Perceptual thresholds for detecting modifications applied to the acoustical properties of a violin

Claudia Fritz[*] and Ian Cross

Centre for Music and Science, Music Faculty, University of Cambridge, West Road, Cambridge CB3 9DP, UK

Brian C. J. Moore

Dept. of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK

Jim Woodhouse

Dept. of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, UK

(Dated: November 7, 2007)

This study is the first step in the psychoacoustic exploration of perceptual differences between the sounds of different violins. A method was used which enabled the same performance to be replayed on different "virtual violins", so that the relationships between acoustical characteristics of violins and perceived qualities could be explored. Recordings of real performances were made using a bridge-mounted force transducer, giving an accurate representation of the signal from the violin string. These were then played through filters corresponding to the admittance curves of different violins. Initially, limits of listener performance in detecting changes in acoustical characteristics were characterized. These consisted of shifts in frequency or increases in amplitude of single modes or frequency bands that have been proposed previously to be significant in the perception of violin sound quality. Thresholds were significantly lower for musically trained than for non-trained subjects but were not significantly affected by the violin used as a baseline. Thresholds for the musicians typically ranged from 3 to 6 dB for amplitude changes and 1.5 to 20% for frequency changes. Interpretation of the results using excitation patterns showed that thresholds for the best subjects were quite well predicted by a multichannel model based on optimal processing.

PACS numbers: 43.66.Jh, 43.66.Lj, 43.75.Cd, 43.75.De

I. INTRODUCTION

There is an extensive literature on the acoustics of the violin, and an even more extensive literature on human perception of sounds in general, and of musical sounds in particular. However, there is virtually no published research on the combined problem of the human capability for perception, discrimination and judgment of the sounds of violins with particular measurable acoustical properties. This is a very significant gap, since perceptual judgments must define what makes a violin different from other bowed-string instruments, and one violin different from another. To the best of the authors' knowledge this paper represents the first attempt to apply rigorous psychoacoustical techniques to a question of this nature.

The ultimate aim underlying the research presented here is to answer the typical question that a violin maker will ask: "Why does this violin sound better than this one", or more specifically "What will happen to the sound if I change such-and-such a constructional detail?" This paper starts the process of attacking that broad aim with a more modest target: to establish the just-noticeable difference for certain particular acoustical changes to the mechanical frequency response of a violin. These changes all relate to quantities previously proposed as significant to the sound quality of a violin. This initial investigation is of admittedly limited scope but is already of interest to instrument makers, telling them for example how far they need to move an individual low body resonance to have an audible effect. Since a violinist will explore a very wide range of bowing as well as types of musical input to a violin when judging its quality, the underlying philosophy of this study is to seek the input which results in the lowest perceptual threshold for each given acoustical change.

There are two stages necessary to such a study: to relate a constructional change to an acoustical change – i.e. a mechanical change of the vibrational properties – and to evaluate the perceptual effect of that change. There is already a significant literature concerned with the first stage (Cremer, 1985; Durup and Jansson, 2005). The experiments reported here concentrate on the second stage, that of establishing quantitative links between acoustical parameters of the instrument body and the perceptions of a listener.

The methodology of the study relies on the large impedance jump between the strings and the bridge of the instrument. The player manipulates the string to vibrate in certain ways, the vibrating string applies a force to the bridge and the body vibrates in response to this

[*]Electronic address: cf291@cam.ac.uk

force, radiating sound to the listener. To a first approximation, the body motion has little backward influence on the string motion. There are exceptions: most obviously the wolf note (Cremer, 1985; Woodhouse, 1993). More generally, if the topic of interest were the "playability" of the violin rather than its sound, then it would certainly not be admissible to ignore this back reaction. Similarly, if the study were concerned with the guitar or the piano, then string/body coupling would be crucial because it determines the decay rates of the various overtones of the string motion. However, for a bowed string such coupling effects can be ignored in the first instance. If strings of the same type are fitted to two different violins, a skilled player will adjust bowing to coerce the vibration into the standard Helmholtz motion with an acceptably short transient (Guettler and Askenfelt, 1997). The force waveforms acting at the bridge in the two cases will be very similar, and one would expect that the major differences in sound between the two instruments could be captured by driving them both with identical forcing waveforms. With this in view, representative force waveforms can be recorded using normal playing on a violin whose bridge is equipped with a piezoelectric force sensor under each string. These predetermined force functions can then be applied to different violins, so that sound differences can be compared with no complications arising from variations in playing.

Such a test could be carried out using different physical violins, applying the force at the bridge with a vibration shaker of some kind. However, for this study a different approach was taken. The mechanical frequency response function of the violin was mimicked using a digital filter, and the output signal for listening tests was generated by applying this filter to the recorded bridge force signal. Once the violin response is represented in digital filter form, it becomes very easy to make controlled variations of a kind which would be virtually impossible to achieve by physical changes to a violin.

Langhoff *et al.* (1995) conducted experiments in which violin performances were filtered digitally in a similar way. However, they equipped the violin with a velocity sensor on the bridge and could not accurately derive the force signal from the velocity signal. They used one violin as a baseline and then modified its frequency response curve (and therefore its impulse response) in several ways, to give enhancement of the Helmholtz resonance and of mid-range frequencies (around 1.7 kHz), and creation of a smoother decay towards higher frequencies. This experiment did show that it is possible to compare violin spectra by listening to digitally filtered signals but it did not address the question of how people perceived the different sounds created. Langhoff *et al.* (1995) only report the "subjective impressions" of one of the authors and no other participant was involved. Here we report the results of psychoacoustic measures of the ability of musically and non-musically trained subjects to discriminate changes in frequency and amplitude of single and multiple resonances.

II. STIMULI

A. Generation principle

In order to create the stimuli, we first recorded input signals (i.e. the force applied by the bowed strings on the bridge) during a live performance on a violin whose bridge was instrumented with a piezoelectric force sensor under each string. Second, we measured a suitable frequency response function for the two chosen violins (a modern instrument of good quality made by David Rubio, and a student-quality instrument used for comparison in the final stage of the work).

The choice of frequency response function raises some important technical issues. The most natural choice would be some kind of pressure response measured by microphone, in response to force applied to the violin bridge. However this choice would be unsuitable for two reasons. First, there is the question of variation with microphone position and the influence of room acoustics: no single measurement can be regarded as giving a "typical" sound. More serious is a theoretical issue. The essence of the experimental methodology is to make controlled changes in the frequency response function to generate the set of stimuli for testing. To make such changes, an explicit mathematical formula is needed for the frequency response in question in which the parameters to be varied appear explicitly. No such formula is known for radiated sound from a complex structure like a violin body. However, such a formula, in terms of mode shapes and natural frequencies, is available for a frequency response describing structural vibration rather than sound radiation (e.g. Skudrzyk, 1981). To take advantage of this, we have chosen to work with a mechanical frequency response, the input admittance function of the violin: this is defined as the ratio between the velocity at the string position on the bridge and the force applied at the same point. This function governs the energy transfer from string to body, and it is the most appropriate structural response for the present purpose. Of course, this raises important questions about whether different results might be obtained using radiated-sound response functions: these questions are currently being considered in ongoing research. The approach used here can be considered as being comparable to listening to a stationary violin with one ear from a fixed position in an unchanging room. These synthesized sounds are realistic enough for listeners to clearly recognize the sound as coming from a violin (for sound examples see (Fritz, 2007)) and for us to investigate relative changes.

