Augmenting Sonic Experiences Through Haptic Feedback

(Article begins on next page)

03 May 2024

Chapter 12
Augmenting Sonic Experiences Through Haptic Feedback

Federico Fontana, Hanna Järveläinen, and Stefano Papetti

Abstract Sonic experiences are usually considered as the result of auditory feedback alone. From a psychological standpoint, however, this is true only when a listener is kept isolated from concurrent stimuli targeting the other senses. Such stimuli, in fact, may either interfere with the sonic experience if they distract the listener, or conversely enhance it if they convey sensations coherent with what is being heard. This chapter is concerned with haptic augmentations having effects on auditory perception, for example how different vibrotactile cues provided by an electronic musical instrument may affect its perceived sound quality or the playing experience. Results from different experiments are reviewed showing that the auditory and somatosensory channels together can produce constructive effects resulting in measurable perceptual enhancement. That may affect sonic dimensions ranging from basic auditory parameters, such as the perceived intensity of frequency components, up to more complex perceptions which contribute to forming our ecology of everyday or musical sounds.

F. Fontana (✉)
Department of Mathematics, Computer Science and Physics, University of Udine, via delle Scienze 206, Udine 33100, Italy
e-mail: federico.fontana@uniud.it

H. Järveläinen · S. Papetti
Institute for Computer Music and Sound Technology, Zurich University of the Arts, Pfingstweidstrasse 96, Zurich 8005, Switzerland
e-mail: hanna.jarvelainen@zhdk.ch

S. Papetti
e-mail: stefano.papetti@zhdk.ch

12.1 Introduction

During a sonic experience, humans give meaning to what is being listened to, based on their perception and cognition of the auditory scene. As other sensory channels normally convey stimuli in parallel to hearing, the human brain integrates a continuous flow of sensations while contextualizing the experience. If, on the one hand, vision, smell, and taste concur in describing an auditory scene, thanks to high-level connections involving our mental imagery [14], on the other hand, touch is often exposed to temporal patterns that exhibit a strong affinity with the acoustic signals hitting the eardrum with respect to their synchronism, amplitude, spectral content, and mutual localization. This similarity is evident, for instance, when a musician plays an instrument, and more in general whenever a human action generates an event producing sound as a (by-)product.

Our chapter is about whether the somatosensory feedback consequence of that action contributes to *augment* the sonic experience. Here, the term augmentation embraces all sorts of enrichment that a sonic experience would benefit from through the somatosensory channel, whether it makes a perceived sound stronger, clearer, more vivid, meaningful, pleasant, or ecologically valid. Such a variety of effects, affecting sound ranging from fundamental physical dimensions until its semantics, can be explained by the tight interactions that sound and vibration establish with one another, as soon as our brain associates them both with a unique event. Understanding such interactions and their effects is the main goal of scientists who investigate the psychophysics of auditory-tactile perception.

Perception psychologists were able to isolate the role of touch, especially during passive auditory tasks. Such tasks in fact lead to generally more robust design, control, and repeatability of the experiments. For this reason, the reference literature introducing this chapter deals mainly with passive touch. However, the most interesting sonic augmentations in an ecological or musical sense involve perception-action loops, in which the listener physically interacts with a sounding object. In the case of active exploration, or when a device reproduces tactile cues, the sense of touch conveys *haptic* feedback. Accordingly, our chapter will focus on effects reported by active listeners, as well as on sonic (either ecologic or musical) experiences resulting from passive tasks in the presence of various haptic interfaces.

12.1.1 Multisensory Processing of Touch and Audition

Multisensory processing—the convergence of information from various sensory channels—happens both in early cortical stages and in high-level structures. These processes can either enhance or depress response relative to the most robust unisensory information. This multisensory integration benefits feature integration, object processing, event detection, and decision-making especially when cues are weak or ambiguous [16, 54] (please refer also to Chap. 10 for a bigger picture on this

topic). There is ample evidence of integration and interaction between the senses of hearing and touch. While somatosensory influence on higher auditory structures is well-known, evidence of low-level influence is more recent and increasing [30]. The cochlear nucleus in the brainstem responds to both somatosensory and auditory stimulation; this way somatosensory input may influence both sound lateralization and the suppression of self-generated sounds [10, 53]. The first cortical stages—previously thought to process unimodal sensory information—are now known to converge and sometimes process heteromodal information. The primary and the belt areas of the auditory cortex receive inputs from various low-level somatosensory areas, while fewer reports point to pathways from auditory to somatosensory areas. Higher-level multisensory areas that process auditory and somatosensory information include the *Superior temporal cortex* and the *Insular cortex* [15].

