Multimodal Playback Device for Violin Sounds

Friedrich Beyer, M. Ercan Altinsoy
Institute of Acoustics and Speech Communication, Chair of Acoustics and Haptics, TU Dresden
E-Mail: friedrich.beyer@tu-dresden.de

Introduction

As with most musical instruments, playing a violin is a multimodal experience that combines auditory and tactile sensation. Tactile feedback from violins is provided by the three physical contact points to the instrument: left hand, chin and shoulder. Playing the instrument with a bow adds another contact point on the right hand, which is usually outside the scope of violin perception studies - as is the case here.

Previous studies have revealed that the felt vibrations at a violin have a strong influence on the player's perception of the instrument according to a number of perceptual attributes (e.g. responsiveness). [1–5] In these studies, a distinction was usually made between conditions with and without (or masked) or normal and amplified tactile feedback.

In a current research project, the influence of vibrations is being investigated in more detail, with the focus on spectral variations of the vibration signals. This research is based on a device that has the ergonomics of a normal acoustic violin (AV), but is able to control sound and vibration independently of each other - also referred to as "Virtual Violin".

In a first phase of the project, this device will only be used for non-interactive playback. Later, interactive experiments are planned in which the participants play the instrument themselves. Then the vibration signal will be processed in real-time.

The device will be based on an electric violin (because of a reduced inherent sound emission) equipped with one or more vibration actuators to generate the intended vibrations at the contact points. In this research project, audio feedback will be provided via headphones to keep the complexity of the device in scope. While in the first, non-interactive part it is possible to prepare the stimuli in advance, in the second part the device will be extended with a signal processing unit (i.e. DSP) to enable real-time processing.

The present work evaluates the proposed method for generating vibration signals by means of attached actuators and investigates:

(I) whether it can reproduce a target vibration signal accurately.

(II) whether it is perceptually valid, tested in a multimodal perception experiment.

The test case for the second research question is not to vary the spectrum of the vibration signal, but to investigate whether a recorded signal of an AV can be reproduced good enough that it is recognized as a real signal by subjects. In the work presented here, the scope of

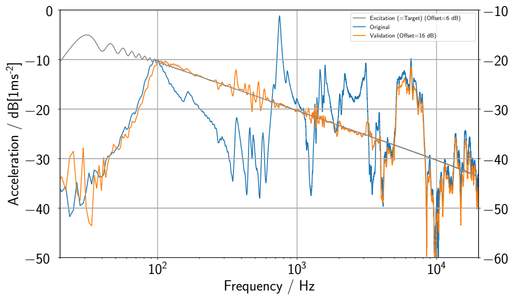


Figure 1: Spectra of excitation voltage signal (gray), measured acceleration at hand position (blue) and measured acceleration resulting from filtered log-sweep (orange).

application is limited to non-interactive cases. Also the vibrations are considered only at the hand position.

Setup

In the study presented here, the playback device was the AV itself, which was to be imitated. Therefore, an AV (*Thomann Classic* with *Wittner* chin rest and *Kun* shoulder rest) was used. Vibrations were generated with a *Tactile Labs MCC-1* actuator (connected to a *t.amp PM40C*) attached to the violin at the head stock. Audio signals were reproduced via headphones (*Sennheiser HD600 + HEAD Acoustics PEQ-V*), which was relevant in the perception experiment. The signals were prepared in a Python script and were played back by *Reaper* via an *RME Fireface UCX* audio interface.

Actuator Calibration and Filter Design

The vibration signal was to be optimized at the hand position (target). As the actuator could no be placed directly at this position (colliding with the hand), the location offset needed to be considered. Therefore, a filter was designed to be applied to the target vibration signals in order to compensate for the transfer function (TRF) between actuator and target position. In addition, also the response of the actuator was equalized with the filter. For this purpose, a log-sweep measurement (20 Hz - 20 kHz) was carried out, to retrieve the TRF between the excitation voltage at the actuator terminals and the acceleration at the target position. [6] For the measurement an *MMF KS901.100B* accelerometer and a *HEAD Acoustics Squadriga III* measurement front-end were used.

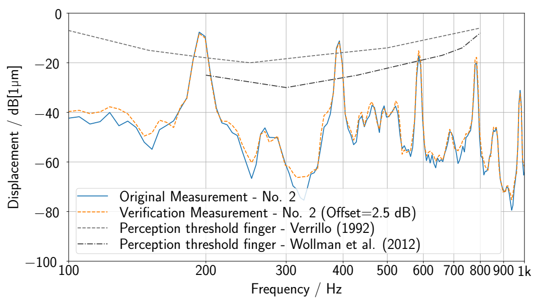
A linear-phase inverse FIR filter was calculated from the TRF [7], the application range of which was limited to the frequency range of 100 Hz to 4 kHz (outside this range the filter response was 1). To validate the filter, another TRF measurement was made with a filtered excitation signal. As depicted in Figure 1, the filtering could equalize the acceleration to a great extent.