The measurement procedure for input admittance was standard (e.g. Jansson, 1997): the bridge was excited with a miniature force hammer (PCB 086D80) at the G-string corner, and at the E-string corner velocity was measured using a laser-Doppler vibrometer (Polytec OFV056/OFV3001). This calibrated input admittance was then processed by modal identification techniques (e.g. Ewins, 2000), and resynthesized from the fitted parameters in the frequency range up to 7000 Hz, to allow parametric modification. To cover this range, 54 modes were needed for the Rubio violin. The frequency response could be modified by manipulating the modal parameters (amplitude, frequency and Q factor), and then used to construct FIR digital filters to obtain the "sound" of modified virtual violins. The filtering (i.e., the convolution of the input signal with the inverse Fourier transform of the frequency response of the violin in the time domain) was carried out using Matlab. It was found that the phase of the frequency response had no audible influence on the output sound, so for improved noise performance, non causal zero-phase filters were used.

B. Acoustical modifications to be tested

Informal tests showed only a very slight perceptual influence of the Q factors of the major modes, for changes up to 40%. Hence, it was decided to limit this initial study to measurements of thresholds for detecting increases in amplitude and frequency of one or several modes of the original admittance function, as described below.

1. Modes A0, B1- and B1+

At low frequencies, the sound of a violin is dominated by three strongly radiating modes. Several authors have suggested that these modes are important for sound quality (e.g. Hutchins, 1962). A0 is a modified Helmholtz resonance ('air mode'), which usually falls around 280 Hz. The two other modes are 'plate modes' which arise primarily from the bending and stretching of the front and back plates: B1- is usually centered in the range 470-490 Hz and B1+ between 530-570 Hz. Collectively B1- and B1+ account for what early researchers on the violin called the "main body resonance".

In the admittance, the A0 mode has a very small amplitude compared to the other low modes, B1- and B1+. However, it plays a very significant role in the radiated sound. To represent this effect approximately, its amplitude was artificially increased by a factor of 5 (14 dB), so that its amplitude was similar to that of B1-, as observed in radiated sound measurement in far-field (e.g. Dünnwald, 1991).

The amplitude and the frequency of each of these three modes were altered individually, as illustrated in Fig. 1 for an increase of frequency for B1+.

2. All modes in each of the four Dünnwald bands

Based on his measurement of the acoustical properties of a large range of violins that had previously been classified as of very good or moderate quality, Dünnwald (1991) proposed four frequency bands which he suggested

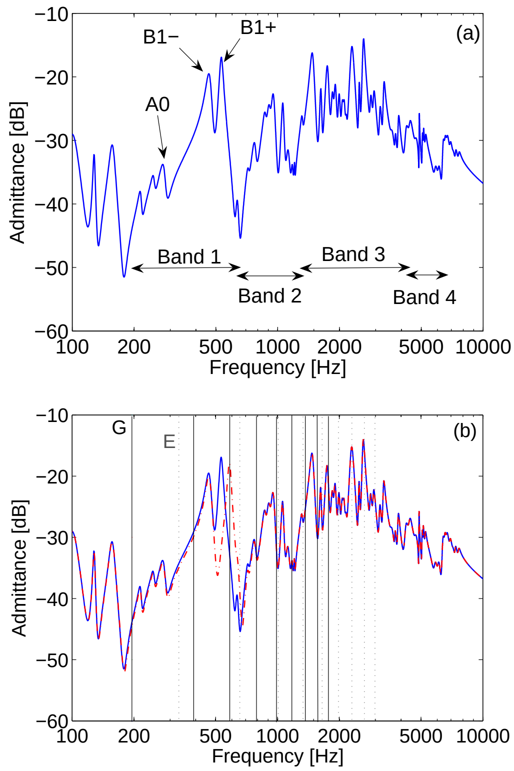

FIG. 1. (a) The resynthesized admittance of the Rubio violin, indicating the modes and the Dünnwald bands which were modified; (b) An example of a shift in frequency of a single mode: mode B1+ of the original admittance (solid curve) is shifted upwards by 14% (dash-dot curve). The positions of the harmonics are shown for each note by the solid vertical lines for G and dotted lines for E.

were important for the judgment of sound quality: 190-650, 650-1300, 1300-4200 and 4200-6400 Hz. The first range includes the lower overtones and may be related to "richness", the second he associated with "nasality", the third with "brilliance" and the fourth with "clarity". We measured detection thresholds for a change, either in amplitude or in frequency, of all the modes within each of these bands. For a shift in amplitude, as a few modes were close to the boundary between two bands, it was decided, instead of using a rectangular passband, to apply a gain function with a flat top and sloping edges (half of a symmetric Hanning window, of total width equal to 100 Hz, outside the nominal range of the band), as illustrated in Fig. 2.

For the frequency modification, the frequencies of all the modes within a given band were shifted by the same factor. However, it should be noted that, for a shift larger than about 20%, the modified modes overlap so much with the modes of the adjacent upper band that the change becomes very artificial.

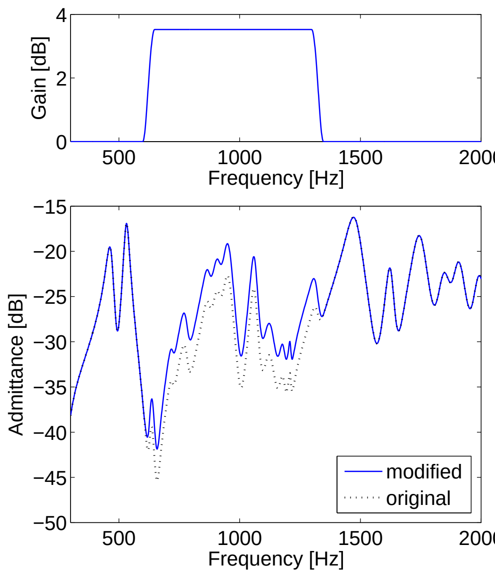

FIG. 2. Illustration of the method for modifying the amplitude of a single Dünnwald band. The gain function shown in the top panel applied to the original admittance function (dashed curve in bottom panel) gives the admittance function illustrated by the solid curve in the bottom panel: all the modes of the second Dünnwald band are shifted in level by 3.5 dB (amplitude increased by 50%).

3. All modes

In a further modification, the frequencies of all modes were simultaneously shifted by the same factor. This simulates, roughly, a change in the size of the violin body. This modification was not done in amplitude as this would only simulate a violin played harder, not a change in the violin body. It would in any case be nullified by the equalization of loudness, to be described shortly.

