremely difficult for some subjects and yet easy for others. Mean accuracy was close to chance (56%), and the standard deviation was large (46%). More evidence for this observation is displayed in Figure 1, which is an ordinal distribution of accuracy scores in the same-pitch–different-timbre condition. As can be seen, the distribution is bimodal, with no scores falling between 37% and 70% correct. Indeed, a majority of the subjects (62%) were at the extremes of the distribution, almost evenly split between those who never made an error and those who never made a correct response. These data indicate that people differ greatly in their ability to disregard changes in timbre while attending to pitch.

An interpretation of these individual differences was sought in the answers provided in the musical-background question-

Table 2
Means and Standard Deviations (in Milliseconds and Percentage Correct) for the Perception (Experiment 2a) and Imagery (Experiment 2b) Sessions With Timbres Varying in Spectral Properties

	Perception				Imagery			
	Same pitch		Different pitch		Same pitch		Different pitch	
Subjects	Same timbre	Different timbre	Same timbre	Different timbre	"Same timbre"	"Different timbre"	"Same timbre"	"Different timbre"
Unselected								
Reaction time	624	547	676	678				
SD	143	430	186	183				
Accuracy (in %)	98	56	91	95				
SD	5	46	20	16				
Qualified								
Reaction time	576	733	631	683	666	692	663	676
SD	118	177	133	190	259	261	237	253
Accuracy (in %)	99	96	95	93	86	72	86	98
SD	2	7	17	20	2	3	3	5

Note. Quotation marks indicate that the timbre of the first note was imagined.

naire that subjects filled out (see the General Method section). Two judges independently rank-ordered the 50 subjects in terms of musical ability. The two orderings were then combined and correlated with accuracy in the same-pitch–different-timbre condition. The resulting correlation, $r(48) = .66$, $p < .001$, indicates that musical experience was associated with subjects' ability to attend to pitch while timbre varied. Subjects with considerable training (more than 6 years) were able to perform this task, whereas those with little or no training had much more difficulty. One straightforward interpretation of this association is that musical experience enables more selective processing of the different dimensions of musical tones. As always with a correlation, though, other interpretations are possible.

The type of acoustic information that is relied on in making the same–different judgment may account for the performance difference between musicians and nonmusicians. If the fundamental frequency of the two tones were the only pieces

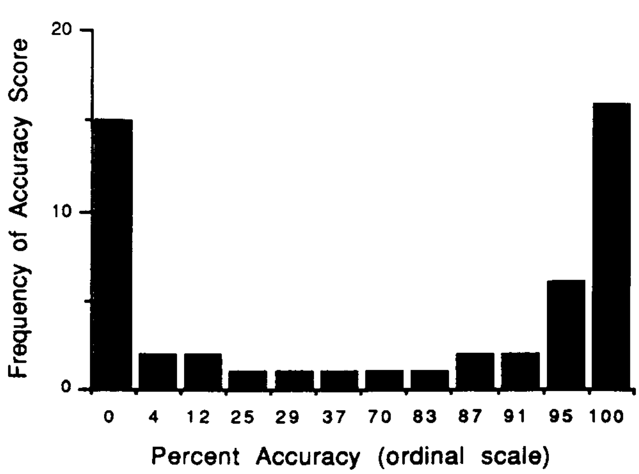

Figure 1. Frequency distribution of percentage accuracy in the same-pitch–different-timbre condition in Experiment 2: Comparison of spectrally varying tones. (Unselected subjects, N=50.)

of information that were used to make the same–different judgment, there should of course have been no difficulty for anyone in performing the task; however, including the harmonics of the tones in the pitch-evaluation process could have hurt performance. Patterson (1989) found that subjects tended to perceive complex tones constructed from a set of upper harmonics (Harmonics 8–24) as being an octave higher than they actually were. One of the timbres used here is similar in harmonic structure to that of Patterson's (that with the fundamental + Harmonics 4–6). This could have resulted in subjects' responding *different* in the same-pitch–"different-timbre" condition because the pitch of one of the timbres would have been perceived an octave higher. If octave misperception is negatively related to musical experience, then this factor may partially account for the performance difference observed between musicians and nonmusicians. Additionally, musicians show greater octave generalization than nonmusicians (Krumhansl & Shepard, 1979), so they might have been able to overlook these octave cues more easily than nonmusicians.