In both measurements, the violin was held in playing position so that the same contact and damping points were present as during normal playing. This is at the expense of a not completely identical violin position throughout different measurements and can explain some of the errors.

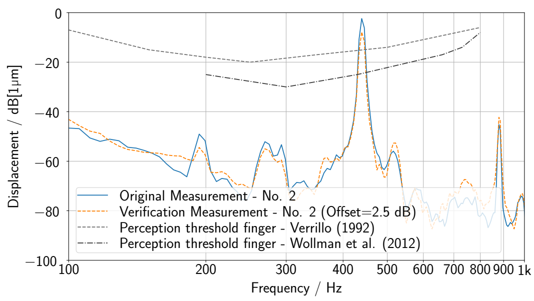
I Vibration Measurement Comparison

In order to compare the vibrations at the hand position caused by a bow-played string with the vibrations reproduced by the actuator, two sets of recordings were made. First, the four open strings were played with the bow in normal holding position while the vibrations were recorded with the measurement setup from before. Each string was recorded twice. Additionally, a *HEAD Acoustics BHS-II* headset was worn by the player to record the audio signals arriving at the ears, in order to be used in the second experiment.
The recorded vibration signals of the first set were at the same time the target signals that needed to be recreated. They were therefore filtered in the next step using the developed filter. A second set of recordings was then made, with the filtered target signals played back via the actuator. Now only the acceleration was recorded.

Results



(a) G-string



(b) A-string

Figure 2: Vibration spectra of open string recordings at hand position, either played with the bow (blue) or reproduced with an actuator (orange). (a) shows a pair of G-string recordings as GOOD example, (b) shows a a pair of A-string recordings as BAD example.

As it is difficult to observe deviations between the time signals, the comparison of each pair of recordings is done in the frequency domain. Figure 2 shows the spectra of recordings of two different strings. In both cases, the original recording of an open string played with the bow (blue) is compared with the corresponding recording of vibrations played via the actuator (orange). In addition, sensation thresholds from the literature were plotted as a reference [8, 9]. These literature references also indicate that vibrations are usually not perceptible at frequencies above 1 kHz, which is why the following plots are truncated at 1 kHz. The lower frequency limit was set to 100 Hz which is well below the lowest possible frequency of 196 Hz for violins (fundamental frequency of the lowest string).
Figure 2(a) depicts the spectra of the recordings of the G-string. It shows, that the deviations between the recordings are very small. In particular, the vibration amplitudes of the tonal components (196 Hz and 392 Hz) match very well. The noise floor between the peaks can be neglected, as the levels here are far below the perception threshold. Figure 2(b) shows the results of an A-string recording with a less optimal playback performance. Here, the amplitude of the fundamental (440 Hz) cannot reach the original amplitude, while the playback amplitude of the first harmonic (880 Hz) exceeds the original amplitude. This may be due to a non identical holding position, but can also be related to structural problems at this frequency.
The deviations of the other recordings lie in between the shown examples. In general,the deviations for the four open strings are less then 3 dB (in the range of just noticeable level difference [10]) for all fundamentals and for most harmonics in the frequency range of interest - with the exception in 2(b). Thus, the concept can be accepted based on acceleration measurements. It should be noted that this study only includes open strings. Things might change if a string is stopped with the fingers of the left hand to play other notes.

II Subjective Comparison

As seen in the previous section, the concept seems very promising on the basis of measurements. In the following, it will be examined whether this also applies on a subjective level.

Method

A perception experiment was carried out to test whether subjects could distinguish between real bowing and playback by an actuator. The subjects were blindfolded during the experiment and were asked to hold the "Virtual Violin" (violin with actuator + headphones)(see Fig. 3). Even though the open headphones (*Sennheiser HD600*) were assumed to be acoustically transparent external sounds were attenuated at high frequencies. Since real stimuli coming from outside the headphones were to be compared with playback stimuli, the latter had to be adjusted at high frequencies by applying a high-shelving filter ($f_c = 2.65\,\text{kHz}$, $gain = -12\,\text{dB}$, $BW = 1\,\text{oct.}$) to the audio signal.
Throughout the experiment samples of different open strings were played either by the actuator (hereafter *virtual*) or with the bow by the experimenter (hereafter *real*). In each trial, one stimulus was played without an option of repeating it. After each trial, the subjects had to decide whether the stimulus was *real* or *virtual*. Between trials, the bow was always lifted and placed back on the string so that it was not possible to detect *vir-*

Figure 3: A subject holding the "virtual violin", which is played by an experimenter.

tual stimuli by a missing bow interaction. During *virtual* stimuli, the bow was held on the played string to prevent it from resonating.