C. Input signal

In a preliminary study (Fritz *et al.*, 2006), performances of single notes and short phrases were played through filter sets corresponding to the measured frequency response curves of three violins, found from blind listening tests to be of very different sound quality. The phrases and single notes were tested for both discrimination and preference. Results showed that all listeners had lower thresholds for single notes than for musical phrases. Since the purpose of the present tests was to establish the thresholds for discrimination of changes which are the lowest that can be achieved under optimal test conditions, it was decided to use two single notes: G3 at 196 Hz and E4 at 330 Hz. The choice of these two notes results from the distribution of their harmonics. In particular, G3 has its second and third harmonics close to the center frequencies of modes B1- and B1+, whereas E4 has no harmonics near these modes. We wished to assess whether this would lead to poorer discrimination of changes in B1- and B1+ when E4 was used. The duration of the single notes was chosen to be relatively short (300 ms) because of echoic memory effects (see III.A). With such short notes, it is very difficult to tell how the violin was bowed and thus, the influence of the way of bowing is reduced.

D. Control of loudness

Large modifications of the modes, in particular of their amplitude, can lead to a change of loudness. One procedure for removing loudness cues is to randomly vary (rove) the overall level from one stimulus to the next. However, to completely eliminate such cues, a large rove range is required (Green, 1988), and this would have made the experience of listening to the violin sounds very unnatural. As an alternative method for ensuring that subjects would discriminate between the sounds on the basis of their spectral shape and not of their loudness, the overall level of each sound file was adjusted to keep the loudness level approximately constant at a value of 93 phons over the whole range of modifications. For each sound file (corresponding to a certain modification), this was achieved by first calculating the maximum value, M, over the duration of the sound of the short-term average loudness level, which was updated every millisecond, using the loudness model developed by Glasberg and Moore (2002). The overall loudness level did not change much over the duration of the sound so working with the maximum was similar to working with the mean. The sound amplitude was then multiplied by $10^{(93-M)/20}$. The result was that each sound file used in the experiment had a calculated loudness level of 93±1 phons. This may appear loud but it should be remembered that the sounds were broadband, and such sounds have a greater loudness than a narrowband sound of the same level (Moore, 2003). The sound levels used were below those typically experienced by violinists when playing (Royster *et al.*, 1991).

III. EXPERIMENT

A. Procedure

Thresholds were estimated using a three-alternative forced-choice procedure. A three-down one-up adaptive tracking rule was used which estimated the 79% correct point on the psychometric function (Levitt, 1971). Three sounds — two the same (the reference violin sound), one different (the modified violin sound) — were played in a sequence, and the subject was asked to choose which one was different. In order to allow echoic memory to operate effectively (Darwin *et al.*, 1972), the sounds and the inter-stimulus intervals were each 300 ms in duration. The amount of modification (either in frequency or in linear amplitude) between the stimuli in a given trial was changed by a certain factor (step size). Eight turnpoints were obtained. A relatively large initial step size of $2^{1/2}$ was applied until the second turnpoint was reached, in order to allow rapid convergence toward the threshold region. After the second turnpoint, the step size was reduced to $2^{1/4}$. Threshold was taken as the mean of the values of the amount of modification at the last 6 turnpoints.

For trials involving frequency shifts, the modification of the single modes was done by moving symmetrically from the original frequency; for instance, a shift of 10% was achieved by moving the center frequency of the mode by +5% and by -5%. This was done to reduce incidences of a shifted mode merging with an adjacent higher or lower mode, especially given that B1- and B1+ are quite close. During the listening test, the reference sound was thus not kept constant. In contrast, for the amplitude test, the reference sound was kept constant and equal to the sound of the original violin, in order to increase discriminability (subjects "learnt" to recognize the reference sound, as it was always the same) and the modification was an upwards shift in amplitude. This last method was used for all the other modifications.

Subjects were given visual feedback during the experiment but did not get any practice or any training beforehand. However, if they performed erratically on the first run, they were asked to do the run again. This applied only to a few subjects, and they were all able to perform the task at the second attempt. However, a few subjects had difficulties in performing certain conditions in the middle of the experiment and were given the opportunity to retry such conditions up to three times. Some succeeded, but not all. The conditions which could not be performed were not always the same among the subjects and this problem could happen even for subjects with low thresholds for other conditions. To avoid this problem, the initial amount of modification would have needed to be significantly larger. However, increasing all the initial values because of one or two people would have considerably increased the duration of the test for the others, and so a compromise was chosen.

The sounds were presented diotically via Sennheiser HD580 headphones, chosen because of their diffuse field response, in a relatively quiet environment. The sampling rate was 44100 Hz and the number of bits was 16.

B. Subjects

Three groups each of 18 subjects were selected according to their musical background. The first group had relatively little musical training (less than 6 years of formal training) and did not practice regularly: this group will be termed "non-musicians" in the following. The two other groups both had considerable musical training (more than 8 years of formal training) and practiced at least weekly. These last two groups were differentiated according to the instrument played: the violinists, viola players and cellists were in one group, and the remaining musicians in the other. Fifty subjects were between 18 and 40 years old, and the four others were between 50 and 60 years old. All subjects reported having normal hearing, although this was not checked. No systematic effect of age was observed in the results. Subjects were paid for their participation. These subjects were available to undertake the tests involving modifications in frequency and amplitude of the Dünnwald bands as well as the modification in amplitude of the single modes.

A different group of subjects was employed for the tests of frequency modification of the three single modes A0, B1+ and B1-. This group consisted of 9 non-musicians, 18 string players and 17 other musicians.

IV. RESULTS

The threshold results for all tests are summarized in Figs. 3 and 4. They are presented only for two categories of subjects, as initial analyses of variance (ANOVAs) did not show any significant difference between string players and other musicians, so the results for these two groups were combined. The average thresholds were calculated as the geometric mean since the standard deviation of the thresholds across subjects tended to increase with the mean value of the threshold. The variability of results among subjects was large, as is evident by some large error bars, representing $\pm$ one standard deviation. This variability was probably partly due to a lack of training. Individual differences could also be observed in the standard deviation of the turnpoints within a run. For the best subjects, the tracking variable went down directly to the region of the threshold and then oscillated closely around it, which gave a low standard deviation. For other subjects (with higher thresholds), the tracking variable fluctuated much more, giving a large standard deviation. Some people had a much lower threshold than average for one modification and a much higher threshold for another.

In the following, first the mean results will be presented and discussed. Then, to reduce the effect of the lack of training, the thresholds of the five best subjects will be presented and interpreted. We propose that these thresholds are close to the best that can be achieved, representing the limits of perceptual performance.

A. General comments on the statistical analysis

ANOVAs were run separately for the amplitude and the frequency modifications. Moreover, separate ANOVAs were performed for each type of modification, i.e., for the three single modes and for the four bands. The ANOVA based on data for the frequency modification of each of the four bands also incorporated, as an additional condition, the simultaneous modification of all modes.

