Exploring violin sound quality: Investigating English timbre descriptors and correlating resynthesized acoustical modifications with perceptual properties

Claudia Fritz[a)]
Université Pierre et Marie Curie, UMR CNRS 7190, Institut Jean Le Rond d'Alembert, 4 place Jussieu, 75005 Paris, France

Alan F. Blackwell
Computer Laboratory, University of Cambridge, 15 JJ Thomson Avenue, Cambridge CB3 OFD, United Kingdom

Ian Cross
Centre for Music and Science, Music Faculty, University of Cambridge, West Road, Cambridge CB3 9DP, United Kingdom

Jim Woodhouse
Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

Brian C. J. Moore
Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, United Kingdom

(Received 23 October 2010; revised 31 May 2011; accepted 7 June 2011)

Performers often discuss the sound quality of a violin or the sound obtained by particular playing techniques, calling upon a diverse vocabulary. This study explores the verbal descriptions, made by performers, of the distinctive timbres of different violins. Sixty-one common descriptors were collected and then arranged by violinists on a map, so that words with similar meanings lay close together, and those with different meanings lay far apart. The results of multidimensional scaling demonstrated consistent use among violinists of many words, and highlighted which words are used for similar purposes. These terms and their relations were then used to investigate the perceptual effect of acoustical modifications of violin sounds produced by roving of the levels in five one-octave wide bands, 190–380, 380–760, 760–1520, 1520–3040, and 3040–6080 Hz. Pairs of sounds were presented, and each participant was asked to indicate which of the sounds was more *bright*, *clear*, *harsh*, *nasal*, or *good* (in separate runs for each descriptor). Increased brightness and clarity were associated with moderately increased levels in bands 4 and 5, whereas increased harshness was associated with a strongly increased level in band 4. Judgments differed across participants for the qualities nasal and good. © *2012 Acoustical Society of America*. [DOI: 10.1121/1.3651790]

PACS number(s): 43.75.De, 43.66.Lj, 43.75.Cd, 43.66.Jh [DD] Pages: 783–794

I. INTRODUCTION

The most in-depth empirical studies of violin timbre have, until now, been conducted by Stepánek (2006) and Stepanek and Otcenasek (2004). In that research, participants listened to selected pairs of violin tones and were asked to describe (in Czech) the differences in timbre that they perceived between the tones. Stepanek and Otcenasek went on to use only four words, reported in English publications as *sharp*, *dark*, *clear*, and *narrow*, for further investigation. The objective of their subsequent research was to correlate acoustical properties with qualitatively identified timbral features (Stepanek and Otcenasek, 2005).

This prior research has some limitations, resulting from the fact that the study was conducted using words in one language as the stimuli, but with results reported as applying to words in a different language. Where research concerns subtle interpretive judgments of verbal classifications, extending the conclusions to translated descriptors may not be fully accurate. Further, the descriptors were obtained by listening to only a selection of violin sounds, and it is hard to be sure that either this sample or participant reports covered the whole timbre space.

Dünnwald (1991) attempted to relate the acoustical properties of violins to perceptual qualities. Based on his measurements of the acoustical properties of a large range of violins that had previously been classified as of very good or moderate quality, Dünnwald proposed four frequency bands that he suggested were important for the judgment of sound quality: 190–650, 650–1300, 1300–4200, and 4200–6400 Hz. The first band includes the lower overtones and may be related to "richness." He associated a high level in the second band with "nasality," a high level in the third band with "brilliance,"

[a)]Author to whom correspondence should be addressed. Electronic mail: claudia.fritz@upmc.fr

and a low level in the fourth band with "clarity." However, the basis on which these particular bands were chosen is not reported and he also did not report any investigation of the judgments of an appropriate number of violinists to test the validity either of the bands or of the associated terms.

We therefore decided to explore the verbal description of violin timbre in English in a way that is complementary to these previous studies, by providing an empirically based descriptive vocabulary. Musical acousticians are ultimately concerned with two different contexts in which a descriptive vocabulary might be applied. The first is related to the range of timbres that can be achieved on a given violin with different playing techniques. As demonstrated by Bellemare and Traube (2006) for the piano and Traube and D'Alessandro (2005) for the guitar, descriptive vocabularies can be used as an aid to characterizing sound production and to instructing players. The second context is the distinctive timbres of different violins regardless of the manner of playing: What descriptors are used to characterize the violins themselves? Although we consider that both these contexts are important, in this paper we focus specifically on the second context, as the terms and their relations provide a tool for acoustical research into the quality of instrumental sound, which is our own primary interest.

Our work can be contrasted with a broader tradition of timbre research, typically using a variety of speech and musical stimuli to isolate attributes that are independent of pitch and loudness, toward a general theory of timbre perception (e.g., von Bismarck, 1974). That more general question is a challenging one, complicated by individual differences of interpretation and by complex dependencies between timbre and other attributes (Houtsma, 1997). Nevertheless, we are able to draw comparisons in our discussion to this broader work. The violin is particularly interesting as a research target because there is such a well-established expert community having trained listening skills and a critical vocabulary, and because it is possible to relate perceptual effects to specific object attributes—in particular, to the resonance characteristics of the violin body. We exploit both of these characteristics in our methods.

We first investigated the English descriptors that are used by experts to characterize violins and then used these terms and their relations to quantify the perceptual effect of some acoustical modifications to violins [bearing in mind the warning about the difficulty of such tests and their interpretation sounded by the work of Faure (2000)]. This latter step was achieved by conducting listening tests using "virtual violins," a method developed by the authors and used in a previous study (Fritz *et al.*, 2007). Briefly, the mechanical frequency response function of the violin was mimicked using a digital filter, and the output signal for listening tests was generated by using as input to this filter a recorded bridge-force signal (i.e., the force applied by the bowed strings to the bridge) during a live performance on a violin whose bridge was fitted with a piezoelectric force sensor under each string. One advantage of this method lies in the fact that once the violin response is represented in digital filter form, it becomes very easy to make controlled variations of a kind that would be virtually impossible to achieve by physical changes to a violin. A second advantage is that the bridge-force input stays constant (having been recorded), which therefore removes the effect of the player and their adaptation to the instrument.

There are two main parts to this paper. The first part is dedicated to the development of a lexicon of timbre descriptors. This involved (1) the collection of a set of candidate words, (2) arrangement of those words by violinists to obtain semantic distance measures, and (3) use of multidimensional scaling to characterize the descriptive space. The second part of the paper describes listening tests that were designed to investigate the relation between acoustical properties of the violin, specifically the relative levels in different frequency bands, and the perceptual qualities as expressed in the lexicon of timbre descriptors.

II. DEVELOPMENT OF A LEXICON OF TIMBRE DESCRIPTORS

A. Objectives

The main objective was to derive a lexicon of timbre descriptors that have validity with respect to the language used by experts: violinists and violin makers/adjusters. However, this lexicon should also have sufficient internal structure to represent the space of possible descriptions without undue duplication or redundancy through inclusion of synonyms and near synonyms.

B. Collection of the descriptors

Interviews were conducted with 19 violinists (native English speakers), in which they were asked to supply between five and ten words that they would use to describe the timbre of a violin and to give, if possible, synonyms and antonyms. Terms were also collected from descriptions of violin sound published in ten full volumes of "The Strad" magazine (Vols. 107–117, published between 1996 and 2007). Although The Strad has been published since 1890, we focused on the ten volumes most recently published at the time of the study, to avoid possible effects of shifts in the use of language over time. The articles reviewed were those dealing with the description of famous violins, with performances on particular violins, and with violin construction techniques. As we wished to develop descriptors that were meaningful to a broad expert community of violin players and makers, we removed those descriptors that occurred fewer than three times (some examples include "silvery," "golden," and "woody"). The final list included 61 descriptors, which are given in Table I (in alphabetical order).

