Isolating the dynamic attributes of musical timbre[a]

Paul Iverson[b] and Carol L. Krumhansl
Department of Psychology, Cornell University, Uris Hall, Ithaca, New York 14853

(Received 28 August 1992; accepted for publication 28 July 1993)

Three experiments examined the dynamic attributes of timbre by evaluating the role of onsets in similarity judgments. In separate experiments, subjects heard complete orchestral instrument tones, the onsets of those tones, and tones with the onsets removed ("remainders"). Ratings for complete tones corresponded to those for onsets, indicating that the salient acoustic attributes for complete tones are present at the onset. Ratings for complete tones also corresponded to those for remainders, indicating that the salient attributes for complete tones are present also in the absence of onsets. Subsequent acoustic analyses demonstrated that this pattern of similarity was due to the centroid frequencies and amplitude envelopes of the tones. The results indicate that the dynamic attributes of timbre are not only present at the onset, but also throughout, and that multiple acoustic attributes may contribute to the same perceptual dimensions.

PACS numbers: 43.66.Jh, 43.75.Cd [WDW]

INTRODUCTION

Timbre is commonly defined as the attribute which allows a listener to discriminate two sounds with the same pitch and loudness. It allows a listener to discriminate between tones played by a trumpet and a violin when they have the same pitch, loudness, and duration. This definition, though, leaves the acoustic basis for timbre almost completely undefined. Any acoustic attribute that does not exclusively contribute to the perception of pitch, loudness, or duration could contribute to the perception of timbre.

Perceptual research has tried to determine which of these possible acoustic attributes are most salient by using similarity scaling techniques. In these experiments, subjects rate on a numerical scale the similarity of all possible pairs of a set of tones. Multidimensional scaling (MDS) (Shepard, 1962a,b; Kruskal, 1964a) converts these similarity judgments to a map of the tones in a low-dimensional geometric space where distances in the space correspond to perceived similarity. This technique reveals relationships between the tones that are difficult to observe directly from the similarity ratings. Researchers examine the MDS map of subjects' judgments and attempt to find acoustic attributes that correspond to relationships between the tones. This technique allows researchers to isolate a small number of acoustic factors that contribute to the perception of timbre.

In all similarity scaling experiments with natural tones, the relative amplitudes of low- and high-frequency harmonics in the static spectrum have contributed to similarity judgments (Grey, 1975, 1977; Grey and Gordon, 1978; Wessel, 1979; Krumhansl, 1989). Grey and Gordon (1978) were successful at quantifying this acoustic attribute. They found that the mean frequency (centroid) of the spectra correlated highly with similarity judgments. Thus the centroid frequency is one of the main contributors to the perception of timbre.

Similarity scaling studies on natural tones have also found dynamic attributes are salient. Wedin and Goude (1972) ran a series of experiments on tape-recorded musical instrument tones. In one experiment, they played subjects complete tones, and in another they removed both onsets (first half second) and decays (last half second) so that most dynamic variation was absent. Although the ratings for the two sets of tones were quite similar, there were some subtle differences. Some tones with similar attacks and dissimilar spectra sounded less similar when the transitions were removed. Some tones with dissimilar attacks and similar spectra sounded more similar when the transitions were removed. Additionally, static-spectral measurements accounted for a greater proportion of variance for edited tones than for complete tones.

Wedin and Goude (1972) did not make measurements of dynamic acoustic attributes, so it is difficult to determine exactly what aspect of the transitions influenced similarity judgments. It is uncertain, for example, whether the differences in judgments were due specifically to the removal of attacks, or if they were due to more global effects of removing temporal variation. Still, the results indicate that dynamic attributes of some sort influenced judgments.

Other researchers have tried to identify specific dynamic attributes. In contrast to centroid frequency, these dynamic attributes have been more variable between studies, and have been considerably harder to quantify. Grey (1975, 1977) found two dynamic attributes that contributed to similarity judgments: the spectral synchrony and fluctuation of the tones (most often at the onsets), and the presence or absence of low-amplitude, high-frequency energy at the onsets. Additionally, Grey (1975; Grey and Moorer, 1977) found that removing the low-amplitude energy at the onsets made the tones sound less realistic. Wessel (1979) found that similarity judgments related to the "bite" of onsets. Krumhansl, Wessel, and Winsberg (reported in Krumhansl, 1989) determined that the rapidity

---

[a] Presented at the 121st Meeting of the Acoustical Society of America [J. Acoust. Soc. Am. **89**, 1988(A) (1991)].

[b] Present address: Department of Speech and Hearing Sciences, WJ-10, University of Washington, Seattle, WA 98195.

of the attack and spectral fluctuation influenced similarity judgments. Although these dynamic attributes have not been consistent or easy to quantify, it seems that the onsets are particularly important.

Experiments on the identification of musical instruments support this conclusion. Clark *et al.* (1963) found that removing steady states and decays did little to impair identification, but removing attacks greatly decreased identification accuracy. Even subjects hearing only the first 60 ms of each tone were able to make relatively accurate identifications. Similarly, Saldanha and Corso (1964) found that recognition of tones with only an attack and an edited steady state was as good as recognition of entire tones; recognition decreased with deleted attacks. Other experiments have supported these conclusions (Berger, 1964; Wedin and Goude, 1972; Elliot, 1975). Onsets seem important for the identification of musical instruments.

Kendall (1986), however, claimed that onsets and decays are not important when tones are present in musical contexts. He examined the recognition of three instruments playing musical passages with legato transitions. Removal of the transitions did not change recognition accuracy, but removal of the time-variant steady states did decrease accuracy. Recognition accuracy also decreased when the time-variant steady states were replaced by static steady states. He concluded that the time-variant steady states were necessary and sufficient for recognition of instruments playing musical sequences. It is difficult, though, to compare the results of this study to previous recognition experiments. The legato transitions may have made the onsets less distinctive, and thus less important for recognition. Also, editing the steady states may have introduced anomalies into the sequences that could have independently decreased accuracy.