We next concentrated on the data of the 28 subjects who had accuracy scores of 70% or higher in the same-pitch–different-timbre condition. Our aim was to determine for these subjects whether spectral properties are represented in the memory traces of the tones. The RT data from these subjects are presented in the bottom half of Table 2 (Qualified subjects). (The accuracy data are tabulated as well but are not particularly informative here because we used only qualified subjects.) The pattern of RT data in the two same-pitch conditions confirms that same-pitch–same-timbre comparisons were faster than same-pitch–different-timbre comparisons. On average, RTs were 157 ms faster when the timbres were the same than when they were different; this difference was reliable by a test of simple effects, $F(1, 27) = 53.15$, $p < .001$. Therefore, even the qualified subjects were influenced by changes in spectral cues, even though these cues are not strictly relevant to pitch judgments.

Other significant results, although not of theoretical interest, were obtained. A main effect of pitch, $F(1, 27) = 35.58$,

$p < .001$, was a result of responses being faster when the pitches of the tones were the same than when they were different. This effect was qualified by a Pitch × Timbre interaction, $F(1, 27) = 26.90, p < .001$. Effects including both correct *same* and *different* responses are (as we have said) not interpretable.

Experiment 2b: Imagery for Timbres Varying in Spectral Properties

That spectral cues were shown to affect pitch judgments in Experiment 2a permits us to ask whether spectral properties can be part of a generated mental image.

Method

Subjects. Given that not all subjects could attend to pitch while timbre was variable, we restricted the subject pool to the qualified subjects from Experiment 2a who could demonstrably do the task. Fifteen of the original 28 returned to participate in the imagery session; each was paid $5.

Procedure. An additional step was included in the introductory phase of this experiment. Because the two synthetic timbres were not taken from familiar real-world sounds, they were not immediately identifiable. Therefore, at the beginning of each testing session, subjects were introduced to the timbres and asked to label them freely. Two candidate labels, *car horn* and *organ,* were also offered. In most cases, subjects opted for these two descriptors. The rest of the experiment proceeded as described. None of the subjects reported difficulty in associating a label with its timbre.

Results and Discussion

The RT and accuracy data are shown in the bottom half of Table 2. Both measures suggest that spectral cues form part of the mental image of a timbre. In the RT data, responses were 25.3 ms faster in the same-pitch–"same-timbre" condition than in the same-pitch–"different-timbre" condition, $F(1, 14) = 5.34, p < .036$. In the accuracy data, subjects were 13.9 percentage points more accurate when the timbres were the same than when they were different, $F(1, 14) = 295.85, p < .001$. All 15 subjects showed the same pattern of results.

Although the omnibus ANOVA produced no other significant results in the RT data, a main effect of pitch, $F(1, 14) = 374.17, p < .001$, and a Pitch × Timbre interaction, $F(1, 14) = 1681.75, p < .0001$, were obtained in the accuracy data. Again, these effects are uniterpretable because different responses were made in the two pitch conditions.

Experiment 3a: Perception of Timbres Varying in Dynamic (Onset) Properties

We examined the representation of dynamic properties in the image of a timbre in Experiment 3a by varying dynamic cues and holding constant the spectral cues. Previous research (Grey, 1977; Miller & Carterette, 1975; Saldanha & Corso, 1964; see Dowling & Harwood, 1986) identified at least two dynamic properties of tones that affect timbre identification and classification: attack rate and changes in attack rates of individual harmonics during note onset. We restricted our investigation to examining attack rate.

Method

Subjects. Fourteen unselected (met no musical-training criteria) introductory psychology students participated as part of a course requirement.

Materials. Two timbres were created by varying the attack parameters of spectrally identical synthesized tones. An "abrupt" timbre was constructed in which peak amplitude was reached within 2 ms of note onset. A "gradual" timbre was created in which peak amplitude was not reached until 220 ms after note onset. The overtone structure of the timbres consisted of the fundamental + Harmonics 1–3 (the car horn timbre from Experiment 2b).