C. Two-dimensional (2D) spatial arrangement of the descriptors

Fifteen experienced violinists (native English speakers) were asked to arrange the 61 words by similarity. The experimental approach was a variant of the card-sort technique for eliciting and structuring experts' knowledge that is frequently used in the knowledge engineering field (e.g., Rugg and McGeorge, 2005). In these techniques, participants are provided with a set of cards, each containing one term. They

TABLE I. Full set of descriptors.

Alive	Balanced	Brash	Bright
Brilliant	Clean	Clear	Closed
Cold	Complex	Dark	Dead
Deep	Dull	Even	Free
Full	Hard	Harsh	Heavy
Interesting	Light	Lively	Loud
Mellow	Metallic	Muffled	Muted
Nasal	Not penetrating	Open	Penetrating
Piercing	Powerful	Pure	Quiet
Raspy	Resonant	Responsive	Rich
Ringing	Rough	Round	Sharp
Shrill	Singing	Smooth	Soft
Sonorous	Steely	Strident	Strong
Sweet	Thin	Tinny	Tiny
Unbalanced	Uneven	Unresponsive	Warm
Weak			

are asked to arrange these cards into piles containing related terms. Participants can create any number of piles that they consider appropriate, thereby capturing natural semantic clusters. However, conventional card-sort methods are relatively insensitive, because similarity between terms can only be rated as either 1 (same pile) or 0 (not the same).

A more sensitive alternative to card sort is the use of pairwise semantic similarity ratings, in which participants report the degree of similarity between two terms using either a Likert scale or continuous slider (e.g., Charles, 2000). There are two disadvantages to this technique. The number of individual comparisons is large (3660 for our data set). More seriously, independent presentation of pairs of terms makes it difficult for participants to calibrate which pairs they consider to be more similar than others, so that subtleties of usage and implication may be lost. Semantic differential comparison of each term to two other terms could mitigate this problem, but requires even larger numbers of individual comparisons (216000 for our data set).

We therefore developed a technique that is more sensitive than card-sort methods, but is practically applicable for large numbers of terms. It also allows participants to directly report clusters of terms, to a greater extent than either similarity ranking or semantic differential techniques, and to at least the same extent as card sort. The experimental setup is shown in Fig. 1. We used an EXCEL spreadsheet, into which participants could arrange terms by cutting and pasting them from an initial randomized list at the side of the screen. Participants were able to construct clusters by placing items in the same columns or rows of the spreadsheet, could express relative distance between clusters by placing them at opposite sides or corners of the grid, and could express relative similarity within clusters by ordering items within the same part of a row or column. Participants were allowed to leave out any words that they considered not relevant to that descriptive context. They were allowed to move each descriptor as many times as required to achieve a satisfactory arrangement.

Although potentially less reliable than semantic differential techniques (because the 2D grid places constraints on the ability to express multiple independent relations), the technique has improved construct validity, as participants are able to construct a layout that directly expresses their intentions to the experimenter. Participants were asked to arrange words in such a way that words with similar meanings were close together and words with different meanings were farther apart. They were told that the overall arrangement should be such that the distance between any two words indicated how similar the meanings of those two

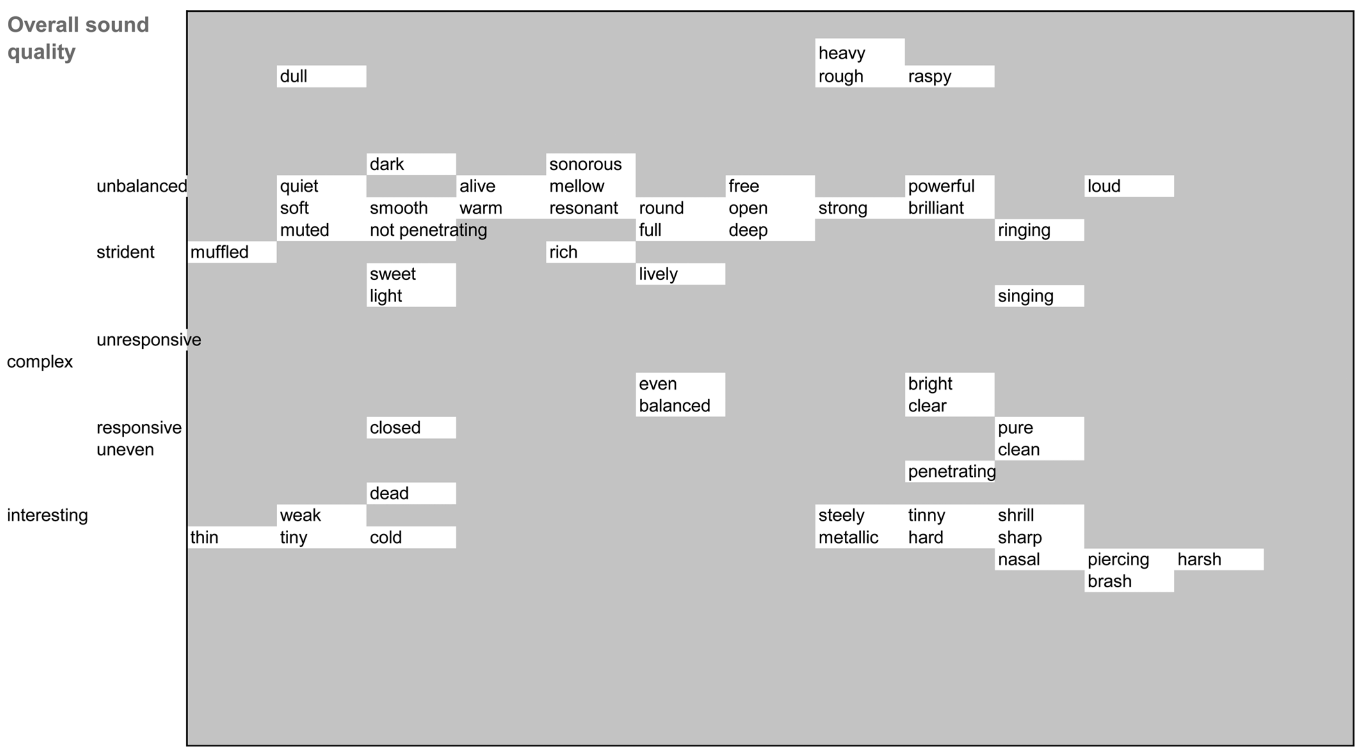

FIG. 1. Illustration of EXCEL spreadsheet used for grouping of words according to similarity of meaning.

words would be, that they could construct any arrangement that would make those distances most accurate with respect to each other, and that they could refine their arrangement until they were happy that they had achieved the best possible arrangement.