The main goal of the experiments reported here is to isolate the dynamic attributes that contribute to timbre. Because onsets have most often been identified as the source for dynamic attributes, the first step was to evaluate what effect onsets have on similarity judgments. In separate experiments, subjects made judgments of complete orchestral tones (experiment 1), onsets of those tones (experiment 2), and of tones with onsets removed (experiment 3). Comparing the results of these experiments tested the influence of onsets on the timbre of complete tones. This comparison guided an investigation of possible attributes which could contribute to perceived similarity.

I. EXPERIMENT 1

This experiment employed the similarity-scaling method used by previous researchers (Grey, 1975, 1977; Grey and Gordon, 1978; Wessel, 1979; Krumhansl, 1989). The results will be used to evaluate the role of onsets through comparisons to subsequent experiments.

The tones were selected from the McGill University Master Samples (MUMS) Library (Opolko and Wapnick, 1989) of digitally recorded musical instruments. Using these tones offered two main advantages. First, they were produced by traditional instruments, so they had the full dynamic complexity of natural tones. Second, these tones are publicly available, so it should be relatively easy for other researchers to evaluate and extend our results.

TABLE I. Instruments, acoustic measurements, and references for the tones used in this study.

Instrument	Edited length (s)	Centroid (Hz)	Time until maximum amplitude (ms)	MUMS reference		
				Volume	Track	Index
Bassoon	2.72	1170	69	2	14	27
Cello	3.20	2853	55	1	11	28
Bb clarinet	2.51	2015	71	2	10	11
English horn	2.88	2180	65	2	09	09
Flute	2.76	1579	74	2	01	01
French horn	2.12	877	61	2	19	23
Muted C trumpet	2.81	7834	69	2	17	07
Oboe	3.29	2258	55	2	08	03
Piano	3.15	1477	31	3	02	40
Tenor saxophone	2.71	1932	56	3	15	13
Tenor trombone	2.28	1232	69	2	22	21
C trumpet	2.73	2374	61	2	16	07
Tuba	2.73	655	67	2	25	24
Tubular bells	2.85	2245	30	3	10	01
Vibraphone	2.00	1569	17	3	06	08
Violin	3.27	2035	64	1	01	06

A. Method

1. Subjects

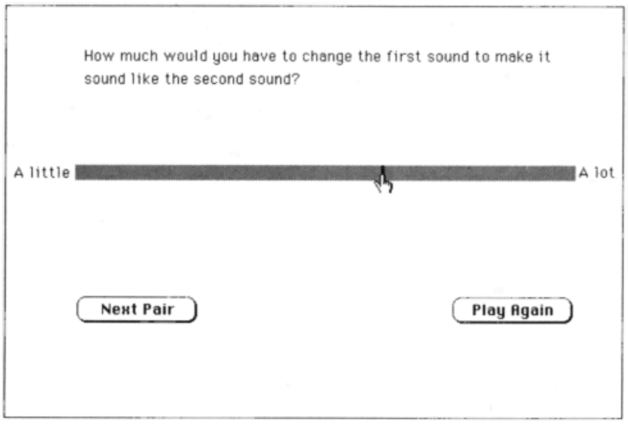
Subjects were nine members of the Cornell University community who participated in the 1-h experiment for course credit. All subjects were amateur musicians.

2. Apparatus

The tones were recorded and reproduced by a DigiDesign Audiomedia digital audio board controlled by a Macintosh IIcx microcomputer. They were played over a single high-quality speaker. Subjects' responses were entered and recorded using the microcomputer that controlled presentation of stimuli.

3. Stimulus materials

The 16 tones used in the experiment were from the MUMS Library (Opolko and Wapnick, 1989), and are listed in Table I. Each tone was digitally sampled monaurally with 44 100 16-bit samples per second. Samples that were longer than 3.5 s were edited by removing portions of the steady states. Care was taken so that the editing process did not introduce transients or other unnatural changes in the tones. Table I lists the edited length of each tone. The frequency of each tone was middle C (262 Hz). The amplitude of each tone was adjusted so that it registered a peak dB meter reading of 70 dBA SPL. Careful listening after the amplitude adjustment determined that the tones were equally loud. Each trial consisted of a pair of tones with an inter-onset interval of 4 s.



How much would you have to change the first sound to make it sound like the second sound?

FIG. 1. Response bar for similarity judgments. Subjects saw this image on a computer screen, and positioned and clicked the appropriate spot on the bar to record their judgments. They clicked on the "Play Again" button to hear the pair again, and clicked on the "Next Pair" button to go on to the next pair.

4. Procedure

Subjects read that they would hear pairs of musical-instrument tones. For each pair of tones, they judged how much they would have to change the first tone to make it sound like the second tone. They were asked to imagine that they had a computer that allowed them to record a sound and change it in any way they wanted. These instructions focused primarily on the similarity of the tones, but were designed to also draw attention to the order of the tones in each pair. Subjects were told that there were no right or wrong answers, and were instructed to base their judgments on their general intuitions rather than on any specific knowledge of instruments or signal processing.

For their judgments, subjects selected the appropriate location on a continuum, shown in Fig. 1, from "a little" to "a lot." A computer screen displayed this continuum, and subjects used a Macintosh mouse to click at the point on the bar that corresponded to their judgment. Their responses could fall on any of the 415 discrete points on this bar. Subjects heard each pair as many times as they needed to make their judgment.