Results and Discussion

Mean RT and percentage accuracy in the pitch–timbre conditions are shown in the top half of Table 3 (unselected subjects) as before, but also as a function of the nature of the probe tone (the second note of the pair). The data were broken down by this variable to tease apart differences in performance that could result from whether the second timbre was abrupt or gradual. After all, the extraction of pitch information from an abrupt tone can be accomplished fairly quickly upon note onset. This same process may take more time when the second timbre has a gradual onset because pitch information could not be easily extracted from the tone until it rose significantly above threshold.

Comparison of accuracy in the abrupt–second versus the gradual–second conditions confirm our suspicions regarding pitch extraction. Response accuracy was significantly better in the abrupt–second than in the gradual–second condition, $F(1, 13) = 8.27, p < .01$. A similar pattern of results is evident in the RT data, though in this case the difference was not reliable, $F(1, 13) = 1.40, p = .26$. This finding, however, is of little theoretical interest to the topic under study because it simply reflects sensory limitations in pitch detection.

Examination of performance in the two conditions of interest reveals that changes in the attack characteristics of a note influenced same–different judgments (collapsed over the type of probe note). Accuracy was 16 percentage points higher in the same-pitch–same-timbre condition than in the same-pitch–different-timbre condition, $F(1, 13) = 8.37, p < .01$. A similar pattern of results is found in the RT data, in which response times were 178 ms faster in the same-pitch–same-timbre condition than in the same-pitch–different-timbre condition; this difference, however, was only marginally significant, $F(1, 13) = 4.34, p < .058$. Overall, these results demonstrate that dynamic properties of timbre are stored in memory, at least until the second tone of a pair is presented. The only other reliable effect that was obtained was a Pitch × Timbre crossover interaction in the accuracy data, $F(1, 13) = 11.99, p < .004$.

Experiment 3b: Imagery for Timbres Varying in Dynamic Properties

Having found that dynamic properties form part of the memory trace of a tone that was presented moments earlier, we went on to assess whether attack-rate information can be

Table 3
Means and Standard Deviations (in Milliseconds and Percentage Correct) for the Perception (Experiment 3a) and Imagery (Experiment 3b) Sessions With Timbres Varying in Dynamic Properties

	Perception				Imagery			
	Same pitch		Different pitch		Same pitch		Different pitch	
Subjects	Same timbre	Different timbre	Same timbre	Different timbre	"Same timbre"	"Different timbre"	"Same timbre"	"Different timbre"
Unselected								
Abrupt–second								
Reaction time	856	1,037	928	900	1,025	1,130	1,190	1,205
SD	238	474	377	278	412	440	445	427
Accuracy (in %)	96	85	85	95	76	75	68	69
SD	5	24	22	6	26	28	28	33
Gradual–second								
Reaction time	971	1,146	1,425	869	1,198	1,025	1,128	1,162
SD	350	626	1,593	265	425	320	334	423
Accuracy (in %)	94	73	77	91	74	77	67	70
SD	7	30	25	11	27	24	31	29
Musicians								
Abrupt–second								
Reaction time	691	792	699	720	1,008	1,030	1,057	1,012
SD	135	173	99	149	180	215	258	218
Accuracy (in %)	95	94	88	85	99	98	98	96
SD	14	19	31	30	2	3	3	7
Gradual–second								
Reaction time	720	780	687	755	1,050	1,064	1,106	1,007
SD	163	196	124	183	235	234	216	206
Accuracy (in %)	98	94	88	85	96	98	99	99
SD	7	10	31	34	9	3	2	4

Note. Quotation marks indicate that the timbre of the first note was imagined.

internally generated by having subjects imagine an abrupt or gradual timbre at a specified pitch.

Method

Subjects. Fourteen subjects were drawn from the same pool as those in Experiment 3a. None participated in the preceding experiment.

Procedure. Because the two timbres from Experiment 3a were not familiar to subjects, the descriptors *abrupt* and *gradual* were used to indicate which timbre was to be imagined on a given trial. No subject reported having any difficulty in using these labels.