As with card-sort and semantic-differential methods, it is necessary to derive a numerical representation of the distance between paired terms in each participant response, in order to construct an aggregated model over the whole sample. This distance encoding should capture the use of rows and columns to express clusters, should take into account the relative visual impact of aspect ratio and blank fields, and must be a function that monotonically increases from the closest possible arrangement (two neighboring cells in the same column) to the furthest possible (opposite corners of the grid). A function that provides these characteristics was defined as an EXCEL macro, and is reported in the Appendix. As noted, the values returned by this function should not be regarded as being equivalent to those for a semantic-differential technique. However, the function does provide a range of values that are greatly superior to those for the conventional card-sort measure, which provides only values of zero or one for each participant/term-pair.

Participants were asked to make four different arrangements, each relating to a different context in which the words might be used to describe timbre. The four contexts were when describing: (i) The overall sound quality of a violin; (ii) the sound quality of the lower violin strings; (iii) the sound quality of the higher strings; and (iv) the ease of playing of an instrument.

D. Analyses

The maps made by each participant were converted into distance matrices using the EXCEL macro defined in the Appendix. Distance data for the words that had been left out were considered to be missing. However, as our planned analyses could not be performed reliably if too many distance estimates were missing, any words that had been left out by more than half of the participants were removed for analysis of that context. The original descriptor list remained complete for the "overall sound quality context," but in other contexts it was reduced as follows.

1) For the "ease of playing" context, only these words remained: *alive*, *balanced*, *clean*, *clear*, *closed*, *dead*, *dull*, *even*, *free*, *full*, *heavy*, *interesting*, *lively*, *open*, *resonan*t, *responsive*, *rich*, *sonorous*, *unbalanced*, *uneven*, *unresponsive*, *warm*, and *weak*.
2) For the "low strings sound quality" context, only these words remained: *closed*, *complex*, *dark*, *dead*, *deep*, *dull*, *full*, *heavy*, *interesting*, *loud*, *mellow*, *muffled*, *muted*, *open*, *powerful*, *quiet*, *raspy*, *resonant*, *rich*, *rough*, *round*, *singing*, *smooth*, *soft*, *sonorous*, *strong*, *warm*, and *weak*.
3) For the "high strings sound quality" context the whole list remained, *except* for: *complex*, *dark*, *deep*, *even*, *full*, *heavy*, *interesting*, *mellow*, *rich*, *round*, *uneven*, and *warm*.

These distances were used as the basis for multidimensional scaling analyses using the ALSCAL algorithm (A-MDS) in SPSS, as described in the following. We expected that the S-stress (a measure of goodness of fit) for each context might be relatively large, and possibly even above the limit of 0.15, below which A-MDS analysis is often considered to be inappropriate, for the following reasons: (1) there were a large number of words; (2) there were individual differences in layout strategy; (3) the distance data were derived from discrete intervals rather than continuous values. We therefore used scree plots (graphs plotting stress as a function of the number of MDS dimensions) to identify the number of dimensions above which stress did not decrease markedly, as recommended by Borg and Groenen (1997).

Figure 2 can be used to judge where stress did not decrease markedly with increasing number of dimensions, and also to determine when the stress value approached the rule-of-thumb figure of 0.15. Both criteria indicated that the appropriate number of dimensions for MDS analysis of overall sound quality was three. The most appropriate number of dimensions for "high-strings quality" and ease of playing was judged to be two. The curve for "low-strings quality" could indicate a choice of three dimensions. However, a trial analysis with three dimensions indicated a very small spread on the third dimension, such that it did not offer useful analytic value. We therefore made an *a posteriori* decision to use two dimensions for this context, thereby allowing direct comparison between the high- and low-strings contexts.

III. RESULTS AND INTERPRETATION

A. Overall sound quality

The results of the A-MDS for overall sound quality are shown in Fig. 3. This corresponds to the original map provided by SPSS, but rotated by the angle vector [2.2 −0.2 0.7] to make it more interpretable. All the terms conventionally used to describe desirable tone qualities are clustered below the positive diagonal. This indicates that dimensions one and two both describe properties that can be classified between one desirable and one undesirable extreme. It also indicates that these two dimensions can be combined in a way that retains their individual degrees of desirability. Finally, the relatively clear separation of upper-left clusters

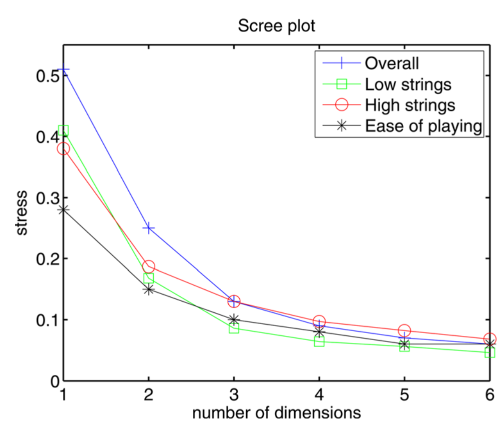

FIG. 2. (Color online) Value of stress as a function of the number of dimensions used in the A-MDS, for each of the four contexts.

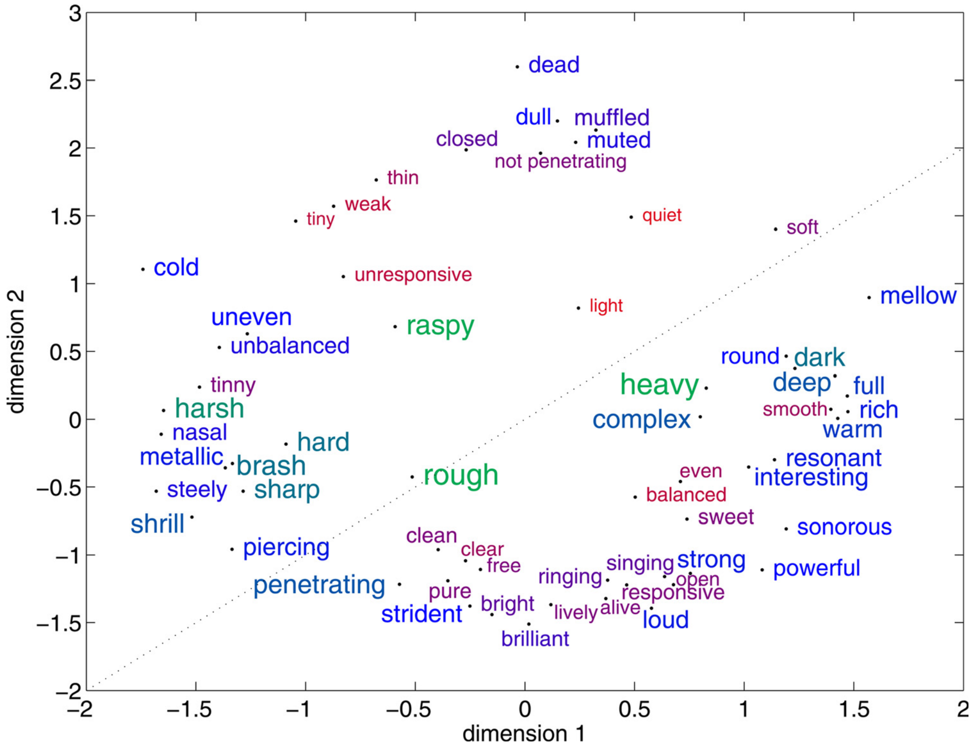

FIG. 3. (Color online) A-MDS map for "overall sound quality." The third dimension is represented by the size of the text (and by color in the online version of the paper). The foreground is represented by larger text and the background by smaller text. The middle range is rendered proportionally as intermediate sizes of text.

and lower-right clusters across this diagonal indicates that almost all violin timbre descriptors incorporate an evaluative judgment as being either good or bad properties.