However, much of the multisensory integration that is necessary for the identification and localization of events takes place in the *Superior Colliculus* (SC), which is located in the midbrain: several subcortical and primary cortical areas project auditory, somatosensory, and visual information to this area. The neurons in the SC can respond differently to cross-modal stimuli than to either of the respective unimodal stimuli. Information is integrated according to a few general principles: spatially and temporally coherent stimuli produce maximal enhancement, and weaker stimuli produce a relatively greater enhancement (inverse effectiveness) [47].

Similar observations have been made on behavioral level: sounds and vibrations have been shown to interact constructively when congruent stimuli are delivered simultaneously [56, 57], with measurable auditory effects of somatosensory feedback [4, 36, 37, 39, 51, 52, 61]. Here congruence is defined depending on the experimental procedure: in general, it refers to conditions in which the multisensory stimulus shares common spatio-temporal as well as spectral features, as if it was originating from a unique source producing sounds and vibrations together. In parallel, simultaneity refers to a stimulus pair whose acoustic and vibratory components are rigorously constrained concerning their mutual synchronization: audio-tactile temporal resolution is superior to audio-visual or visuo-tactile combinations [20]. In this regard, it must be kept in mind that hearing and touch are both very sensitive to temporal delays, and detect especially low latency values relative to each other. By varying these values in the range 5–70 ms, Kaaresoja et al. have been able to change the perceived quality of virtual buttons during a clicking gesture [29]. More in general, mutual unsynchronization and/or delocalization of the acoustic and vibratory components leads to disparate effects that must be dealt with case by case [50], revealing the complexity of audio-tactile interactions. As this chapter focuses on haptic feedback, we will instead describe experiments where stimuli are simultaneous and co-localized.

Spatial collocation seems in fact somewhat less critical than temporal synchrony, judging by the presence of audio-tactile interactions and enhancement in many experiments where participants receive vibrotactile feedback through the hand and auditory stimuli through headphones [31]. Nevertheless, humans have good spatial discrimination ability between auditory and tactile stimuli: lateral angles of $\geq 5.3°$ were detected between electrotactile stimulation at the fingertip and sound source in an

experiment by Altinsoy [2]. (To put this in context, auditory localization blur for the scraping sounds used as stimuli in the experiment was 3.9°.) Indeed, the seeming failure of some studies to demonstrate spatial modulations of audio-tactile interactions may be due to the fact that stimuli have been presented at hands or otherwise at some distance from the head; more recently, spatial modulation effects have indeed been observed especially in the space close to the head [31]. However, these phenomena are not thoroughly known yet; note that in the peripersonal space, even unimodal auditory localization differs from that at greater distances [6–8].

The psychophysical literature specifically dealing with the effects of touch on auditory perception is sparse, mostly focusing on intensity and pitch as primary objects of investigation. As opposed to the previously described constructive effect valid for multisensory cues of intensity, the interactions between auditory pitch and tactile frequency discrimination are more complex [5, 59]. In particular, tactile frequencies do not need simultaneity nor co-localization to affect pitch perception [62]. As part of their study on the audio-tactile pitch and loudness interactions, Yau et al. found separate mechanisms for tactile influence on loudness and pitch, with audio-tactile loudness perception depending more on the timing of the stimuli [60]. Anyhow, pitch is perceived much more accurately through the hearing system, hence touch in general plays no supportive role during the perception of frequency components in an audio-tactile signal. Still, tactile frequency discrimination ability has been ascertained [23, 55], with surprising accuracy in congenitally deaf individuals [35]. This evidence naturally leads to the question about musical sensations induced by touch, an issue which has fascinated several scientists [49] and, hence, occupies an important part of this chapter.