For each of the analyses, the type of modification (the band or the single mode which was modified) and the note (G or E) were used as the dependent variables in a mixed (one between, two within) repeated-measures ANOVA. The between-group variable, the subject-type, had three values, non-musicians, musicians, and string players. The first of the two within-group variables, the modification, had either three or four values, while the second, the note, had two. In cases where the condition of sphericity was violated, we report values using the Huynh-Feldt correction.

The variance was not homogeneous for the modification in frequency, either for the Dünnwald bands or for the single modes. Therefore, the ANOVAs were conducted using the logarithms of the thresholds, which kept the violation of homogeneity at a level where ANOVA was still reasonably robust (for a given within variable, the ratio of the standard deviations across the three between-groups did not exceed 4).

For each of the four analyses, up to three subjects were sometimes removed as their data were incomplete: either they could not finish in the arranged time or could not do one of the tasks, even after several attempts.

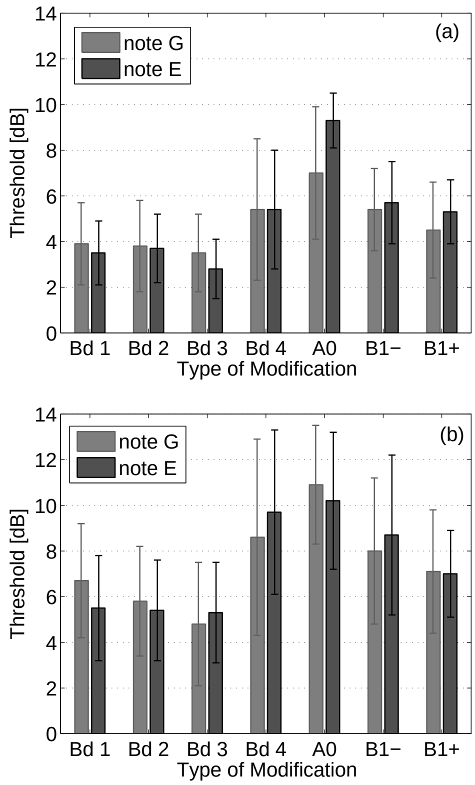

FIG. 3. Mean thresholds for detecting a modification in amplitude, expressed in dB. The upper (a) and lower (b) panels show results for musicians and non-musicians, respectively. The type of modification is indicated by the label under each pair of bars. Light and dark bars show results for the notes G and E, respectively. Error bars represent ± one standard deviation across subjects for each category of subject.

B. Influence of subjects' musical training

As mentioned above, the ANOVAs showed no significant difference between string players and other musicians, so these two groups are treated as a single group for subsequent analyses. However, musicians performed significantly better than non-musicians (all p values less than 0.001), except for the test involving modification in frequency of the single modes [$F(2, 38) = 1.7, p = 0.19$].

C. Thresholds

1. Modification in amplitude

Thresholds for both categories of subjects and for the two single notes G and E are given in Fig. 3. The x-axis represents the various conditions: amplitude modification of all modes of the ith Dünnwald band "Bd i" or of one single mode A0, B1- or B1+. The mean thresholds range from about 3 dB (musicians, band 3, note E) to over 10 dB (non-musicians, modification of A0, both notes).

The ANOVA of the results for the bands showed a main effect of group, with musicians performing significantly better than non-musicians [$F(1, 50) = 18.5,\ p < 0.001$]. There was a main effect of modification [$F(2, 100) = 6.3$, $p = 0.003$] but no main effect of note [$F(1, 50) = 2.6$, $p = 0.110$]. There was no significant interaction between modification and note [$p > 0.1$] but there was a significant interaction between group, modification and note [$F(2, 100) = 3.6,\ p = 0.03$]. This three-way interaction arises from the fact that non-musicians consistently performed poorly (compared to musicians) for each band with the note E, while musicians' performance with that note was better for band 3 than for either band 1 or band 2.

In the analysis of results for modes, there was a main effect of modification [$F(2, 98) = 83.9$, $p < 0.001$], and a main effect of note [$F(1, 49) = 4.7$, $p = 0.035$]. There was no significant interaction between modification and note [$p > 0.1$] but there was a significant interaction between note and group [$F(1, 49) = 14.8, p < 0.5$]. Non-musicians performed consistently poorly for both notes whereas musicians performed rather better (1 dB) for note G than for note E.

2. *Modification in frequency*

Fig. 4 shows corresponding results to Fig. 3 for the frequency-modification tests. A shift in frequency of the modes in Dünnwald band 4 was not detectable at all, so no threshold is given for this case nor is it included in the analyses below. "All" means that all modes were shifted in frequency. For this particular modification, a 2-s long musical phrase (the first two notes of the third theme of the Glazunov *Concerto for violin in A minor op.82*) was also tested (with the inter-stimulus interval kept to 300 ms) and the threshold obtained (also shown in Fig. 4) confirms what was shown in the preliminary study: people are less sensitive to subtle changes with a musical phrase than with single notes. This is consistent with the finding that the threshold for detecting a change in center frequency of a single formant in speech sounds (a formant corresponds to a resonance in the vocal tract) is higher when sentences are used than when isolated vowels are used (Liu and Kewley-Port, 2004).

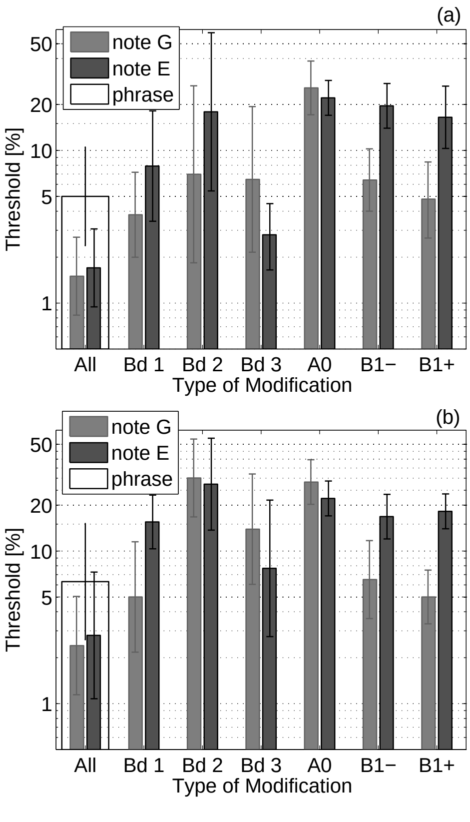

FIG. 4. As Figure 3, but showing thresholds ((a) for musicians and (b) for non musicians) for detecting a modification in frequency, expressed as a percentage of the center frequency and plotted on a logarithmic scale. The large open bars show thresholds when a musical phrase was used rather than a single note.

Remember that thresholds above 20% have no real meaning (see section II.B.2). A high threshold usually indicates that the corresponding modification was not perceptible.