The experiment began with 8 practice pairs of sounds. These pairs were randomly selected for each subject such that each of the 16 tones occurred once during the practice. Subjects were instructed to observe the range of similarity in this set of tones. They were asked to distribute their responses so that they would click near the "a little" end of the bar for the smallest changes (most similar), near the "a lot" end of the bar for the largest changes (least similar), and in the appropriate spots in-between for other size changes. After the practice was over, they completed an experimental session with 240 trials. The trials were composed of each possible pair of tones presented in a different random sequence for each subject. They were asked to keep their judgments as consistent as possible throughout the experimental session.

B. Results and discussion

Each subject's responses were put into the form of a triangular matrix giving the similarity for each of the 120 pairs averaged across order. The order in which the tones were presented was found in preliminary analyses to have little effect. Inter-subject correlations examined consistency of ratings between subjects. Of the 36 inter-subject correlations, the average was $r=0.65$ ($df=118$), and each was significant at the $p<0.001$ level. Thus the ratings were highly consistent between subjects.

The ratings were averaged across subjects, and analyzed using the Kruskal (1964a,b) multidimensional scaling technique implemented in the SYSTAT computer program (Wilkinson, 1987). The MDS analysis used Kruskal's stress formula 1 (Kruskal, 1964a), a Euclidean distance metric, and a monotonic regression function. A two-dimensional solution modeled the responses with a stress of 9.9%. On this measure 0% stress indicates a perfect monotonic relationship between distances in the MDS space and similarity judgments (Kruskal, 1964a). Figure 2 displays the solution. The horizontal dimension tended to

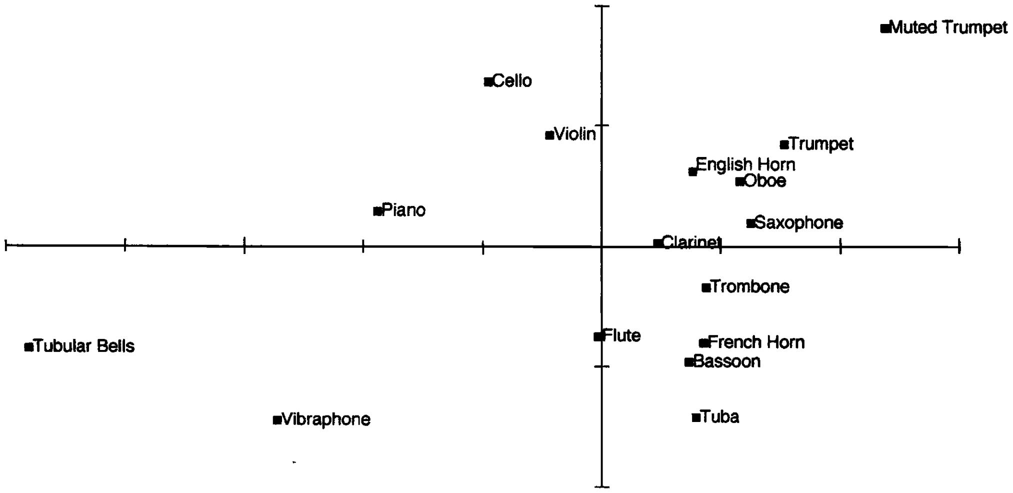


FIG. 2. MDS solution for similarity judgments on complete tones. The horizontal dimension corresponds to dynamic attributes, and the vertical dimension corresponds to static spectral attributes.

separate the percussive instruments, like the tubular bells, from the blown instruments, like the saxophone. This was probably due to dynamic properties of the tones. The vertical dimension tended to separate the brighter tones, like the muted trumpet, from duller tones, like the tuba. This was probably due to static spectra of the tones. The acoustic correlates of these dimensions will be examined later in greater detail.

Even though the tones varied somewhat in duration, as shown in Table I, these variations did not seem to correspond to spatial relationships in the MDS solution. Also, the average similarity ratings were not significantly correlated with the differences in durations of the tones, $r = -0.078$ $(df = 118)$, $p > 0.05$, further indicating that duration did not influence judgments.

II. EXPERIMENT 2

This experiment examined the perception of the onset portions of the tones and compared them to the complete tones. If the similarity of onsets corresponds to the similarity of complete tones, then onsets and complete tones have the same salient attributes. If, however, there is no correspondence, then we can conclude that onsets do not influence the similarity of complete tones.

In this experiment, subjects heard the first 80 ms of the tones in experiment 1. This length of time captured nearly all of the initial fluctuation of the tones. The more percussive tones had attacks shorter than 80 ms, so these edited tones had brief steady states. It was necessary to equalize duration in this manner, because playing attacks alone would have confounded attack rapidity and tone length. Experiments that have evaluated the identification of onsets have edited tones in a similar manner (Clark *et al.*, 1963; Saldanha and Corso, 1964). As in experiment 1, subjects rated the similarity of each pair of tones.

A. Method

1. Subjects

Subjects were ten members of the Cornell University community who were each paid $4.00 for participating in the 1-h experiment. They were all amateur musicians.

2. Apparatus

Same as in experiment 1.

3. Stimulus materials

Each tone consisted of the first 80 ms of a tone used in experiment 1, followed by a 10-ms linear decay from full amplitude to silence. This decay ramp prevented transient noises resulting from the editing process. The amplitude of each tone was adjusted to 70 dBA SPL, following the procedure used in experiment 1.

4. Procedure

Same as in experiment 1.