Some deaf musicians show an indisputable ability to "feel the vibrations" during music performance, not merely for entraining with other musicians [12, 22, 25, 26], but also for sharing melody and timbre with them. This ability seems to be the result of the long training *any* (i.e., including the normally able) good musician has accumulated with all senses on their instrument [34] during a continuous perception-action process. Such a training, hence, refines a multisensory acuity for the instrument quality, not limited to its sound [21, 58].

Non-musicians can also discriminate musical timbre and relative pitch intervals from vibrotactile cues, to some extent even without training [24, 49]. However, generalizing the above-mentioned higher-level phenomena to musically untrained individuals is not obvious [3]. Being inherently psychophysical, there is no reason to think that the summation of auditory and tactile cues of intensity would not apply to non-musicians. In parallel, musical training seems to facilitate more subtle audio-tactile synergies mediated by higher nervous system levels, such as those linking pitch and tactile frequency recognition [11, 33]. Amid these two facts, the possibility for touch to enable the detection in normal listeners of frequency components otherwise inaudible, due to masking or threshold effects, is yet to be systematically explored.

Table 12.1   Key characteristics of the experiments forming the chapter

Experiment	Conditions	Input gesture	Haptic feedback	Audio feedback	Results
Ball bouncing on everyday materials [13]	Passive	Finger-pad contact	Recorded stimuli	Recorded stimuli	Audio-tactile summation leads to best identification
Reproduction of target pressing forces [27]	Interactive, musical	Finger pressing	Sine wave	Sine wave	Audio-tactile summation offers best performance
Perception of musical scales [48]	Passive, musical	Hand-instrument contact	Musical scale	Masking noise	Differences between scales are discriminated
Perception of plucked strings [44]	Interactive, musical	String plucking	Guitar string	Guitar string	Haptic feedback does not increase sound quality
Piano playing [18]	Interactive musical	Free playing	Piano vibration on and off	Piano	Natural vibrations alter perceived piano quality
Digital piano playing [19]	Interactive, musical	Free playing	Piano vibration, filtered noise	Piano	Vibrations alter perceived keyboard quality
Playing experience on a haptic surface for musical expression [41]	Interactive, musical	Free playing	None, sine wave, filtered audio, band-passed noise	Expressive/dull	Vibrations coherent with audio feedback improve perceived interface quality/playing experience

12.1.2   Chapter Outline

In their respective interaction contexts and with different confidence levels, hence, the experiments chosen for this chapter share the general assumption that a sonic experience can be influenced by somatosensory cues. Some of them (e.g., [19, 48]) contributed to give form to the musical haptics research methodology and, hence, led to inevitably less robust conclusions. For this reason, they are certainly more suggestive than conclusive.

In an aim to orient the reader to the experiments which reflect his or her interests, Table 12.1 summarizes their key characteristics. Moreover, the table labels the experiments with gray tones classifying their dependence on specific elements. According to this classification, the first two experiments define an abstract context which is in principle applicable to multiple interaction contexts. The third and fourth experiments limit these contexts respectively to musical scales and plucked strings perception. The fifth and sixth ones further restrict the context respectively to acoustic and digital pianos. Finally, the seventh experiment specifically targets haptic versions of sound wave templates.

More in detail, the first experiment suggests a role of tactile frequency discrimination in enhancing the auditory perception of near-threshold frequency components; this role emerged during the audio-tactile identification of everyday materials from their response to a ball hitting them [13]. Next, we present an experiment conducted using an audio-tactile interface [40], showing that individuals performing a basic musical gesture such as finger pressing were able to reproduce previously learnt target forces more accurately if receiving contextual audio-tactile feedback instead of auditory or tactile feedback alone [27].

The third and fourth experiments link the aforementioned effects to musical experiences. As evidence of the power of the vibrotactile channel to deliver musical information, we first review a test in which Western and Indian musicians categorized and even identified music scales from both traditions by touching the surface of a harmonium [48]. Then, a robotic stringed instrument prototype called Keytar is described, in which the accurate haptic rendering of its virtual strings was significantly appreciated by users, however with no significant improvements for the perceived sound quality [44].