Stimuli

The set of stimuli consisted of the 8 recordings from before, used as *virtual* stimuli after applying the equalization filter. In addition, for each *virtual* stimulus, a *real* stimulus of the same string was added to the set. Thus, there were 8 stimuli played with the bow in the set, two of each open string. Each stimulus was repeated three times, resulting in a total of 48 stimuli. It should be noted that it was not possible to reproduce the *real* stimuli completely identically, so there were 3×2 slightly different representations of each bow-played string. In comparison, there were 2 different versions for the *virtual* stimuli, each repeated three times. The set of stimuli was randomized for each subject before the start of the experiment.

A Python program carried out the randomization and guided through the experiment. It included a graphical user interface to tell the experimenter which string and whether a *real* or a *virtual* stimulus was to be played. Depending on the subject's decision, "r" (for *real*) or "v" (for *virtual*) was entered into the program whereupon the next trial was started.

Subjects

A total of 14 subjects (average age: 32.9 years) took part in the experiment, without any special requirements to their profile. Among the subjects, two were experienced amateur violinists, further six were playing other music instruments. Eight of the subjects had experiences in vibrations and tactile perception (including perception experiments), intersecting with the musically trained people. Two were completely naive.

Results

To evaluate the results, the rate of correct answers - or detection rate - is investigated. If 50% of the answers were correct, subjects could not distinguish between *real* and *virtual* stimuli at all. If 100% were correct, perfect distinction was possible. The same would apply to 0% correctness, even though in this case the mapping would be wrong for some reason.

The overall detection rate of the present sample was found to be **58.8%**. This appears to be a good tendency, even if it is not equal to chance level. A Binomial-Test

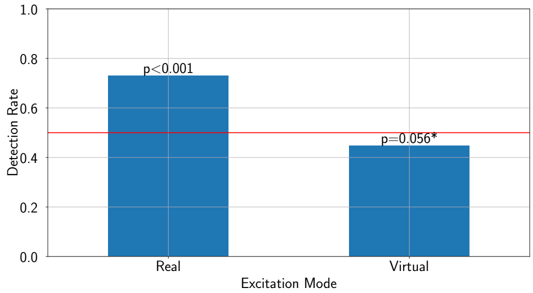


Figure 4: Detection rate of the data set grouped by excitation mode.

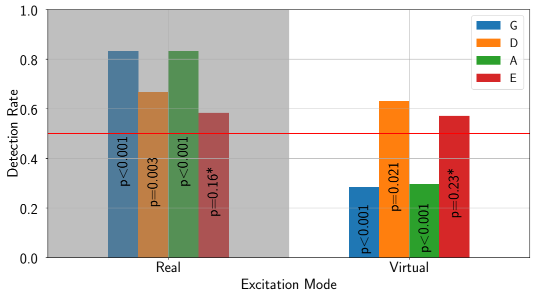


Figure 5: Detection rate of the data set grouped by excitation mode and excited string.

(H_0: proportion == 0.5, $p < 0.001$) confirms that it is unlikely that the detection rate of the entire population can be assumed to be at chance level.

A closer look at the data, grouping it by excitation mode, shows more promising results (see Fig. 4). Obviously, *real* stimuli were detected much better than *virtual* ones. This could be due to the methodology of the experiment involving an experimenter who cannot perfectly reproduce the stimuli and could introduce further errors. Although the experiments were conducted very carefully, it was not possible to completely suppress small side noises. To mitigate this, the bow was also put on the strings in *virtual* trials, as mentioned before.

Regarding the *virtual* stimuli, the results are closer to the 50% line, and can be interpreted as statistically significant (Binomial-Test $p > 5\%$). In addition, the subjects gave wrong answers for *virtual* stimuli in more than 50%. Thus, they interpreted them as *real* stimuli, which supports the idea that "real-appearing" stimuli can be generated with this method.

In Figure 5 the level of detail was increased by one more step and now also groups the data by the string that was excited. Focusing on the *virtual* stimuli (right side) shows that for G- and A-string the results are even more promising, as more than 70% of the answers are actually interpreted as *real* stimuli. The D- and E-string come close to chance level. For the E-string, the same is seen for *real* stimuli. This could be due to the fact that the tactile feedback no longer has influence on this string, as the fundamental frequency (659 Hz) is almost outside the perceptual range and the level is around or even below the perceptual threshold. Therefore, in these trials the experiment is constraint to a comparison in the auditory perception, whereby the audio playback was binaurally recreated very accurately with the aforementioned setup (this was also confirmed in comments of the subjects, see below). The fundamental frequency of the D-string is much lower and is actually within the most sensitive frequency region for tactile sensations. However, looking at the displacement spectrum (not depicted here) revealed that the level of the fundamental frequency is near or below the threshold. Therefore, a similar reasoning applies here as for the E-string.