Looking more closely at the individual dimensions, it appears possible to characterize these dimensions in terms of qualitative similarity between the clusters of descriptors at their extremes. We wish to emphasize that the dimensions need not correspond to any single acoustic or perceptual property, but should be regarded as a "rich" description requiring interpretation of the concurrence of this group of terms. Table II summarizes the dimensions that we have found, and gives interpretive comments from both perceptual and acoustic perspectives. We offer some convenient acronyms, having a degree of onomatopoeic correspondence to the dimension extremes.

TABLE II. Characterization of sound quality dimensions.

Dimension	Undesirable	Desirable	Interpretation
1 MCH/WRM	Metallic	Warm	Spectral balance
	Cold	Rich	Undesirable associated with excessive high-frequency content or too little low-frequency content
	Harsh	Mellow	
2 MDD/BRL	Muted	Bright	"Amount of sound" produced by the instrument
	Dull	Responsive	Energy in the spectrum, especially in middle and upper ranges
	Dead	Lively	
3 BRR/ESL	Brash	Even	Noisy character
	Rough	Soft	Width of distribution of spectral energy
	Raspy	Light	

B. High-strings sound quality context

The results of the A-MDS for high-strings sound quality are shown in Fig. 4. The two dimensions obtained for the high-strings sound quality context appear to correspond very well to the first two dimensions of the "overall sound quality" context: "spectral balance" for the first dimension, and "amount of sound" for the second dimension. However, comparisons between these multiple MDS results should not be overinterpreted, because the scales of each axis may differ from one map to another, making interpretations hard to visualize, especially in three dimensions. A more straightforward comparison can be made in terms of the A-MDS input data, recording

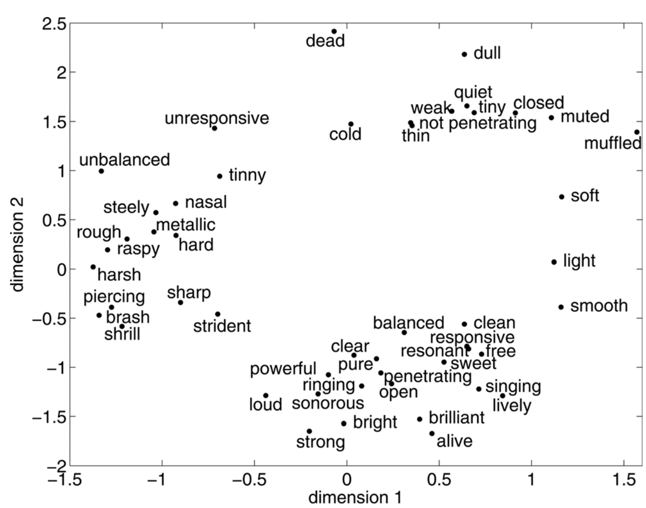

FIG. 4. A-MDS map for "high-strings sound quality."

average distance between descriptors in each context. Those with the lowest average distance are most often clustered with other descriptors, so can be considered to provide a reference frame of timbre description for that context. The descriptors with the lowest average distance from other descriptors in the high strings context are *clean*, *clear*, *free*, and *hard*, which provide a plausible central frame for description of high-strings timbre. It is also interesting to consider those descriptors that are least often included in clusters, either placed alone, or at the edges of the layout. These descriptors can be identified because they have the highest average distance from descriptors. In this context, *dead* and *dull* have the highest average distance. This is discussed further in the following.

C. Low-strings sound quality context

The results of the A-MDS for low-strings sound quality are shown in Fig. 5. The dimensions derived for this context correspond well to the dimensions obtained for the two previous contexts. Dimension 1 relates mainly to spectral balance, whereas dimension 2 relates mainly to the amount of sound. However, in this context words relating to strong high-frequency content, such as *sharp*, *shrill*, and *harsh*, were not used. The descriptors with the lowest average distance from other descriptors in the low-strings context are *deep*, *dark*, *warm*, and *mellow*. These provide a plausible central frame for the description of low-strings timbre, which is clearly distinct from the reference frame for the description of high-strings quality. This is in agreement with the observation that players judge the low and the high strings independently, an observation that led us to differentiate these two contexts. Meanwhile *dead* and *raspy* are most often placed alone, with the highest average distance from other descriptors. It is interesting that the word dead plays this role in both the high- and low-strings context, and this is discussed further in the following.

D. Ease of playing

The results of the A-MDS for ease of playing are shown in Fig. 6. The first dimension corresponds to the positive or

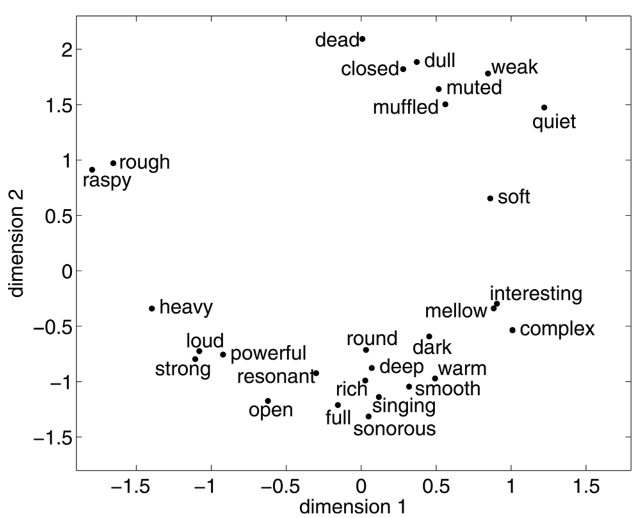

FIG. 5. A-MDS map for "low-strings sound quality."

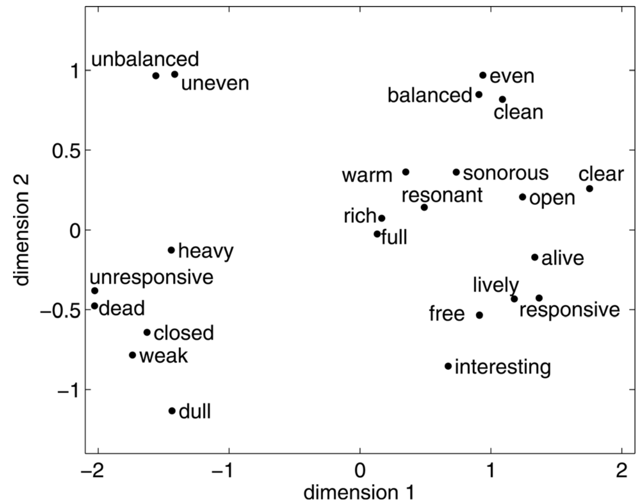

FIG. 6. A-MDS map for "ease of playing."

negative extent to which a violin is easy to play in different respects. Words relating to good playing characteristics always appear on the right-hand side, and words associated with a violin being hard to play appear on the left-hand side. The second dimension is related to the various respects in which a violin can be either easy or hard to play. High values on dimension 2 correspond to factors relating to balance and evenness. Low values on dimension 2 seem to be related to overall responsiveness; the extent to which effort is needed to obtain a given output.