The ANOVA for bands was based on the results obtained for Dünnwald bands 1, 2 and 3 and for the case when all modes were shifted simultaneously. There was a main effect of modification [$F(2.0, 97.7) = 118.1$, $\epsilon = 0.67$, $p < 0.001$], and a main effect of note [$F(1, 49) = 6.2$, $p = 0.016$]. There was a significant interaction between modification and note [$F(2.4, 118.4) = 26.4$, $\epsilon = 0.81$, $p < 0.001$] due to the fact that the average performance for musicians and non musicians was better for G than E when all modes were shifted and for bands 1 and 2, but this pattern reversed for band 3. There was also a significant interaction between note, modification and group [$F(2.4, 118.4) = 4.9$, $\epsilon = 0.81$, $p = 0.006$]. Musicians and non-musicians exhibited the same pattern of responses with respect to note when all modes were shifted and for bands 1 and 3, but not for band 2, for which musicians performed better for G than for E, while the results for non-musicians showed a slight trend in the opposite direction.

The ANOVA of the results for single modes showed a significant main effect of modification [$F(2, 78) = 130.4$, $p < 0.001$] and a main effect of note [$F(1, 39) = 124.7$, $p < 0.001$]. There was a significant interaction between modification and note [$F(2, 78) = 134.3$, $p < 0.001$], reflecting the fact that performance was better for note G than for note E for modes B1- and B1+ but not for mode A0.

The significant influence of note on the thresholds can be explained as follows. Note G has more harmonics in Dünnwald bands 1 and 2 than E, and G has its second and third harmonics close to B1- and B1+ which makes a slight change in the frequency of the corresponding modes much more noticeable for note G than for note E. Regarding the manipulation of band 3, it is not clear at first why there is such a difference between thresholds for E and G. However, some insight can be gained by calculation of excitation patterns, which can be defined as the relative response of the auditory filters plotted as a function of the filter center frequency (Moore and Glasberg, 1983). Excitation patterns calculated according to the procedure described by Glasberg and Moore (2002) (see Fig. 5) show that the difference between excitation patterns for the modified sound and the reference sound is much larger for E than for G for a given amount of modification, which explains why the threshold is larger for G. Note that the difference in excitation level between the reference and the modified sounds outside band 3 is due to the control of loudness. If the amplitude is lower in band 3, the amplitude has to be higher for other frequencies to give a constant overall loudness. This compensation effect is of course bigger when the difference in band 3 is large (which is why it is more noticeable for E than for G). A more systematic and quantitative study of excitation patterns is presented in section IV.E.

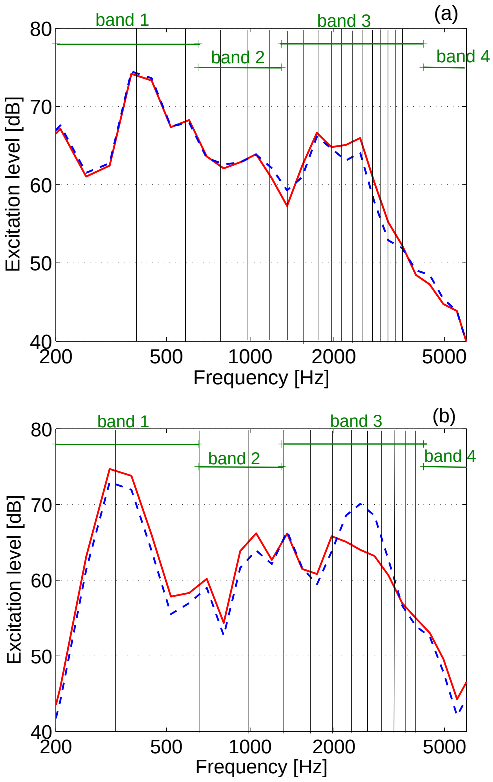

FIG. 5. (color online) Excitation patterns for notes G (a) and E (b), for the reference sound (dashed line) and the modified sound (solid line) when the third band was shifted in frequency by 5%. The vertical lines show the positions of the harmonics for each note.

D. Results based on the five best subjects

The results for each type of modification varied markedly across subjects. At least some of this variability arose from differences in musical experience. It probably depended also on whether or not subjects had previously taken part in auditory discrimination tasks, especially tasks requiring analytical listening. The subjects used here were given very little training before testing began. Previous research has shown that training can lead to improved performance for many aspects of auditory discrimination (Irvine and Wright, 2005). However, those subjects who initially show relatively good performance tend to show only small improvements with practice, while those who initially show relatively poor performance tend to improve markedly with practice (Fitzgerald and Wright, 2005; Irvine and Wright, 2005; Micheyl *et al.*, 2006; Moore, 1976). Hence it seems likely that the performance of the "best" subjects tested here would improve little with practice and would be representative of the thresholds that can be achieved by trained subjects attending to the optimal detection cues. To assess what this "best" performance was, we selected the five subjects who had the lowest average thresholds for the amplitude modification and the five who had the lowest average thresholds for the frequency modification (recall that different subjects were used for the two types of modification), and we determined mean scores for those subjects only. The results are shown in Fig. 6, together with predicted results which will be explained in section IV.E below.

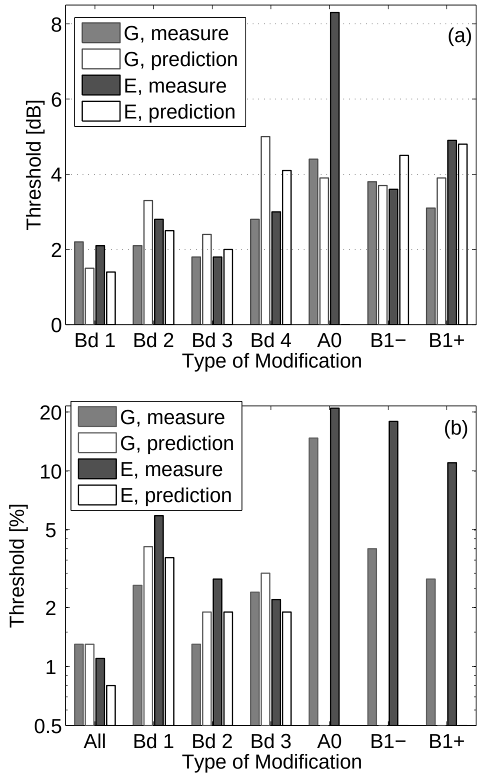

FIG. 6. Comparison between obtained mean thresholds for the five "best" subjects for each type of modification – (a) in amplitude and (b) in frequency – and the thresholds predicted by model 2. The upper and lower panels show results for the modifications in amplitude and frequency, respectively.