Comments of Subjects

The participants had the opportunity to comment on the experiment after it was completed. A frequent comment was that the task was difficult and the distinction was rather based on guessing (4x). One subject said that everything was feeling real. Another common comment was that vibrations were not as important as the auditory feedback (4x). The audio feedback itself was rated as very convincing.

Only in a few cases structural problems were reported. In one case the virtual acoustic cue was perceived too far on the left side, indicating inaccurate placement of the instrument. Two subjects stated that the *virtual* stimuli decayed unnaturally, while another one claimed that the real bow-playing was less smooth. One subject reported feeling more transients in *real* trials.

Finally, it was reported that *real* stimuli are more likely to be perceived in the shoulder or closer in general (2x), which may be due to the fact that chin and shoulder vibrations were not controlled by the playback device in the current study. However, at least for one of the subjects who reported this, the detection rate was worse than average (48%) and hence probably not an effective cue. Lastly, it was reported that stimuli with vibrations were considered more real (2x). It should be noted that vibrations were always present in the current experiment.

Finally, there were also general comments stating that there was a lack of experience with violin playing (2x) or that the experiment was too long and unpleasant (1x).

Summary

In summary, the results showed that in average a distinction between *real* and *virtual* stimuli was possible only to a limited extent. Considering only *virtual* stimuli, the detection rate almost reached chance level or even tended towards lower rates for some strings. This is positive for the intended application in playing back *virtual* stimuli, that are supposed to feel *real*. Thus, this concept can basically be used in upcoming experiments.

Conclusion & Outlook

In the work presented, the concept of a "Virtual Violin" was introduced, which can independently reproduce or modify the tactile and auditory feedback of a violin. Two experiments were presented in which the validity of this concept was evaluated for non-interactive scenarios. First, acceleration measurements were carried out at the hand position on the neck of a violin to compare real bow-played and playback vibrations. It was shown that the resulting spectra matched very well for most recordings. Second, a perception experiment was carried out with 14 participants. In this experiment, the subjects were asked to distinguish between real bow-played and playback stimuli in a multimodal context (audio-tactile). The results showed that discrimination performance between real and virtual stimuli was very limited (58.8% correct). When focusing only on virtual stimuli, the detection rate dropped to more or less chance level for two strings and even lower for the other two strings, indicating that virtual representations of these strings were considered real stimuli. Hence, it was concluded, that the concept is a valid method for further research.

In the future, the "Virtual Violin" will be extended to also control chin and shoulder contact vibrations. Furthermore, the system will be ported to an electric violin. Perception experiments are then carried out investigating the influence of different vibration feedback on perceptual attributes.

Acknowledgments

This project is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 493698227.

The authors want to thank all subjects that took part in the experiment.

References

[1] Saitis, C., Fritz, C., Scavone, G. P., Guastavino, C., Dubois, D.: Perceptual Evaluation of Violins: A Psycholinguistic Analysis of Preference Verbal Descriptions by Experienced Musicians. J. Acoust. Soc. Am., 2017, 13.

[2] Wollman, I., Fritz, C., Poitevineau, J., McAdams, S.: Investigating the Role of Auditory and Tactile Modalities in Violin Quality Evaluation. PLOS ONE 9, no. 12 (April 12, 2014)

[3] Wollman, I., Fritz, C., Poitevineau, J.: Influence of Vibrotactile Feedback on Some Perceptual Features of Violins. The Journal of the Acoustical Society of America 136, no. 2 (August 1, 2014)

[4] Fritz, C., Muslewski, A., Dubois, D.: A Situated and Cognitive Approach of Violin Quality. 2010, 5.

[5] Flueckiger, M.: Vibrotactile Feedback in Piano and Violin Playing: Actuation, Interaction, and Sensation. Doctoral Thesis, ETH Zurich, 2018.

[6] Farina, A.: Advancements in Impulse Response Measurements by Sine Sweeps. 2007, 21.

[7] Smith, S. W., The Scientist and Engineer's Guide to Digital Signal Processing, Ch. 17.

[8] Wollman, I, Fritz, C., Frelat, J.: Vibrotactile Feedback in the Left Hand of Violinists. 2012, 6.

[9] Verrillo, R. T.: Vibration Sensation in Humans. Music Perception 9, no. 3 (April 1, 1992)

[10] Israr, A., Tan, H., Reed, C.: Frequency and Amplitude Discrimination along the Kinesthetic-Cutaneous Continuum in the Presence of Masking Stimuli. The Journal of the Acoustical Society of America 120 (December 1, 2006)