This interpretation is supported by the positioning on dimension 2 of those pairs of words that are ordinarily antonyms, such as *even*/*uneven*, *balanced*/*unbalanced*, *lively*/*dead*, and *responsive*/*unresponsive*. However, although words like *heavy* and *weak* were used, *light* and *strong* were not. This indicates that, when working with semantic scales, if a preparatory lexical study such as this one has not been performed, then one should use monopolar scales (e.g., *open–not open*) rather than dipolar scales (e.g., *open–closed*), because the words at each end of the scale may not be relevant antonyms for judging violins. The necessity of using monopolar scales for timbre judgments, in contrast to the bipolar scales traditionally used in differential semantic scaling, has already been demonstrated by Kendall and Carterette (1992a,b) in their study of verbal timbre attributes of wind instruments.

In this context, the descriptors with the lowest average distance from other descriptors are *full* and *rich*. These provide a plausible central frame when describing ease of playing, because they indicate a broad range of timbre qualities that can be accessed in performance. The descriptors that were most often placed alone in this context are once again *dull* and *dead*, together with *unresponsive* and *weak*, which are more directly related to the effort that the player must apply to produce a given amount of sound. In all three of the specific contexts, expert violinists have placed these "deadness" descriptors furthest away from other terms, suggesting that—as they have been selected for inclusion in the layout, but are not considered similar to the other selected descriptors—they are used to express lack of similarity to any interesting timbre quality in the instrument, and correspondingly increased effort that the player must apply to produce the desired timbre effects in performance.

IV. CORRELATIONS BETWEEN ACOUSTICAL MODIFICATIONS AND PERCEPTUAL PROPERTIES

A. Aim and general methodology

The aim of this experiment was to determine the effect of acoustical modifications, specifically changes in frequency response of the violin body, on the perceptual qualities described by the words established in the first part of the study. The general approach was to present pairs of modified sounds to expert violinists, who were asked to choose which sound corresponded better to a verbal descriptor. The results were used in a correlation analysis to determine which aspects of the violin frequency response contribute to particular dimensions of timbre description.

Instead of modifying a real violin acoustically we worked with "virtual modified violins," as in an earlier work (Fritz *et al.*, 2007). As will be described in detail in Sec. IV C, the response of a real violin was emulated by an accurately fitted digital filter model and the desired changes were then made to this digital filter. In the present experiment, we wished to investigate the relationship of our descriptors to a wide variety of possible acoustical modifications. Mechanical changes to a violin generally modify the body response over a wide frequency range. In order to explore the perceptual effect of such changes, we separated the vibration modes into five frequency bands, each one octave wide, and imposed independent modifications on each band. The frequency ranges of the bands were 190–380 Hz (band 1), 380–760 Hz (band 2), 760–1520 Hz (band 3), 1520–3040 Hz (band 4), and 3040–6080 Hz (band 5). The lowest edge frequency of 190 Hz corresponds approximately to the violin's lowest note (G3, 196 Hz), and is low enough to include the first important resonance around 280 Hz in band 1.

B. Design of the listening tests

The aim of the listening tests was to determine the correspondence between the space of possible changes in violin frequency response and the dimensions of timbre description investigated in the first part of this paper. The general strategy was to choose two stimuli, and ask a listener to compare them in terms of a particular descriptor X, judging which of the sounds is "more X" and which is "less X."

This experimental design requires a relatively large number of pairs of sounds to perform correlation analysis, so in practice the experiment can only be conducted for a relatively small set of descriptors. By using the A-MDS maps reported in the first part of this paper, it was possible to reduce the 61 candidate descriptors to a representative set that was sufficiently small to allow a large number of perceptual judgments for each descriptor. We chose three descriptors that were widely spread in the context of overall sound quality, as illustrated in Fig. 3, and that were thought likely to be related to differences in spectral shape. These were: *bright*, *harsh*, and *nasal* (note that harsh and nasal were separated on dimension 3). We used an additional descriptor from Dünnwald (1991), namely *clear*, to allow comparison with his work. We also used the general descriptor *good*, to assess whether any specific spectral shape was associated with overall preference; as von Bismarck (1974) notes, "an aesthetic-evaluative" dimension is often discovered in factor analyses of the ratings of complex sounds. The word *good* was chosen to allow clear preference judgments while being completely neutral with respect to violin and musical terminology. This contrasts with some of the other terms, such as *bright* or *nasal*, for which it is not obvious that "more" of the quality is always good or bad.

For each descriptor, 150 pairs of sounds, played successively, were presented. On each trial, the participant was asked to judge whether the first or the second sound was more X than the other (with X being one of the descriptors). The tests were administered using MATLAB (The MathWorks, Natick, MA). Participants were first provided with a short explanatory text to read before starting the test. The background of the study as well as the instructions were given in this text.

The test itself was divided into three sections for each descriptor, with mandatory breaks between them. Participants were also encouraged to take a break whenever they felt tired. In each section, 50 pairs of sounds were rated along one of the scales. The order in which the scales were presented to the participants was randomized across participants, and different stimuli were generated for each participant.

C. Generation of the stimuli

The acoustical response of a good quality violin made by David Rubio was measured by applying a calibrated impulse to the violin bridge and measuring the velocity response at the bridge with a laser Doppler vibrometer. Modal identification techniques (Ewins, 2000) were used to fit this response in the frequency range up to 7000 Hz using 54 modes, each defined by an amplitude A_i, a natural frequency ω_i, and a quality factor Q_i. The impulse response of the violin body is then

$$g(t) = \sum_{i=1}^{54} A_i \cos(\omega_i t) e^{-\omega_i t / Q_i} \quad (1)$$

and this can be used directly to give a finite impulse response (FIR) digital filter to simulate the transient response of the violin.

Each modified FIR digital filter for the listening tests was created as follows. Five random numbers were generated using the MATLAB function RAND, drawn from a population whose decibel level had a uniform distribution in the range ± 10 dB. The first of these numbers was used to scale the amplitudes A_i of all vibration modes with natural frequency falling within frequency band 1, the second for band 2 and so on. Quality factors and natural frequencies were not changed. Equation (1) with the modified levels was used to generate a filter. Adjustment was done via modal amplitudes, rather than equalization filters, in order to preserve the physical relevance of the results (in principle, the modifications could be achieved in a physical violin by adjustment of the structural details), and also to minimize possible artifacts that might have resulted from the use of filters with sharp boundaries. The scaling range ± 10 dB was chosen because it is roughly twice the threshold value obtained by Fritz *et al*. (2007) for detection of changes within the Dünnwald (1991) frequency bands. It results in a set of stimuli for which typical differences are clearly audible, whereas the overall manipulation remains within the range of plausible variation for a natural violin sound.

A force signal directly recorded from the bridge of a violin during normal playing was fed into the filter to resynthesize the sound of the modified virtual violin. Here, a single recorded bridge signal was used to create the audio stimuli: this signal was a two-note phrase extracted from a legato performance of the third theme of the Glazunov *Concerto for violin in A minor op. 82*. The passage was played *mezzo-forte* with an amount of vibrato typical for this type of music. The two notes [A♭3 (207 Hz) followed by F4 (349 Hz)], have a total duration of 3 s. This passage has been used in previous studies and was retained for consistency. It was originally chosen on the basis of its effectiveness in live-performance tests of discrimination between different violins. Further details of measurement, recording and processing can be found in Fritz *et al*. (2007), Sec. II B 2.

For each participant, a set of 300 stimuli was synthesized. The random scaling factors were chosen independently for each band, and were varied randomly from stimulus to stimulus, and from participant to participant. However, the same 150 pairs were used for the five descriptors for a given participant. The general concept is similar to that used in studies of "molecular psychophysics" (Ahumada and Lovell, 1971; Berg, 1989; Richards and Zhu, 1994).