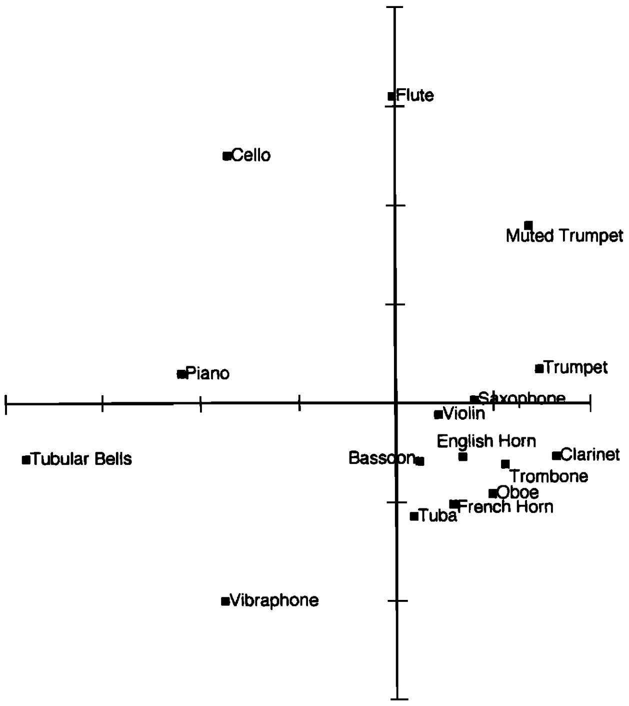


FIG. 3. MDS solution for similarity judgments on onsets. The horizontal dimension corresponds to dynamic attributes, and the vertical dimension corresponds to static spectral attributes.

B. Results and discussion

Each subject's responses were put into the form of a triangular matrix as in experiment 1. Inter-subject correlations showed a high degree of consistency between subjects. Of the 45 inter-subject correlations, the average was $r = 0.63$ $(df = 118)$, and each was significant at the $p < 0.001$ level.

The ratings were averaged across subjects, and analyzed using the multidimensional scaling method used in experiment 1. A two-dimensional solution modeled the responses with a stress of 15.5%. Figure 3 displays the solution. Even though subjects only heard onsets, the solution was very similar to that of complete tones. The horizontal dimension tended to separate the percussive instruments from the blown instruments, and the vertical dimension tended to separate the brighter instruments from the duller instruments. Although the exact location of each tone is not the same as in experiment 1, the general patterns of the two solutions correspond. This indicates that the perception of onsets and complete tones depends on similar attributes.

Correlations between the experiments support this conclusion. The triangular matrix from experiment 1 correlated $r = 0.74$, $p < 0.001$ $(df = 118)$ with that from experiment 2. This correlation was made on the mean subject ratings rather than on the computed MDS distances, so it is independent of the MDS algorithm.

Given the similarity of experiments 1 and 2, it can be concluded that onsets contain many of the same salient acoustic attributes as complete tones. One possible explanation for this is that listeners attend only to onsets when making similarity judgments on complete tones. Before concluding this, it is necessary to examine what effect removing the onsets has on similarity judgments. Only if removing the onsets significantly changes judgments can we conclude that listeners are basing their judgments of complete tones solely on the onsets.

III. EXPERIMENT 3

This experiment assessed the perceived similarity of tones with the onsets removed ("remainders"). If similarity judgments of remainders correspond to similarity judgments of complete tones, then remainders contain the same salient attributes as complete tones. If there is no correspondence, then we can conclude that remainders do not greatly affect the perception of complete tones, and that the relationship between experiments 1 and 2 depends on attributes present only at the onsets.

In this experiment, subjects heard the tones used in experiment 1 with the first 80 ms removed. As in experiments 1 and 2, subjects rated the similarity of each pair of tones.

A. Method

1. Subjects

Subjects were nine members of the Cornell University community who were each paid $4.00 for participating in the 1-h experiment. They were all amateur musicians.

2. Apparatus

Same as in experiments 1 and 2.

3. Stimulus materials

The tones were edited versions of those used in experiment 1. Each tone had the first 80 ms deleted, and began with a 10-ms linear onset ramp from silence to full amplitude. This onset ramp prevented transient noises resulting from the editing process. The amplitude of each tone was adjusted to 70 dBA SPL, following the procedure used in experiment 1.

4. Procedure

Same as in experiments 1 and 2.

B. Results and discussion

Each subject's responses were put into the form of a triangular matrix as in experiments 1 and 2. Inter-subject correlations showed high consistency between subjects. For the 36 inter-subject correlations, the average was $r=0.64$ $(df=118)$, and each was significant at the $p<0.001$ level.

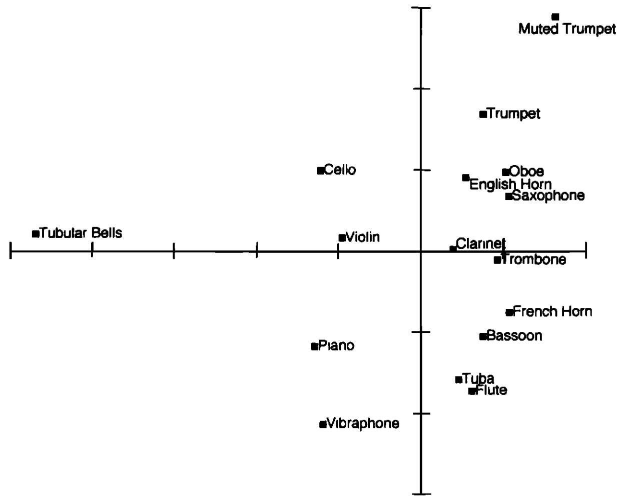


FIG. 4. MDS solution for similarity judgments on remainders. The horizontal dimension corresponds to dynamic attributes, and the vertical dimension corresponds to static spectral attributes.

The ratings were averaged across subjects, and analyzed using the multidimensional scaling technique used in experiments 1 and 2. A two-dimensional solution modeled the responses with a stress of 13.4%. Figure 4 displays this solution. Even though the onsets were removed, the results were very similar to those for both complete tones and onsets. The horizontal dimension tended to separate the percussive tones from the blown instruments. The vertical dimension tended to separate bright instruments from dull instruments. Correlations also demonstrated this correspondence between the results of the experiments. The mean triangular matrix from experiment 3 correlated $r=0.92$ $(df=118)$. $p<0.001$ with experiment 1, and $r=0.75$ $(df=118), p<0.001$ with experiment 2. Thus ratings on remainders correspond quite closely with those on complete tones and onsets.