Results and Discussion

Mean RT and percentage accuracy are shown in the top half of Table 3. Unlike in the perception experiment (Experiment 3a), no main effect of probe tone (abrupt or gradual) was evident in either the accuracy or the RT data: $F(1, 13) < 1$ in both cases. Also unlike in Experiment 3a, no evidence was obtained for a matching effect, which would have suggested that dynamic cues to timbre can be imagined. In the RT data, we found the expected pattern of results in the abrupt–second trials, with faster RTs in the same-pitch–"same-timbre" condition than in the same-pitch–"different-timbre" condition. We obtained the opposite pattern of data in the gradual–second condition, however, yielding a nonsignificant result overall, $F(1, 13) < 1$. Statistical analyses of the accuracy data also yielded a nonsignificant effect, with performance being virtually identical in the same-pitch–"same-timbre" and same-pitch–"different-timbre" conditions, $F(1, 13) < 1$.

Two other effects were reliable. A main effect of pitch was obtained in the accuracy data, $F(1, 13) = 6.09$, $p < .03$, and a three-way Probe Tone × Pitch × Timbre interaction was obtained in the RT analysis, $F(1, 13) = 8.16$, $p < .01$. The interaction fits with the null results reported earlier if we consider the onset characteristics of the first tone presented on a trial, the sine-wave tone. The onset of the sine-wave tone was abrupt. Because of this, on one half of the trials subjects were asked to imagine an onset (abrupt) of which they had just received a token. In the same-pitch conditions, RTs were faster when the imagined timbre matched that of the sine wave (abrupt–"abrupt": 1,025 and 1,025 ms) than when it did not (abrupt–"gradual": 1,130 and 1,198 ms). (The term in quotation marks refers to what subjects were instructed to imagine.) This finding suggests that imagining a timbre opposite in dynamic value to what has just been presented requires considerable effort, which manifests itself in the form of slower RTs to the following probe tone (note that this result holds regardless of whether the probe tone was abrupt or gradual). The failure to find a similar pattern of results in the different-pitch condition appears to be the cause of the triple interaction.

The results of Experiments 3a and 3b suggest that dynamic properties, specifically attack rates, might not form part of the mental image of a timbre. Drawing such a conclusion from a negative finding is warranted under the current circumstances: The same stimuli were used in both parts of the experiment. When the first timbre was presented auditorily (Experiment 3a), evidence was obtained indicating that dynamic properties form part of the momentary memory trace of a timbre. Only when these properties had to be internally generated, top-down, was no evidence found to suggest that dynamic cues are represented in memory. Statistical evidence supporting this conclusion is provided by the results of an analysis performed on the data from the two experiments. We ran a 2 × 2 ANOVA on the same-pitch RT data (same and different timbre conditions collapsed over probe tone) across the two experiments. The Experiment × Timbre interaction was reliable, $F(1, 26) = 4.20$, $p < .05$, indicating that same responses were differentially affected by timbre in the two experiments. The different-timbre condition had a large effect on RTs in Experiment 3a (178-ms difference between different-timbre and same-timbre conditions) but a much smaller one in Experiment 3b (−34 ms). Thus, it seems safe to conclude that an imagined timbre is far richer in spectral properties than in dynamic properties.

With all experimentation on such musical qualities as timbre, nothing is more natural than to wonder whether the results would have been the same for trained musicians as they were for unselected subjects. In connection with Experiment 2a, we have already shed some light on this question. The results of Experiments 3a and 3b make the question even more urgent, however, because the latter study showed that unselected people were incapable of generating a mental image for dynamic timbre cues (Experiment 3b), even though the perceptual residue from hearing such cues could be used in comparing one tone with another (Experiment 3a). Perhaps subtler musical cognition on the part of trained subjects would allow such generation. Beethoven's case may not demonstrate a general human capacity for musical imagery. In other words, it may speak more to the mastery of one very highly gifted individual in this respect.

Accordingly, we repeated Experiments 3a and 3b with musically trained individuals. Our subjects were 12 students from the Ohio State University School of Music, all of whom had had at least 5 years of formal training on a single instrument (median number of years = 8.5). All subjects had had additional training through course work and the regular music-school curriculum.

The methods were identical to those of Experiments 3a and 3b. Subjects participated in both experiments, first 3a and then 3b.