D. Participants

Fourteen participants took part. They were recruited separately from those described in the first part of this paper. Twelve were expert violinists, with a pass at Associated Board Grade VIII as a minimum requirement. The other two were a leading English violin maker, and one of the authors, who has considerable experience in listening to violin sounds. All participants were native British speakers. Their hearing was checked to be normal using a Grason-Stadler (Eden Prairie, MN) GSI 16 audiometer (defined as audiometric thresholds below 15 dB HL at the standard audiometric frequencies). The 12 violinists were paid for their participation.

Some data were lost for one participant, for *harsh* and *nasal*, as the result of a software error. In addition, one participant did not do the test for the descriptor *good*. Hence for these three descriptors, only the data for 13 participants are available.

E. RESULTS AND DISCUSSION

We determined the correlation between the acoustical modifications to the resonance of the baseline violin and the direction assigned to these modifications when they were considered as descriptor vectors. The answer $a(i)$ of a participant when listening to the ith pair of sounds was coded 1 if the first sound was considered more X than the second sound and -1 if the second sound was considered more X than the first sound. The correlation for octave band k was then calculated as

$$\mathrm{corr}(k) = \frac{\sum_{i=1}^{150} C(k,i)a(i)}{\sqrt{\sum_{i=1}^{150} C(k,i)^2 \sum_{i=1}^{150} a(i)^2}}. \quad (2)$$

$C(k,i)$ is defined in terms of $A'_k(i)$ and $B'_k(i)$, which are the modified amplitude values of band k for, respectively, the first sound and the second sound of pair i:

$$C(k,i) = \frac{A'_k(i)}{\sum_{j \neq k} A'_j(i)} \frac{\sum_{j \neq k} B'_j(i)}{B'_k(i)}. \quad (3)$$

This corresponds to the ratio between the band-k modified amplitudes of the two sounds normalized by the sum of the modified amplitudes in the other bands.

To check whether the number of pairs of sounds presented to each participant was sufficiently large to give stable results, the results for each participant were divided into two halves (α and β), and the results for each half were analyzed separately. The quadratic means (qm) of the differences between the correlations obtained for each descriptor for the two halves were calculated for each subject:

$$qm = \sqrt{\frac{\sum_{k=1}^{5} \left[corr_\alpha(k) - corr_\beta(k)\right]^2}{5}}. \quad (4)$$

These means were averaged across subjects and across descriptors to obtain a split-half reliability index $I = 0.17$ with a standard deviation of 0.06. The maximum possible value of I being 2, the index obtained was small enough to consider the results to be stable and thus the number of pairs of sounds to be sufficient.

As correlations fall between -1 and $+1$, they cannot be normally distributed. For statistical comparisons, we used the r to z transform proposed by Fisher (1915), which makes the numbers more normally distributed:

$$z = \frac{1}{2}\frac{\ln(1+r)}{\ln(1-r)}, \quad (5)$$

where r is the correlation (Fig. 7 shows the actual correlation values, rather than z-transform values). Hereafter all the statistical analyses are performed on the z-transform values and so in the following, "data" refer to the z transform of the correlations.

A first step in the analyses was to check the consistency between participants' judgments for each descriptor. To do this, the Intraclass Correlation Coefficient (ICC) for each descriptor was computed using SPSS. The results are given in Table III. Although the consistency between participants was

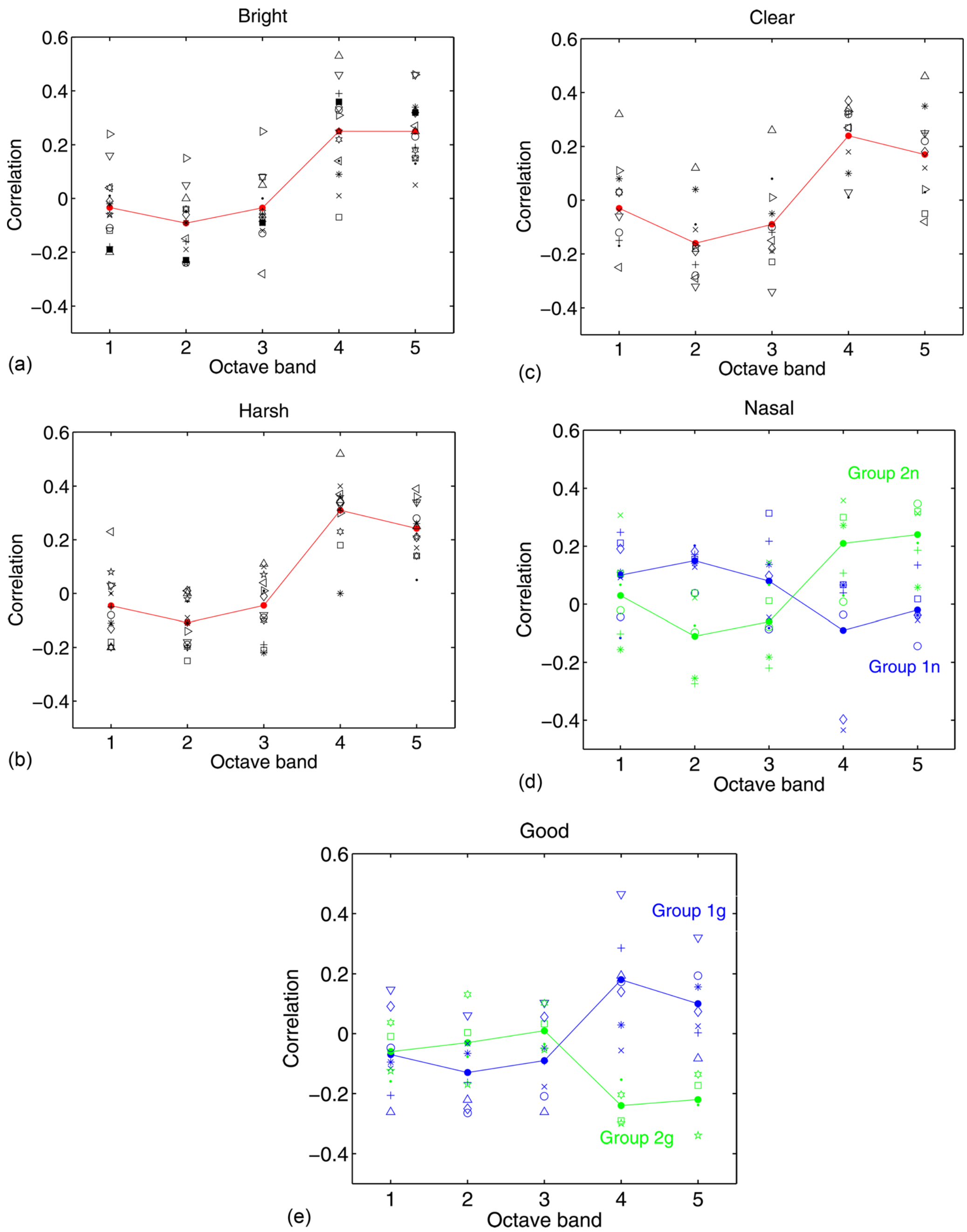

FIG. 7. (Color online) Plots of correlation versus band number for the five descriptors. Symbols show individual data and lines show the means.

very high for *bright*, *harsh*, and *clear*, the consistency was extremely poor for *nasal* and *good*. It is not sensible to average results across participants when the ICC is very low. The data were inspected to determine whether the participants could be divided into subgroups with higher consistency within groups. For *nasal*, two groups were found: The first group, labeled *1n*, contained seven participants and the second group, *2n*, contained the remaining six participants. For *good*, two groups were also found (*1g* with seven participants and *2g* with four participants); the remaining two participants did not fit in either group. Groups *1n* and *1g* shared five participants, whereas groups *2n* and *2g* shared four participants. Table IV shows the ICC values obtained for these subgroups. The ICC values were high except for group *1n*. For *clear*,

TABLE III. ICC and *p* values for each descriptor.