A multiple correlation was run with the results of experiment 1 as the dependent variable and the results of experiments 2 and 3 as independent variables. This yielded a multiple $R=0.924$. Both judgments on remainders $[t(117)=15.488, p<0.001]$ and on onsets $[t(117)=2.389, p<0.05]$ significantly contributed to the model. Although judgments on onsets and remainders both correspond to judgments on complete tones, judgments on remainders account for more variance.

As in experiment 1, the average similarity judgments were not significantly correlated with the differences in tone durations, $r=0.032$ $(df=118)$, $p>0.05$, indicating that duration differences did not influence judgments.

In summary, removing the onsets did not greatly change similarity judgments, and removing the remainders also had little effect. Thus, the salient attributes for complete tones can not be isolated to either onsets or remainders. This constrains the problem of finding acoustic attributes that contribute to perceived similarity.

IV. ACOUSTIC ANALYSES

Two types of attributes can account for the pattern of results found in this study. First, acoustic attributes that are relatively constant throughout the tones could potentially account for similarity. For example, the spectra may be similar at the onsets and remainders. Second, multiple acoustic attributes could contribute to a single dimension of similarity judgments. For example, dynamic attributes of the onsets and remainders could have the same relative

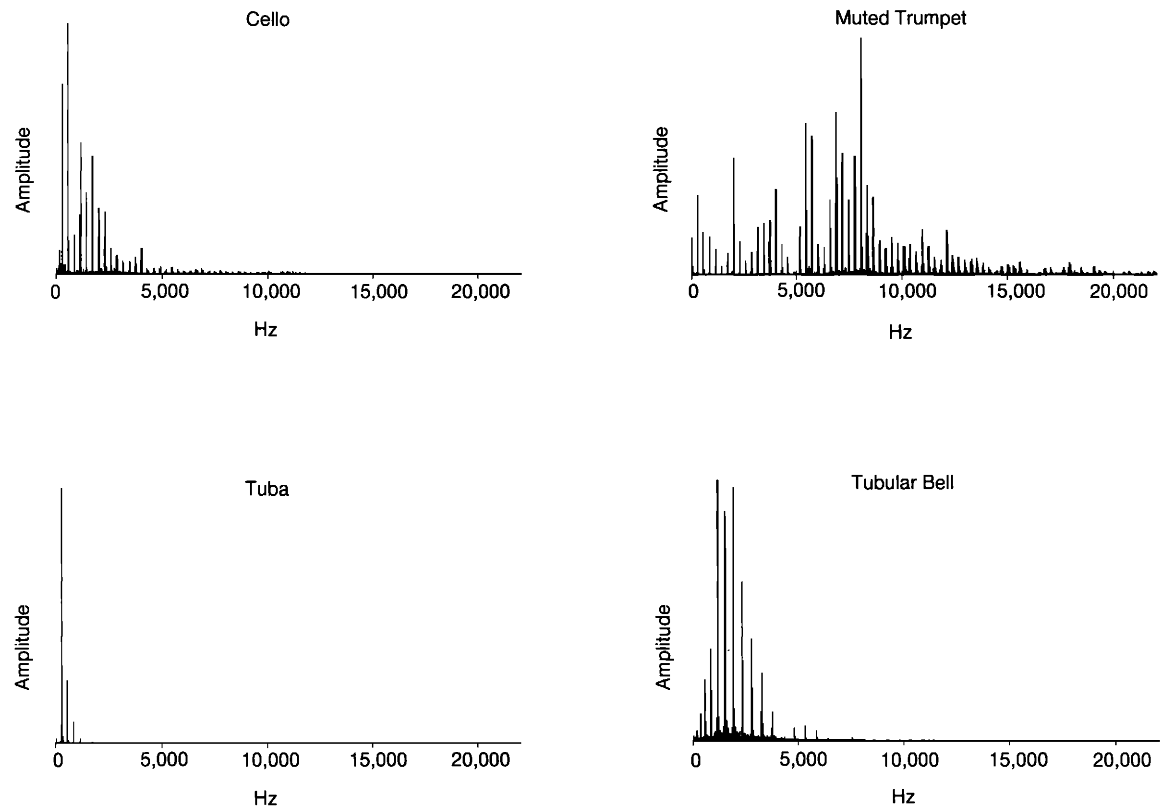


FIG. 5. Frequency spectra for selected tones. The horizontal dimensions correspond to frequency, and the vertical dimensions correspond to amplitude (in arbitrary linear units). The muted trumpet tone has unusually strong high-frequency harmonics (high-centroid frequency), and the tuba has quite weak high-frequency harmonics (low-centroid frequency).

effect on similarity, although these attributes could be acoustically different. An examination of the acoustics of the tones should help identify which types of attributes are salient.

Interpreting the "brightness" dimension of the similarity judgments was relatively straightforward. Previous studies (Grey, 1975, 1977; Grey and Gordon, 1978; Wessel, 1979; Krumhansl, 1989) have demonstrated that the centroid frequency corresponds to one dimension of timbre judgments. A Fourier transform (Press *et al.*, 1988) was performed on each entire tone, and frequency centroids were calculated on the linear frequency and amplitude values. It was calculated by summing the products of the amplitude and frequency values for each component, and dividing this by the sum of the amplitudes. Table I lists centroid frequencies, and Fig. 5 displays spectra for example tones.