For the amplitude modification, the pattern of results is generally similar to that obtained for all subjects (Fig. 3), except that thresholds are lower. However, for the modification to A0 using note E, the mean threshold remained relatively high (about 8 dB) even for the five best subjects. The reason was found by looking more carefully at the spectra of the stimuli: the fundamental component of E corresponds to the trough lying above the A0 resonance, which is shifted in frequency when the amplitude of the resonance increases. Thus, the amplitude of the fundamental of E does not vary monotonically when the amplitude of A0 increases: the frequency of the trough moves from below the fundamental frequency of E to above it. Therefore, two different modifications can have the same effect on the amplitude of the fundamental component of E, if the corresponding troughs lie on either side of fundamental frequency of E. For modifications to bands 1-4, thresholds for the five best subjects were relatively small, in the range 2-3 dB, while thresholds for the single modes were mostly around 4 dB, again with the notable exception of A0 with the note E.

The thresholds for the amplitude modification obtained here can be compared to those obtained in experiments on "profile analysis" (Green, 1988). In such experiments, thresholds are measured for the detection of a change in level of a single "target" frequency component relative to the levels of other components, which form a kind of "background" or "profile". To prevent subjects from using the change in level of the target component as a cue, the overall level of the whole stimulus is randomly varied from one stimulus to the next. When the background contains a large number of equal-amplitude components, the threshold for detecting a change in relative level of the target is typically only about 1-2 dB (Green, 1988). However, if the components in the background do not have equal amplitudes, i.e., the profile is irregular in some way, the thresholds increase to 2-4 dB (Kidd *et al.*, 1986), values comparable to those found here. Our thresholds are also similar to thresholds for detecting a change in amplitude of a single formant in synthetic speech sounds (Pols, 1999).

For the frequency modification, the pattern of results for the five best subjects is generally similar to that obtained for all subjects (Fig. 4), except that the thresholds for the five best subjects are not markedly higher for band 2 than for the other Dünnwald bands. Thresholds are lowest (about 1%) for the modification to all bands.

The thresholds for the modifications to a single Dünnwald band are comparable to thresholds for detecting changes in the center frequency of a single formant in synthetic speech sounds (Kewley-Port and Watson, 1994; Lyzenga and Horst, 1997). The relatively low thresholds found here for discrimination of a change in all modes simultaneously are consistent with the finding that, for synthetic speech sounds, thresholds are lower when all formants are shifted together in the same direction than when only a single formant is shifted (Hawks, 1994).

E. Interpretation using excitation patterns

It is of interest to explore the extent to which existing auditory models of intensity and frequency discrimination might be capable of accounting for our results. Accordingly, we have attempted to model the results using three different, empirically grounded, auditory models.

Excitation patterns have been used as the basis of models for predicting the ability to detect changes in frequency and/or level of sounds (Florentine and Buus, 1981; Moore and Sek, 1992, 1994; Zwicker, 1956). In this section, we compare how well three different excitation-pattern models can account for the results obtained in the present experiment. For the analysis presented here, the excitation patterns were calculated with filter center frequencies spaced at 1-ERB_N intervals, using the ERB_N-number scale given by Glasberg and Moore (1990). To calculate excitation patterns from the waveforms of the sounds, we used the method described by Glasberg and Moore (2002). The excitation patterns were calculated at 50-ms intervals. The analysis that follows is based on the results of the five "best" subjects for each type of modification, as described in the previous section.

The models make use of the detectability index, d' (Green and Swets, 1974; Macmillan and Creelman, 1991). We assume here that the contribution to detectability in the ith frequency channel, d'_i, is proportional to the excitation level difference ΔL_i in the ith ERB_N between the reference sound and the modified sound, when the modified sound is at the threshold value. Because of this assumption we actually calculated a quantity for each model, D_1, D_2, and D_3, which was based on the ΔL_i values and was proportional to the value of d' for each model. The models differ in whether and how "information" is combined across channels, and therefore with our assumption, whether and how excitation level differences are combined.

The single-channel model (Zwicker, 1956, 1970) is based on the assumption that detection depends on monitoring the single place on the excitation pattern that changes the most:

$$D_1 = \max(\Delta L_i) \tag{1}$$

For this model, excitation patterns were determined every 50 ms and then averaged over time. The largest average excitation level difference given by a single channel was chosen as an estimate of D_1.

The multichannel model with optimal processing (Florentine and Buus, 1981; Moore and Sek, 1992)) is based on the assumption that information from different parts of the excitation pattern can be combined in an optimal manner:

$$D_2 = \sqrt{\sum_{i=1}^{n} (\Delta L_i^2)}. \tag{2}$$

For this model, D_2 was calculated every 50 ms and all quantities were then averaged to give the final estimated value of D_2.

The multichannel model without optimal processing (Moore and Sek, 1992) assumes that observers base their decision on an unweighted sum of decision variables (d'_i values) across all channels:

$$D_3 = \frac{\sum_{i=1}^{n} (\Delta L_i)}{\sqrt{n}}. \tag{3}$$

For this model, the averaging across time was done in the same way as for model 2.

For all three models, only "active channels", i.e., channels with excitation level above an assumed absolute threshold, were considered. The threshold excitation level was chosen to be 5 dB.

If any of the models worked perfectly, and there were no errors of measurement, then the calculated D values would be constant across all conditions except for the frequency modification of the single modes. This condition was conducted differently as the reference sound was not kept constant (see III.A) so this condition is not included in the following analysis. The modification in amplitude of mode A0 is excluded as well as this modification did not have a monotonic effect. The calculated D values were not constant for any of the models. However, we can assess which model is the best by evaluating, for each model, the coefficient of variation (CV) of the values of D, which is defined as the ratio of the standard deviation to the mean, across all conditions. The CV was 0.4, 0.3 and 0.4 for models 1, 2 and 3, respectively, so model 2 was marginally the best.

Using the mean value of D given by model 2, $D_2 = 6.1$, we can predict the threshold for each task which would give a value of D equal to this average value D_2. The predictions were generated by successive iterations to find the amount of modification which would give a value of D equal to D_2. Predictions are shown in Fig. 6. The root mean square (rms) deviation of the data from the predicted values is slightly less than 1 dB for the modifications in amplitude and 1% for the modifications in frequency, so the model is quite accurate and is equally good for G and E. This is particularly clear for the predictions obtained when all modes are shifted where, on average, both thresholds and predictions are lower than any that could be predicted on the basis of results obtained in response to shifts in individual bands. This suggests that listeners did indeed integrate information across a wide range of frequencies in a way which was close to optimal.

V. EXPERIMENTS WITH ANOTHER VIOLIN

All of the preceding results were obtained using only one violin. It is a possibility that the acoustical particularities of that instrument had a significant impact on listeners' responses. In order to test this, and also to test the robustness of the excitation-pattern model, a subset of the listening tests was repeated using a student violin judged to be of relatively poor quality, and which had acoustical characteristics that differed considerably from those of the Rubio violin. Its resynthesized input admittance is compared to that of the Rubio violin in Fig. 7.

It was decided to restrict this additional experiment to the shift in amplitude of all modes in each of the Dünnwald bands. The subjects for this second experiment were 15 musicians, among whom 5 had participated in the experiment with the Rubio violin.