Descriptor	Bright	Harsh	Clear	Nasal	Good
ICC	0.97	0.98	0.89	−1.1	0.3
P	<0.001	<0.001	<0.001	0.71	0.37

TABLE IV. ICC and *p* values for the subgroups for the descriptors nasal and good.

Descriptor	Nasal		Good	
Subgroup	*1n*	*2n*	*1g*	*2g*
ICC	0.76	0.93	0.91	0.94
P	0.01	<0.001	<0.001	<0.001

three participants gave patterns of responses that were clearly different from that for the remainder. When the data for these three were omitted, the ICC increased to 0.96 with $p < 0.001$.

Correlation is plotted as a function of band number for each descriptor in Fig. 7. The graphs are based on analyses excluding outliers, so there were 14 participants for *bright*, 13 for *harsh*, 11 for *clear*, two groups of 7 and 6 participants for *nasal*, and two groups of 7 and 4 participants for *good*. Symbols show the results for individual participants and lines show means. An analysis of variance (ANOVA) was performed on the data for each descriptor. For all descriptors, there was one within factor, band number, and, for *nasal* and *good*, there was one between factor, group. The influence of band number was significant for *bright* [$F(2.5,32.8) = 34.3$, $\varepsilon = 0.63$, $p < 0.001$], *harsh* [$F(2.7,32.6) = 49.9$, $\varepsilon = 0.68$, $p < 0.001$[1]] and *clear* [$F(4,40) = 21.4$, $\varepsilon = 1$, $p < 0.001$]. For the other two descriptors, *good* and *nasal*, there was no significant effect of band number or group. However, there were significant interactions: for *nasal*, $F(4,44) = 14.7$, $p < 0.001$; for *good*, $F(4,36) = 17.2$, $p < 0.001$. This demonstrates that individuals differ in the physical characteristics they use to judge the qualities *nasal* and *good*. Where the pattern of results was similar across participants (for *bright*, *clear*, and *harsh*), this means that participants used similar criteria and physical characteristics to make their judgments.

The quality *bright* was associated with an increased level in bands 4 and 5 and slightly decreased levels in bands 2 and 3. The results for quality *clear* showed a similar pattern, suggesting that, within the limited range of manipulations used in our tests, judgments of the qualities *bright* and *clear* were based on similar physical spectral changes. This is consistent with the results of the A-MDS for overall sound quality, as shown in Fig. 3; *bright* and *clear* were very close together in the MDS space, indicating that the terms have similar meanings for violinists. The results for *harsh* showed a similar pattern to those for *bright* and *clear*, greater harshness being associated with higher levels in bands 4 and 5. However, the A-MDS results in Fig. 3 suggest that *harsh* has a very different meaning from *bright* or *clear*. The distinction between *harsh* and *bright* or *clear* may be related to the relative amount of energy in band 4. For *harsh* the correlation of judgments with the level in band 4 was about 0.3, whereas for *bright* and *clear* the correlation was only about 0.2. It seems likely that a moderately high level in band 4 is associated with increased brightness or clarity, but if the level in band 4 becomes too high, this is associated with harshness.

For judgments of *nasal*, the results for the two groups were approximately complementary. For group *1n*, increased nasality was associated with increased levels in bands 1, 2, and 3, and decreased levels in bands 4 and 5. For group *2n*, increased nasality was associated with decreased levels in bands 1, 2, and 3, and increased levels in bands 4 and 5. For judgments of *good*, the results for the two groups were also approximately complementary. Increased goodness was associated with an increased level in bands 4 and 5 for group *1g* and a decreased level in bands 4 and 5 for group *2g*. Evidently, the participants in the different groups disagreed about what constitutes a *nasal* or a *good* sound quality. It is interesting that the results for *good* and *nasal* were related: Participants who found a modification more nasal liked it less and participants who found a modification less nasal liked it more.

Our results are not consistent with the proposal of Dünnwald (1991) that a high level in the band from 650 to 1300 Hz, corresponding roughly to our band 3, is associated with nasality. *Nasal* cannot be associated with one band, and seems to be related to different criteria depending on the participant. Dünnwald also proposed that a high level in the band from 1300 to 2400 Hz is associated with brilliance or brightness. Our results suggest that the frequency range associated with *brightness* extends above this, as an increase in level in band 5 (3040–6080 Hz) made the sound brighter. Finally, Dünnwald proposed that increased clarity was associated with a lower level in the range 4200–6400 Hz, whereas our results show that clarity was associated with an increase in level in bands 4 and 5 (1520–6080 Hz).

V. DISCUSSION AND CONCLUSIONS

The study of the verbal attributes used by English-speaking violinists to describe the timbre of violins shows how these different attributes relate to each other, depending on the context. Four contexts were used: overall sound quality, high- and low-strings quality, and ease of playing. The results of multidimensional scaling showed consistent use among violinists and suggested that the dimensions identified had an evaluative aspect; they were related to properties that are regarded as good or bad. It would be desirable to establish the relationship between these verbal dimensions and acoustical properties of violin sounds, but this may be a very difficult task. Following Plomp (1970), a technique that has been used to model the perception of musical timbre is to construct a multidimensional perceptual space from similarity judgments for pairs of sounds. Most studies (Grey, 1977; Krumhansl, 1989; McAdams *et al.*, 1995) have found a three-dimensional space, based purely on perceived similarity of acoustical stimuli. The authors have then searched for acoustical correlates of the perceptual dimensions. There is general agreement about the acoustic correlates of two of the dimensions, which are thought to represent the spectral center of gravity and the rise time of the temporal envelope (the attack). However, there is some disagreement about the acoustical correlates of the third dimension.

Our own approach has been to construct a semantic space independently of specific stimuli, so that similarity judgments can be analyzed in terms of the described experience, in addition to the acoustic properties of the stimuli. The research of Faure (2000) demonstrates the difficulty in establishing a clear correspondence between semantic descriptors and acoustic correlates, and gives some cause for caution in studies of this kind. Nevertheless, our results suggest that dimensions related to overall sound quality appear to be associated with (a) spectral balance [related to the spectral centre of gravity, and comparable to von Bismarck's (1974) dimension of "sharpness"], (b) the amount of energy across the spectrum of resonances produced by the instrument (comparable to von Bismarck's dimension "full–

empty"), and (c) "noisiness" (comparable to von Bismarck's dimension "compact–scattered").

In the second part of the study, a selection of the terms identified and the term *good* were used to investigate the perceptual effect of acoustical modifications of violin sounds produced by roving of the modal levels in five one-octave wide bands, 190–380, 380–760, 760–1520, 1520–3040, and 3040–6080 Hz. Pairs of sounds were presented, and the participants (mostly violinists) were asked to indicate which of the sounds were more *bright*, *clear*, *harsh*, *nasal*, or *good* (in separate runs for each descriptor). Increased brightness and clarity were associated with moderately increased levels in bands 4 and 5, whereas increased harshness was associated with a strongly increased level in band 4. Judgments differed across participants for the qualities *nasal* and *good*. For these two qualities, the participants could be divided into two groups; the patterns of responses were consistent within each group, but differed across groups.