To examine the effect of centroid on the experimental results, the centroid frequencies were correlated with the vertical dimension of each MDS solution. Additionally, triangular matrixes were formed by calculating the difference in centroid frequencies for each pair of tones, and these matrixes were correlated with the raw similarity judgments. The centroid frequency for complete tones correlated $r=-0.70$ $(df=14)$, $p<0.01$ with the vertical dimension of the MDS solution, and the differences in centroid frequency correlated $r=0.23$ $(df=118)$, $p<0.05$ with the raw similarity judgments. The centroid frequency for onsets correlated $r=-0.61$ $(df=14)$, $p<0.05$ with the vertical dimension of the MDS solution, and the differences in centroid frequency correlated $r=0.25$ $(df=118)$, $p<0.05$ with similarity judgments. The centroid frequency for remainders correlated $r=-0.75$ $(df=14)$, $p<0.001$ with the vertical dimension of the MDS solution, and the differences in centroid frequency correlated $r=0.34$ $(df=118)$, $p<0.001$ with similarity judgments. Centroid frequency contributed to judgments in each experiment.

Centroid frequency was similar for the complete tones, onsets, and remainders. The centroids of the complete tones correlated $r=0.95$ $(df=14)$, $p<0.001$ and $r=1.00$ $(df=14)$, $p<0.001$ with that of the onsets and remainders, respectively. Additionally, the centroids of the onsets and remainders correlated $r=0.95$ $(df=14)$, $p<0.001$. This demonstrates that centroid frequency is relatively constant throughout each tone. Since this attribute is salient for similarity judgments, it is not surprising that ratings for complete tones, onsets, and remainders are similar.

The other dimension of the similarity matrixes was somewhat harder to identify, but it seemed to correspond to how the amplitude of each tone changes over time. An amplitude envelope was calculated for each tone by taking the rms of the original wave form and the Hilbert transform of that wave form (Bracewell, 1965), and low-pass filtering it at 20 Hz. This method calculates the amplitude at each point in the sound with better temporal resolution than other techniques. Figure 6 displays the amplitude envelopes for example tones.

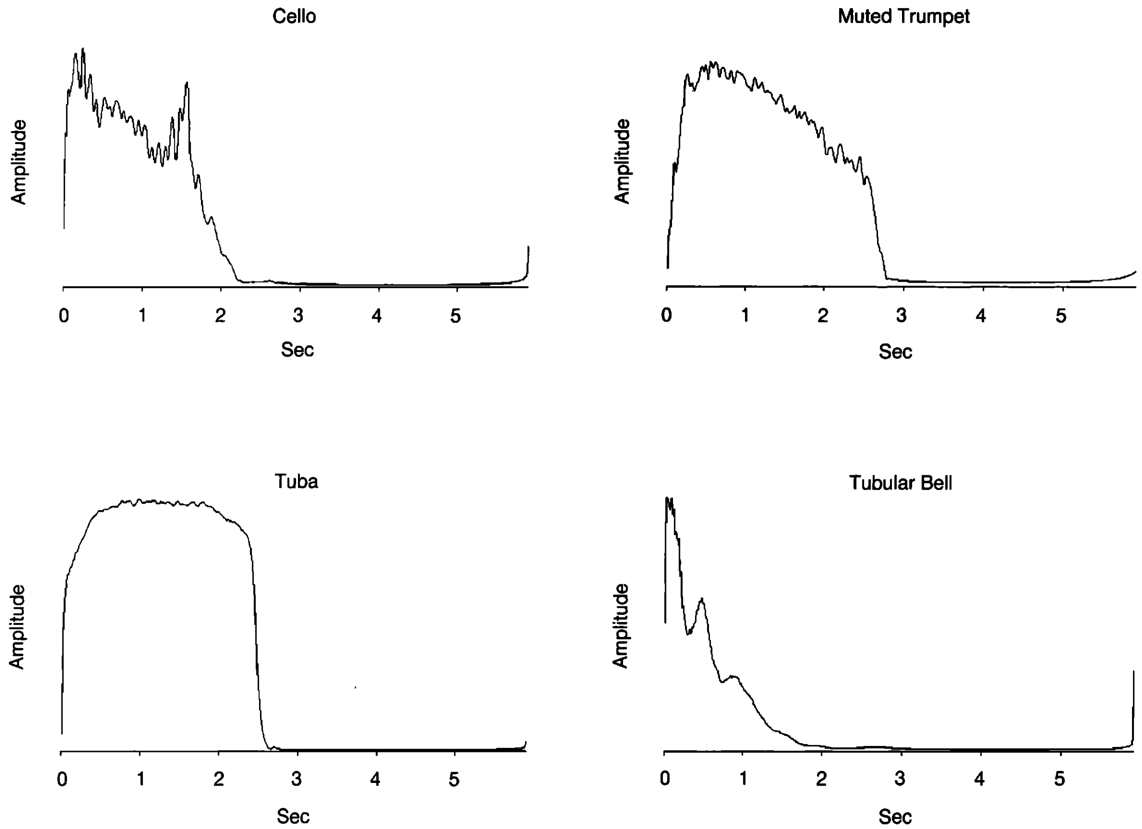


FIG. 6. Amplitude envelopes for selected tones. The horizontal dimensions correspond to time, and the vertical dimensions correspond to amplitude (in arbitrary linear units). The tubular bell has a rapid attack and a gradual decay, while the tuba has a more gradual attack, a relatively constant steady state, and a rapid decay.

Quantifying the salient aspects of the amplitude envelope presented a problem, because no single attribute could account for similarity judgments. Onsets have very different dynamic qualities than remainders. For onsets, tones to the left of the dimension increased in amplitude more rapidly than those to the right. For the remainders, tones to the left of the dimension fell in amplitude gradually, and tones to the right of the dimension stayed relatively steady, then had a rapid decay. Because a single attribute could not account for judgments, multiple attributes were sought.