The results are shown in the bottom half of Table 3. Consider the perception (Experiment 3a) data first: As we would have expected from Experiment 2, accuracy scores in the same-pitch–different-timbre condition were higher for musicians than for unselected subjects (94% and 79%, respectively). Moreover, RTs were prompter for the trained subjects in the perception experiment. The main result was that the effect of timbre on same-pitch judgments was as robust with the musicians, $F(1, 10) = 15.10$, $p < .003$, as it had been for untrained subjects.[2]

This was the precondition for examination of the imagery data (Experiment 3b). For these data, as in perception, accuracy scores were much higher for the trained musicians than for the unselected subjects, as we would expect. Similarly, musicians' RTs were slightly faster than nonmusicians' RTs. The main question was whether the new subjects would be able to generate mental images sufficient to show the same advantage of matching timbres as they did in perception. The answer is, They could not. Neither the 22-ms difference in the abrupt–second data nor the 14-ms difference in the gradual–second data was close to statistical reliability ($p > .30$ in both cases). In this experiment, timbre seemed to make a difference in different judgments, with conflicting timbre and pitch cues (different-pitch–"same-timbre") producing slower RTs than nonconflicting cues (different-pitch–"different-timbre"), $F(1, 11) = 12.33$, $p < .005$. This difference, however, appeared neither in the imagery data for unselected subjects nor in the musicians' perceptual data, so we are not inclined to stress it. The main conclusion is that an inability to generate images of dynamic properties of timbre appears now to be a limitation of human beings in general and not just of unselected subjects.

The final experiment extends the finding that the attack rate of notes cannot be imagined. The difference between two timbres that vary only in attack rate is in their loudnesses at any point during note onset. That is, one timbre will reach the steady-state portion of the note before the other, resulting in a local difference in loudness between the two timbres. If a difference in attack rate can be considered a locally defined difference in loudness, then the results of Experiment 3 may be indicative of a much wider phenomenon: Loudness information cannot be generated in auditory images.

To examine this possibility, we used the two-part procedure of the previous experiments and varied the amplitudes (instead of the timbres) of the pitches that were presented. If attack rate is tantamount to loudness, then these data should mirror those of Experiment 3: Changes in loudness should affect pitch judgments in Experiment 4a but not in Experiment 4b.

Experiment 4a: Perception of Tones Varying in Loudness

Method

Subjects. Twelve undergraduates from an introductory psychology course participated in exchange for course credit.

Materials. The three pitches (F4, G4, and A4) were recorded twice, once at a low intensity (≅70 dBA SPL) and once at a moderate intensity (≅82 dBA SPL). The intensity level at which the stimuli were presented varied slightly across subjects because of differences in preferred listening level; however, the difference in loudness between the loud and soft tones, the manipulation of interest, would not have changed as a result of these fluctuations in intensity level. The notes were recorded with a single timbre (the abrupt timbre from Experiment 3) to hold constant the spectral and dynamic properties of the stimuli.

[2] A portion of 1 subject's data was lost because of equipment failure. The discrepancy in degrees of freedom (10 instead of 11) is due to the removal of this subject's data from the analysis.

Results and Discussion

The data were broken down and analyzed in the same manner as in the preceding experiment, except that the two variables involving timbre were now loudness variables (loudness [same–different] and probe tone [loud–soft]). The results are shown in Table 4. Inspection of the RT data in the same-pitch conditions reveals that variations in loudness influenced same–different judgments. On average, RTs were 291 ms faster when the loudnesses were the same than when they were different, $F(1, 11) = 14.16$, $p < .003$. Although the accuracy data show a similar trend, with performance being 9 percentage points higher in the same-pitch–same-loudness condition than in the same-pitch–different-loudness condition, the difference was not statistically reliable, $F(1, 11) = 2.31$, $p < .16$. The RT data demonstrate clearly that the momentary residue of a tone presented auditorily retains loudness information.

Other effects of less interest emerged from the analyses. A main effect of loudness showed same-loudness responses to be reliably different from different-loudness responses, $F(1, 11) = 5.10$, $p < .05$. This effect was qualified by the now-familiar Pitch × Loudness (previously Pitch × Timbre) interaction, $F(1, 11) = 5.24$, $p < .04$. The omnibus ANOVA performed on the accuracy data yielded no significant results.