Except for *bright*, these results are not consistent with the assumptions of Dünnwald (1991), which have been often used as a reference among violin makers.

It seems that violinists differ in the way that they use the terms nasal. It would be of interest in future work to understand better what nasality means for players: Are there different kinds of nasalities, e.g., associated with a twangy voice or the production of nasal vowels? Is the use of the term nasality language dependent? A study is presently being conducted using the French language to address these issues.

Violinists also differ in what they judge to be a "good" spectral shape for a violin sound. It is possible that we found this result because of the short synthetic sounds used here. However, the finding seems to agree with the preliminary results of a study that is currently being conducted using real violins to investigate the consistency of judgments of violin players.

ACKNOWLEDGMENTS

We are grateful to the Leverhulme Trust for funding this research and to Andrew Norman, Wendy Mak, and Eugene Feygelson for helping to design and run the experiment. We thank two anonymous reviewers for detailed constructive comments.

APPENDIX

This appendix describes an EXCEL algorithm defining a metric for distance between two cells of a spreadsheet in which values have been arranged into separated clusters. This metric is defined to take into account three observed layout strategies—the use of neighboring cells to define (categorical) clusters, separation of individual items or clusters by (ordinal) numbers of rows or columns, and (ratio) diagonal distance across the screen.

```
Const AspectRatio As Double = 1.5
Const ClusterDiscount As Double = 0.5
  If (firstRow = secondRow) Then
    ' The two cells are in the same row, so
    look for a row cluster
    If (firstCol < secondCol) Then
        leftCell = firstCol
        rightCell = secondCol
    Else
        leftCell = secondCol
        rightCell = firstCol
    End If
    distance = ClusterDiscount * Aspec-
    tRatio * (rightCell - leftCell)
    For intervening = (leftCell + 1) To
    (rightCell - 1)
        If(IsEmpty(Cells(firstRow,
        intervening))) Then
          distance = rightCell - leftCell
          Exit For
          End If
    Next intervening
  ElseIf (firstCol = secondCol) Then
    ' The two cells are in the same column,
    so look for a column cluster
    If (firstRow < secondRow) Then
        topCell = firstRow
        bottomCell = secondRow
    Else
        topCell = secondRow
        bottomCell = firstRow
    End If
    distance = ClusterDiscount*
    (bottomCell - topCell)
    For intervening = (topCell + 1) To
    (bottomCell - 1)
        If(IsEmpty(Cells(firstCol,
        intervening))) Then
          distance = bottomCell - topCell
          Exit For
        End If
    Next intervening
  Else
    ' The cells are in different rows and
    columns - calculate distance
          distance =
          gap + Sqr
          ((secondRow - firstRow)
          ^2 + (AspectRatio
           * (secondCol - firstCol))^ 2)
  End If
```

[1]Here and in the following, the Huynh–Feldt correction was applied.

Ahumada, A., and Lovell, J. (**1971**). "Stimulus features in signal detection," J. Acoust. Soc. Am. **49**, 1751–1756.

Bellamare, M., and Traube, C. (**2006**). "Investigating piano timbre: relating verbal description and vocal imitation to gesture, register, dynamics and articulation," in *Proceedings of the Ninth International Conference on Music Perception and Cognition*, Bologna, Italy, pp. 59–60.

Berg, B. (**1989**). "Analysis of weights in multiple observation tasks," J. Acoust. Soc. Am. **86**, 1743–1746.

Borg, I., and Groenen, P. J. F. (**1997**). *Modern Multidimensional Scaling: Theory and Applications* (Springer, New York), pp. 37–38.

Charles, W.G. (**2000**). "Contextual correlates of meaning," Appl. Psycholinguist. **21**, 505–524.

Dünnwald, H. (**1991**). "Deduction of objective quality parameters on old and new violins," Catgut Acoust. Soc. J. **1**(series II), 1–5.

Ewins, D. (**2000**). *Modal Testing: Theory, Practice and Application* (Research Studies Press, Baldock, UK), pp 153–192.

Faure, A. (**2000**). "Des sons aux mots, comment parle-t-on du timbre musical? (From sounds to words, how do we talk about musical timbre?)," Ph.D. dissertation, Ecole des Hautes Etudes en Sciences Sociales, Paris, France.

Fisher, R. A. (**1915**). "Frequency distribution of the values of the correlation coefficient in samples of an indefinitely large population," Biometrika **10**, 507–521.

Fritz, C., Cross, I., Moore, B. C. J., and Woodhouse, J. (**2007**). "Perceptual thresholds for detecting modifications applied to the acoustical properties of a violin," J. Acoust. Soc. Am. **122**, 3640–3650.

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

Houtsma, A. J. M. (**1997**). "Pitch and timbre: Definition, meaning and use," J. New Music Res. **26**, 10–115.

Kendall, R. A., and Carterette, E. C. (**1992a**). "Verbal attributes of simultaneous wind instrument timbres. I. Von Bismarck's adjectives," Music Percept. **10**, 445–468,.

Kendall, R. A., and Carterette, E. C. (**1992b**). "Verbal attributes of simultaneous wind instrument timbres. II. Adjectives induced from Piston's orchestration," Music Percept. **10**, 469–502.

Krumhansl, C. L. (**1989**). "Why is musical timbre so hard to understand?," *Structure and Perception of Electroacoustic Sound and Music*, edited by S. Nielzén and O. Olsson (Elsevier, Amsterdam), pp. 43–53.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (**1995**). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psych. Res. **58**, 177–192.

Plomp R. (**1970**). "Timbre as a multidimensional attribute of complex tones," *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp 397–414.

Richards, V. M., and Zhu, S. (**1994**). "Relative estimates of combination weights, decision criteria, and internal noise based on correlation coefficients," J. Acoust. Soc. Am. **95**, 423–434.

Rugg, G., and McGeorge, P. (**2005**). "The sorting techniques: A tutorial paper on card sorts, picture sorts and item sorts," Expert Syst. **22**, 94–107.

Stepánek, J. (**2006**). "Musical sound timbre: Verbal description and dimensions," in *Proceedings of the International Conference on Digital Audio Effects (DAFx-06)*, Montréal, Quebec, Canada, pp. 121–126.

Stepanek, J., and Otcenasek, Z. (**2004**). "Interpretation of violin spectrum using psychoacoustics experiments," in *Proceedings of the International Symposium on Musical Acoustics*, Nara, Japan, pp 324–327.

Stepanek, J., and Otcenasek, Z. (**2005**). "Acoustical correlates of the main features of violin timbre perception," in *Proceedings of the Conference on Interdisciplinary Musicology*, Montreal, Canada, pp 1–9.

Traube, C., and D'Alessandro, N. (**2005**). "Vocal synthesis and graphical representation of the phonetic gestures underlying guitar timbre description," *Proceedings of the Eighth International Conference on Digital Audio Effects (DAFx"05)*, Madrid, Spain, September 20–22, pp. 104–109.

von Bismarck, G. (**1974**). "Timbre of steady sounds: A factorial investigation of its verbal attributes," Acustica **30**, 146–159.