Fig. 8 compares thresholds for the two violins. For the Rubio violin, thresholds correspond to the average across fifteen musicians, including the five musicians who also did the second experiment, and ten others chosen arbitrarily from the 31 other musicians. Although the thresholds tend to be lower for the "bad" violin (except for the modification of band 1 for note E, for which thresholds are equal), an ANOVA showed no significant difference between thresholds for the two violins ($F(1,28)=2.3$, $p=0.145$), as the standard deviations were quite large.

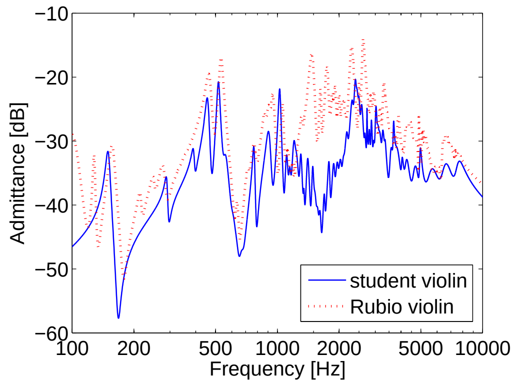

FIG. 7. Comparison of the resynthesized input admittances of the Rubio violin and the student violin of relatively poor quality.

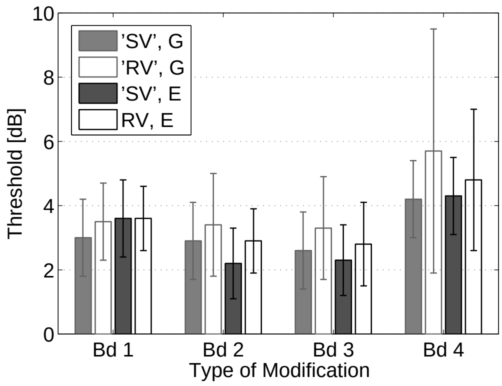

FIG. 8. Comparison of the thresholds obtained by musicians for a manipulation in amplitude of the modes within each Dünnwald band, for the Rubio violin ('RV') and the student violin ('SV'). See the key in the figure for details of the conditions. The error bars represent ±1 standard deviation across subjects.

Fig 9 shows a comparison between the predictions obtained with model 2, using the same value for D_2 and the average results for the five best subjects. The rms deviation is 0.8 dB so the fit is quite good.

VI. SUMMARY AND CONCLUSIONS

The work described in this paper represents the first stage of a project to provide quantitative information about the discriminability of and perceptual preferences between violins. The eventual aim of the project is to

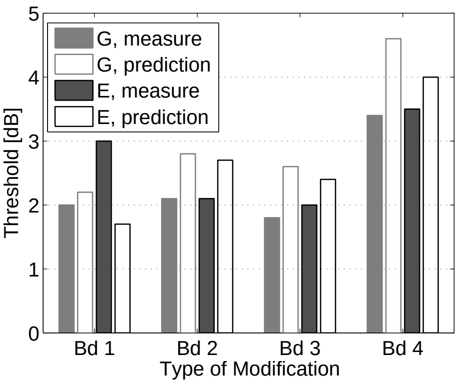

FIG. 9. Comparison between results and predictions of model 2, for the five best subjects, for a modification in amplitude of the student violin's modes. See the key in the figure for details of the conditions.

make direct links between the perceptual results and parameters relevant to instrument makers: materials choices, constructional geometry and set-up details. The results will also inform efforts to improve the quality of computer-synthesized string sounds.

This initial study explored two aspects of violin acoustics which have received great prominence in the earlier literature as possible indicators of aspects of 'quality': the three individual low-frequency modes of vibration (below 700 Hz), which dominate the sound of a violin and are usually labeled A0 (a modified Helmholtz resonance), B1- and B1+ (two strong 'wood modes'); and a set of four frequency bands proposed by Dünnwald (190-650 Hz, 650-1300 Hz, 1300-4200 Hz and 4200-6400 Hz) on the basis of measurements of a large number of violins of varying quality. Tests were conducted to establish thresholds for the perception of a change in frequency or amplitude of each of the three modes separately and for blocks of modes lying in the four "Dünnwald bands". Finally, a test was conducted in which the frequencies of all modes were varied simultaneously.

Results were presented for two groups of listeners: with and without extensive musical training. As might have been anticipated, the musically trained listeners had consistently lower thresholds. A series of ANOVAs was performed to investigate the significance of different main effects and interactions, in particular the influence of the type of modification and of the single note used as input on the thresholds. To obtain an estimate of the discrimination thresholds attainable by trained listeners attending to the optimal detection cues, results were calculated for the best five subjects in each group of tests. For modifications of amplitude, these 'best' thresholds were in the range 3-5 dB for individual modes and 1-3 dB for the Dünnwald bands. For modifications in frequency, the best listeners had thresholds around 3-5% for individual modes, 1-3% for the first three Dünnwald bands, and around 1% when all frequencies were varied simultaneously. Frequency changes in the 4th Dünnwald band were not detectable.

Predictions of threshold were made using three different models based on excitation patterns. The best performance was obtained using a multichannel model based on optimal combination of information across channels. This model reproduced the main results well, including the remarkably low threshold for detection of a simultaneous frequency shift of all modes. This success allows tentative predictions to be made for any combination of input signal and filter modification. This may allow a more systematic design of future tests, since input signals could be optimally chosen to allow a listener to best discriminate a given type of filter modification.

The best choice of stimulus sounds for these threshold tests proved to be very short single notes, probably partly because they allowed echoic memory to assist discrimination. When tests were repeated using a short musical phrase, higher thresholds were obtained. This finding is in some respects counter-intuitive, but is consistent with what is known about "informational masking"; it is difficult to detect a subtle change in a sound when the sound itself is varying strongly (Watson, 1987).

There is strong anecdotal evidence that certain subtle differences between violins can be perceived by violinists, and have great importance to them. It is sufficient to note that the market values of superficially similar violins range over some four orders of magnitude: in round numbers, from about \$100 to \$1000000. The authors are conscious of the fact that the tests described here are based on sounds which are very unmusical - our short single notes are barely recognizable as violin sounds. It would surely not be possible to obtain subtle judgments of quality and preference from such sounds. The likely conclusion is that the thresholds obtained here only tell part of the story of violin discrimination, and that higher-level perceptual processes are brought into play when a trained violinist compares instruments in a musical setting - for example during the process of choosing a new instrument.