Experiment 4b: Imagery for Tones Varying in Loudness

Method

Subjects. Twenty students were drawn from the subject population that was used in Experiment 4a. None participated in the preceding experiment.

Materials. The three sine-wave tones were redigitized at a loudness level that was judged by two listeners to be close to that of the soft tones used in Experiment 4a. Equating the loudness level of the sine wave with that of one of the loudnesses anticipates the analysis used in the previous experiment, wherein data were inspected separately for the two versions of the probe (second) tone. We were again prepared to see whether the task of imagining a tone whose loudness properties conflicted with the sine wave just presented was a limiting factor in responding.

Procedure. The labels *loud* and *soft* were used to indicate to the subjects whether to imagine the sine wave to be loud or soft. A training phase was initiated at the beginning of each testing session to provide subjects with practice in imagining loud and soft tones. Otherwise all details matched those of comparable earlier experiments in the series.

Results and Discussion

Mean RT and percentage accuracy are shown in Table 4. The similar pattern of performance observed in the loud-second and soft-second conditions when the pitches were the same suggests that the loudness of the sine-wave tone had a similar impact on performance in the two probe-tone conditions.

Neither the RT nor the accuracy data provide evidence that loudness information is represented in the generated image of a tone. Although a trend in this direction is visible in the data, with RTs being 41 ms faster in the same-pitch–"same-loudness" condition than in the same-pitch–"different-loudness" condition, the means were not reliably different from one another, $F(1, 19) < 1$. No hint of such an effect was evident in the accuracy data, as the means were virtually identical in the two loudness conditions in which the pitches were the same ("same loudness" = 88%; "different loudness" = 87%), $F(1, 19) < 1$. These results suggest that imagery for loudness is vanishingly weak if it exists at all.

The only statistically reliable result was a Probe Tone × Pitch interaction obtained in both the RT and the accuracy data, $F(1, 19) = 5.56$, $p < .03$, and $F(1, 19) = 5.97$, $p < .03$, respectively. When the probe tone was loud, performance was

Table 4
Means and Standard Deviations (in Milliseconds and Percentage Correct) for the Perception (Experiment 3a) and Imagery (Experiment 3b) Sessions With Tones Varying in Loudness

	Perception				Imagery			
	Same pitch		Different pitch		Same pitch		Different pitch	
Result	Same loudness	Different loudness	Same loudness	Different loudness	"Same loudness"	"Different loudness"	"Same loudness"	"Different loudness"
Loud–second								
Reaction time	813	1,043	1,079	1,066	837	902	1,006	865
SD	196	393	461	366	218	303	560	273
Accuracy (in %)	92	87	91	90	92	89	80	86
SD	19	24	14	18	11	21	26	25
Soft–second								
Reaction time	847	1,199	972	1,066	932	949	842	890
SD	271	532	311	324	329	302	250	311
Accuracy (in %)	94	81	91	89	84	84	90	89
SD	16	21	17	21	26	25	16	19

Note. Quotation marks indicate that the loudness of the first note was imagined.

best when the pitches were the same. The reverse occurred when the probe tone was soft.

Our claim here that loudness cannot be imagined is again based on a null result. As mentioned previously, this conclusion is justified by the experimental procedure. The results of Experiment 4a demonstrated that loudness persists long enough to affect performance when the tones are presented auditorily. The difficulty, therefore, lies in the generation of loudness as a top-down process. The negative finding should not be attributed to weak statistical power, for 20 subjects were tested in Experiment 4b as opposed to 12–14 in most of the earlier studies in the series. A comparison of the data from Experiments 4a and 4b adds additional support to our claim. As in Experiment 3, we ran a 2 × 2 ANOVA on the same-pitch (same and different loudness) RT data across the two experiments. The reliable Experiment × Loudness interaction, $F(1, 31) = 5.45$, $p < .03$, indicates that the effect of loudness on same-pitch judgments in Experiment 4a was significantly greater than that in Experiment 4b. (The RT differences between the different-loudness and same-loudness conditions were 291 ms and 41 ms, respectively.)

Conclusions

The goal of this project was to determine the acoustical basis for mental imagery of timbre. We used an experimental paradigm in which imagery is inferred from whether performance with two audible stimuli mirrors that when the first of the pair must be imagined. We found that the image of a timbre appears to be based primarily on spectral properties. No evidence was obtained which suggested that dynamic properties, at least attack rate, are represented in an auditory image.