A Euclidean distance metric measured the similarity of each pair of amplitude envelopes for complete tones. This metric calculated the rms of the amplitude differences between envelopes. The complete tones were of somewhat different lengths, so only the middle 1.5 s of the envelopes were compared. This measurement correlated $r=0.76$ $(df=118)$, $p<0.001$ with similarity judgments on complete tones, and $r=0.56$ $(df=118)$, $p<0.001$ with those on remainders. We tried various simplifying formulations, but failed to adequately isolate a single aspect of the amplitude envelopes that most contributed to similarity. It seems that general differences in amplitude envelopes contribute to similarity judgments.

We also calculated the Euclidean distance between amplitude envelopes of each pair of onsets. This correlated $r=0.47$ $(df=118)$, $p<0.001$ with similarity judgments. We isolated the time until maximum amplitude as the factor most contributing to this correlation, and these values are listed in Table I. The time from the start of each tone to its maximum amplitude level correlated $r=0.79$ $(df=14)$, $p<0.001$ with the dynamic dimension of the MDS solution. Calculating the difference between these times for each pair of tones formed a triangular matrix, and this correlated $r=0.55$ $(df=118)$, $p<0.001$ with the similarity judgments of experiment 2. The rapidity of the attacks appeared to be most salient for similarity judgments of the onset portions.

Although no single attribute can account for the dynamic dimension of experiments 1, 2, and 3, the general differences in amplitude envelopes corresponded to similarity judgments. The Euclidean distance between amplitude envelopes of complete tones correlated $r=0.806$ $(df=118)$, $p<0.001$ with differences in the time until maximum amplitude of the onsets, demonstrating that tones with similar attacks tend to have similar amplitude envelopes for the rest of the tones. Even though onsets have very different dynamic properties than remainders, the relative similarities of their amplitude envelopes remain the same.

Multiple regression analyses further compared the measurements of acoustic attributes to similarity judgments. For similarity judgments of complete tones, differences in centroid frequencies $[t(117)=-6.246,\ p<0.001]$ and Euclidean distance between amplitude envelopes $[t(117)=-15.118,\ p<0.001]$ each significantly contributed to judgments, and had a multiple $R=0.824$ [$F(2,117)=124.015$, $p<0.001$]. For judgments on onsets, differences in centroid frequencies [$t(117)=-4.752$, $p<0.001$] and differences in the time until maximum amplitude [$t(117)=-8.399$, $p<0.001$] each significantly contributed to judgments, and had a multiple $R=0.645$ [$F(2, 117)=41.710$, $p<0.001$]. For judgments of remainders, differences in centroid frequencies [$t(117)=-6.471$, $p<0.001$] and Euclidean distance between amplitude envelopes [$t(117)=-10.230$, $p<0.001$] each significantly contributed to judgments, and had a multiple $R=0.724$ [$F(2,117)=66.476$, $p<0.001$].

These acoustic measurements accounted for a substantial portion of the variance in similarity judgments, although some of the variance was unexplained. An improved fit to the data might have been obtained if the measurements were based on psychological measures of spectral and amplitude discrimination. Also, additional acoustic attributes, which we did not quantify, could have made some small contribution to judgments. Last, some of the unexplained variance may have been due to random variation of responses. In any case, the acoustic attributes we quantified accounted quite well for the judgments.

The acoustic attributes were similar to those found in previous studies employing similarity judgments. Centroid frequency has consistently been found to influence similarity judgments in previous studies (Grey, 1975, 1977; Grey and Gordon, 1978; Wessel, 1979; Krumhansl, 1989), and it was also influential in this study. Studies (Wessel, 1979; Krumhansl, 1989) have found that attack rapidity influences judgments. Our dynamic dimension was similarly related to attack rapidity, although the Euclidean distances between amplitude envelopes were also related to this dimension.

We did not find, however, an influence of spectral fluctuation or high-frequency onset energy like that found by Grey (1975, 1977). Due to the inclusion of percussive tones, our set of tones had a greater range of amplitude envelope differences than the tones used by Grey (1975, 1977). It is possible that this increased the salience of amplitude envelopes, and this may have obscured more subtle dynamic attributes. Additional acoustic attributes may be more salient in sets of tones with more homogeneous amplitude envelopes and frequency centroids.

Although the acoustic attributes of this study are comparable to previous studies, it is difficult to directly compare similarity ratings. Our study employed instruments used in previous experiments by other researchers, but the tones used in other studies may have actually been quite different. Different tones from the same type of instrument can have quite different timbres, since timbre is dependent on factors such as register, loudness, and playing technique. Thus it cannot be assumed that tones played by the same instrument in different studies also had the same timbre.

V. CONCLUSION

Three main conclusions can be drawn from these experiments. First, the attributes that are salient for timbral similarity judgments are present throughout tones. Similarity judgments on complete tones corresponded both to those on onsets and remainders. Second, centroid frequency contributes to similarity judgments. This attribute remains relatively constant throughout each tone, so it can not be localized to any particular part of a tone. Last, differences in amplitude envelopes also contribute to similarity judgments. Although no single feature of these amplitude envelopes was isolated as contributing to the similarity judgments of complete tones, the general pattern of similarity persists throughout each tone. Tones that have similar onsets also have similar remainders.

These conclusions raise some important issues for timbre perception research. It may be that similarity and identification judgments are based on different attributes. As detailed in the introduction, researchers (Clark *et al.*, 1963; Saldanha and Corso, 1964; Berger, 1964; Wedin and Goude, 1972; Elliot, 1975) have found that identification relies on onsets. Identification is accurate if listeners hear onsets, and poor if they don't. For the similarity judgments in the present experiments, onsets did not have this special status. Also, the complete tones, onsets, and remainders in our experiments did not seem equally recognizable. The complete tones were easily recognizable and natural sounding. The onsets sounded less natural, but were identifiable. The remainders sounded similar to the complete tones, but the lack of natural onsets made it harder to identify the blown instruments in particular. The similarity judgments did not reflect these differences among the three sets of tones.