There are many directions for possible further work along the lines explored in the study reported here. There are many more parameters which could be varied to establish perceptual thresholds. Of particular interest might be parameters with a direct interpretation in terms of a physical modification to a violin: for example, the parameters influencing the "bridge hill" (e.g. Woodhouse, 2005), or material properties of the wood used to build the violin body. Another important way of extending the study would be to use radiated-sound transfer functions rather than structural response (input admittance) as in this study. If suitable models can be formulated it would be possible, for example, to test the suggestion of Hill *et al.* (2004) in the context of the guitar, that low-frequency modes may have an important influence on the radiated sound well above their resonant frequencies, analogous to the broad-band sound radiation from a loudspeaker. Finally, informal tests have confirmed what one might guess, that a player is more acute and reliable at distinguishing two violins than is a non-playing listener. It would be instructive to repeat these tests using live playing on an electric violin, with sets of real-time digital filters instead of off-line filtered sound files.

VII. ACKNOWLEDGMENTS

This research was supported by the Leverhulme Trust. We thank Brian Glasberg and Hugh Greenish for their help in the design of the experiment, and Emily Smith for having conducted part of the experiment. We are also grateful to all the subjects who took part in this study. We thank two anonymous reviewers for detailed constructive comments.

Cremer, L. (**1985**). *The Physics of the Violin* (MIT Press, Boston, MA).

Darwin, C. J., Turvey, M., and Crowder, R. G. (**1972**). "An auditory analogue of the Sperling partial report procedure: Evidence for brief auditory storage", Cognitive Psychology **3**, 255–267.

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

Durup, F. and Jansson, E. V. (**2005**). "The quest of the violin bridge-hill", Acta Acustica - Acustica - **91**, 206–213.

Ewins, D. (**2000**). *Modal Testing: Theory, Practice and Application* (Research Studies Press LTD, Baldock, England).

Fitzgerald, M. B. and Wright, B. A. (**2005**). "A perceptual learning investigation of the pitch elicited by amplitude-modulated noise", J. Acoust. Soc. Am. **118**, 3794–3803.

Florentine, M. and Buus, S. (**1981**). "An excitation-pattern model for intensity discrimination", J. Acoust. Soc. Am. **70**, 1646–1654.

Fritz, C., Cross, I., Smith, E., Weaver, K., Petersen, U., Woodhouse, J., and Moore, B. C. J. (**2006**). "Perceptual correlates of violin acoustics", in *Proc. of 9th Int. Conf. on Music Perception and Cognition* (Bologna, Italy).

Fritz, C. (**2007**). "Experimental investigation of the perceptual correlates of violin acoustics", `http://www2.eng.cam.ac.uk/~cf291/project.htm`.

Glasberg, B. R. and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched-noise data", Hear. Res. **47**, 103–138.

Glasberg, B. R. and Moore, B. C. J. (**2002**). "A model of loudness applicable to time-varying sounds", J. Audio Eng. Soc. **50**, 331–342.

Green, D. M. (**1988**). *Profile Analysis* (Oxford University Press, Oxford).

Green, D. M. and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Krieger, New york).

Guettler, K. and Askenfelt, A. (**1997**). "Acceptance limits for the duration of pre-Helmholtz transients in bowed string attacks", J. Acoust. Soc. Am. **101**, 2903–2913.

Hawks, J. W. (**1994**). "Difference limens for formant patterns of vowel sounds", J. Acoust. Soc. Am. **95**, 1074–1084.

Hill, T. J. W., Richardson, B. E., and Richardson, S. J. (**2004**). "Acoustical parameters for the characterisation of the classical guitar", Acustica - Acta Acustica **90**, 335–348.

Hutchins, C. M. (**1962**). "The physics of violins", Scientific American **207**, 79–93.

Irvine, D. R. and Wright, B. A. (**2005**). "Plasticity of spectral processing", Int. Rev. Neurobiol **70**, 435–472.

Jansson, E. V. (**1997**). "Admittance measurements of 25 high quality violins", Acustica **83**, 337–341.

Kewley-Port, D. and Watson, C. S. (**1994**). "Formant-frequency discrimination for isolated English vowels", J. Acoust. Soc. Am. **95**, 485–496.

Kidd, G., Mason, C. R., and Green, D. M. (**1986**). "Auditory profile analysis of irregular sound spectra", J. Acoust. Soc. Am. **79**, 1045–1053.

Langhoff, A., Farina, A., and Tronchin, L. (**1995**). "Comparison of violin impulse responses by listening to convoluted signals", in *Proc. of International Symposium on Musical Acoustics* (Paris).

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics", J. Acoust. Soc. Am. **49**, 467–477.

Liu, C. and Kewley-Port, D. (**2004**). "Vowel formant discrimination for high-fidelity speech", J. Acoust. Soc. Am. **116**, 1224–1233.

Lyzenga, J. and Horst, J. W. (**1997**). "Frequency discrimination of stylized synthetic vowels with a single formant", J. Acoust. Soc. Am. **102**, 1755–1767.

Macmillan, N. A. and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (**2006**). "Influence of musical and psychoacoustical training on pitch discrimination", Hear. Res. **219**, 36–47.

Moore, B. C. J. (**1976**). "Comparison of frequency DLs for pulsed tones and modulated tones", Brit. J. Audiol. **10**, 17–20.

Moore, B. C. J. (**2003**). *An Introduction to the Psychology of Hearing, 5th Ed.* (Academic Press, San Diego).

Moore, B. C. J. and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns", J. Acoust. Soc. Am. **74**, 750–753.

Moore, B. C. J. and Sek, A. (**1992**). "Detection of combined frequency and amplitude modulation", J. Acoust. Soc. Am. **92**, 3119–3131.

Moore, B. C. J. and Sek, A. (**1994**). "Effects of carrier frequency and background noise on the detection of mixed modulation", J. Acoust. Soc. Am. **96**, 741–751.

Pols, L. C. W. (**1999**). "Flexible, robust, and efficient human speech processing versus present-day speech technology", in *Proc. XIVth International Congress of Phonetic Sciences ICPhS'99*, volume 1, 9–16 (San Fransisco, USA).

Royster, J. D., Royster, L. H. and Killion, M. C. (**1991**). "Sound exposures and hearing thresholds of symphony orchestra musicians ", J. Acoust. Soc. Am. **89**, 2793–2803.

Skudrzyk, E. (**1981**). *Simple and complex vibratory systems* (Pennsylvania State University Press, University Park, PA, U.S.A.).

Watson, C. S. (**1987**). "Uncertainty, informational masking, and the capacity of immediate auditory memory", in *Auditory processing of complex sounds*, edited by W. A. Yost and C. S. Watson (Lawrence Erlbaum Associates, Hillsdale N.J.).

Woodhouse, J. (**1993**). "On the playability of violins: Part 2. Minimum bow force and transients.", Acustica **78**, 137–153.

Woodhouse, J. (**2005**). "On the "bridge hill" of the violin", Acta Acustica - Acustica **91**, 155–165.

Zwicker, E. (**1956**). "Die elementaren Grundlagen zur Bestimmung der Informationskapazität des Gehörs (The foundations for determining the information capacity of the auditory system)", Acustica **6**, 356–381.

Zwicker, E. (**1970**). "Masking and psychological excitation as consequences of the ear's frequency analysis", in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg, 376–396 (Sijthoff, Leiden).