However, before it can be concluded categorically that dynamic properties are never part of an auditory image, other types of dynamic cues (and their combinations) should be examined. Only one type of dynamic cue was manipulated here. The ability to imagine other factors such as spectral changes during note onset and rate of note offset should be examined.

Dynamic cues to timbre may exist, but their representation may be different from that of spectral cues. Spectral cues are for the most part time invariant, whereas dynamic cues to timbre evolve over time. This temporal property of dynamic cues may form part of the representation of a timbre, resulting in an image that unfolds over time. Our failure to find evidence of imagery for dynamic cues may be due to the temporal constraints of the experimental paradigm. In Part B of the procedure, subjects had to imagine a specified timbre and then make a speeded response to the second tone. Because image generation was self-paced, and there was a 1,500-ms pause prior to the second tone, subjects may have finished imagining the onset of the note by the time the probe tone was presented. Furthermore, requiring subjects to make a speeded response may have allowed subjects to make only a partial comparison of the imagined and presented timbres. An experimental paradigm that is more sensitive to the time-varying properties of timbre might be better suited for studying imagery for dynamic cues.

The fact that Beethoven was able to compose his Ninth Symphony suggests that he must have been able to imagine differences in loudness. That we failed to demonstrate imagery for loudness was surprising to us at first. Who would have thought that such a salient perceptual dimension could *not* be imagined! The results of Experiment 4 allow us to conclude that loudness is probably not coded in the auditory image of a timbre, not that it is not coded in memory. Loudness may be represented in another form. For example, because variations in intensity are produced on a natural instrument by changing the physical force used to sound it, loudness may be coded in imagery in motor form.

As discussed earlier, the failure to find imagery for the attack rate of a note may be a result of the more general inability to imagine loudness. If this is the case, then the results of Experiments 3b (imagery for dynamic cues) and 4b (imagery for loudness) strongly suggest that loudness cannot be imagined auditorily. Not 1 subject, however, mentioned having difficulty in imagining abrupt or gradual tones, or loud or soft tones. They were uniformly confident that they were following instructions to do so. Clearly, further research is necessary to resolve these issues.

In closing, this study sheds some light on the nature of imagery for timbre and auditory imagery in general. Currently, we are exploring the cause of the individual differences observed in Experiment 2 (where nonmusicians were not able to dissociate pitch from timbre) and intend to examine systematically imagery for other dynamic properties of timbre.

References

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

Cutting, J. E., Rosner, B. S., & Foard, C. F. (1976). Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology, 28,* 361–378.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition.* New York: Academic Press.

Farah, M. J., & Smith, A. F. (1983). Perceptual interference and facilitation with auditory imagery. *Perception & Psychophysics, 33,* 475–478.

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

Halpern, A. R. (1988). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 434–443.

Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition, 17,* 572–581.

Hubbard, T. L., & Stoeckig, K. (1988). Musical imagery: Generation of tones and chords. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 656–667.

Krumhansl, C. L., & Shepard, R. N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 579–594.

Miller, J. R., & Carterette, E. C. (1975). Perceptual space for musical structures. *Journal of the Acoustical Society of America, 58,* 711–720.

Paivio, A. (1979). *Imagery and verbal processes.* Hillsdale, NJ: Erlbaum.

Patterson, R. D. (1989). The tone height of multi-harmonic tones. *Proceedings of the First International Conference on Music Perception and Cognition* (pp. 119–124). Kyoto, Japan.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. (1969). Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monographs, 79*(1, Pt. 2).

Rosen, S. M., & Howell, P. (1981). Plucks and bows are not categorically perceived. *Perception & Psychophysics, 30,* 156–168.

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

Seashore, C. E. (1967). *Psychology of music.* New York: Dover. (Original work published 1938)

Segal, S. J., & Fusella, V. (1970). Influence of imagined pictures and sounds on detection of visual and auditory signals. *Journal of Experimental Psychology, 83,* 458–464.

Received August 22, 1990
Revision received August 7, 1991
Accepted August 8, 1991 ■