Similarity judgments may rely on the comparison of gross acoustic attributes, but identification judgments may rely on acoustic attributes that are more informative of their source. Acoustic characteristics of the onset may be uniquely important for determining if an instrument was struck, blown, or bowed, but gross acoustic differences may be less informative. The centroid frequency changes for different notes on the same instrument (Sandell, 1991), and different playing techniques change the amplitude of a tone over time. These differences would affect similarity judgments, but they would probably not affect the identification of those tones. Thus the results from similarity judgments may not be important for determining what aspects of the acoustics are most important for identification.

The pattern of results in this study would not necessarily generalize to a different set of tones. For example, tones could be synthesized with attributes of their onsets uncorrelated with attributes of their remainders. The onsets and remainders of these tones would then produce different patterns of similarity judgments. For the tones used here, however, judgments on onsets and remainders did correspond. It is possible that the salient acoustic attributes for natural tones generally correlate throughout their duration. A survey on the acoustics of natural tones is needed to test this possibility.

The suggestion that multiple acoustic attributes of timbre lead to the same percept allows for interesting comparisons to speech perception. In speech perception, disparate acoustic attributes can lead to the same phonetic distinction. For example, aspiration noise, formant transitions, and vowel length all correspond to voicing (Lisker, 1957). This is taken as evidence for the claim that speech is processed differently than other sounds since there is no "purely acoustic" reason for combining these attributes (Repp, 1982). Our work suggests that a similar combining of acoustic cues may occur for timbre. Although dynamic attributes of onsets and remainders are acoustically different, they lead to similar perceptions. This work is insufficiently developed in itself to seriously challenge theories of speech perception, but it suggests that disparate acoustic cues may be equivalent in nonspeech sounds.

ACKNOWLEDGMENTS

This research was supported by a grant to Paul Iverson from Sigma Xi, the scientific research society. We would like to thank two anonymous reviewers for comments on an earlier draft of this paper.

Berger, K. W. (**1964**). "Some factors in the recognition of timbre," J. Acoust. Soc. Am. **36**, 1888–1891.

Bracewell, R. N. (**1986**). *The Fourier Transform and Its Applications* (McGraw-Hill, New York).

Clark, M., Luce, D., Abrams, R., Schlossberg, H., and Rome, J. (**1963**). "Preliminary experiments on the aural significance of parts of tones of orchestral instruments and choral tones," J. Audio. Eng. Soc. **11** (1), 45–54.

Elliot, C. (**1975**). "Attacks and releases as factors in instrument identification," J. Res. Music. Educ. **23**, 35–40.

Grey, J. M. (**1975**). "An exploration of musical timbre," Ph. D. dissertation, Stanford University, Stanford, CA.

Grey, J. M. (**1977**). "Multidimensional scaling of musical timbres," J. Acoust. Soc. Am. **61**, 1270–1277.

Grey, J. M., and Gordon, J. W. (1978). "Perceptual effects of spectral modifications of musical timbres," J. Acoust. Soc. Am. **63**, 1493–1500.

Grey, J. M., and Moorer, J. A. (**1977**). "Perceptual evaluations of synthesized musical instrument tones," J. Acoust. Soc. Am. **62**, 454–462.

Kendall, R. A. (**1986**). "The role of acoustic signal partitions in listener categorization of musical phrase," Music Percept., **4**(2), 185–214.

Krumhansl, C. L. (**1989**). "Why is musical timbre so hard to understand?," in *Structure and Perception of Electroacoustic Sound and Music*, edited by S. Nielzen and O. Olsson (Elsevier, Amsterdam).

Kruskal, J. B. (**1962a**). "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," Psychometrika **29**, 1–27.

Kruskal, J. B. (**1964b**). "Nonmetric multidimensional scaling: A numerical method," Psychometrika **29**, 115–129.

Lisker, L. (**1957**). "Closure duration and the intervocalic voiced-voiceless distinction in English," Language **33**(1), 42–49.

Opolko, F., and Wapnick, J. (**1989**). *McGill University Master Samples User's Manual* (McGill University Faculty of Music, Montreal).

Press, W. H., Flannery, B. P., Tuekolsky, S. A., and Vetterling, W. T. (**1988**). *Numerical Recipes in C: The Art of Scientific Computing* (Cambridge U. P., New York).

Repp, B. H. (**1982**). "Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception," Psychol. Bull. **92**, 81–110.

Saldanha, E. L., and Corso, J. F. (**1964**). "Timbre cues and the identification of musical instruments," J. Acoust. Soc. Am. **36**, 2021–2026.

Sandell, G. J. (**1991**). "A library of orchestral instrument spectra," in *Proceedings of the 1991 International Computer Music Conference* (Computer Music Association, San Francisco).

Shepard, R. N. (**1962a**). "The analysis of proximities: Multidimensional scaling with an unknown distance function. I.," Psychometrika, **27**, 125–140.

Shepard, R. N. (**1962b**). "The analysis of proximities: Multidimensional scaling with an unknown distance function. II.," Psychometrika, **27**, 219–246.

Wedin, L., and Goude, G. (**1972**). "Dimension analysis of the perception of instrumental timbre," Scand. J. Psychol., **13**, 228–240.

Wessel, D. L. (**1979**). "Timbre space as a musical control structure," Comput. Music J. **3**, 45–52.

Wilkinson, L. (**1987**). *SYSTAT: The System for Statistics* (SYSTAT, Evanston, IL).