Conversely, a constructive effect was measured in pianists playing an acoustic piano whose natural vibrations could be switched on and off, thanks to peculiar engineering of the keyboard: in this case, the inclusion of vibrotactile feedback resulted in a measurable improvement of the instrument sound quality [18]. A similar effect was measured in musicians playing an actuated digital piano when this instrument reproduced vibrations recorded on a real piano [19].

Finally, using a force-sensitive haptic surface for musical expression which controlled a synthesizer, the effect of various vibration types on perceived quality attributes and the playing experience was assessed [41].

12.2   Ball Bouncing on Everyday Materials

Two experiments [13] studied the role of impact sounds and vibrations for the subjective classification of three flat objects, which were respectively made of wood, plastic, and metal—see Fig. 12.1.

The task consisted of feeling an actuated surface and listening through headphones to the recorded feedback of a ping-pong ball hitting such objects (Fig. 12.2, left), after they had been experienced during a training task (Fig. 12.2, right).

In Experiment 1, sounds and vibrations were recorded by keeping the objects in mechanical isolation. In Experiment 2, recordings were taken while the same objects stood on a table, causing their resonances to fade faster due to mechanical coupling with the support. Twenty-five subjects, aged between 23 and 61 years ($M = 32.1$, $SD = 10.1$), participated in Experiment 1, and twenty-seven (21–54 years old; $M = 29.0$, $SD = 6.8$) in Experiment 2. Eight subjects participated in both experiments. Roughly one-third of the participants were female. In terms of musical training, participants were not screened, and they reflect the general population average.
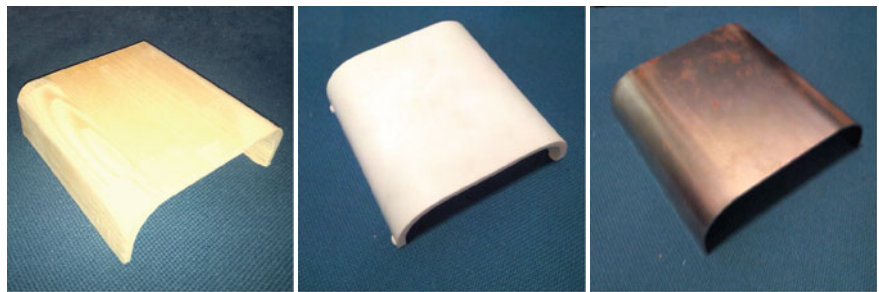
Fig. 12.1 Materials used in the experiment. Left: wood. Center: plastic. Right: metal

Fig. 12.2 Experimental tasks. Left: perceptual task. Right: training task
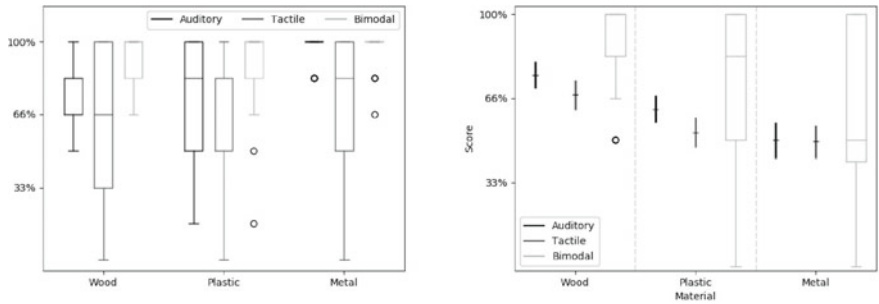

Fig. 12.3 Boxplot of and mean proportions correct with SE bars for all condition combinations. Left: Experiment 1. Right: Experiment 2

As a general result, in both experiments tactile identification was less accurate than auditory identification. In parallel, the *bimodal* (i.e., simultaneously auditory and tactile) identification ranked significantly better in both experiments, providing evidence of support from touch to auditory material identification (Fig. 12.3).

This conclusion was not contradicted by a control experiment, in which participants were asked to identify the materials from real bounces as during the training shown in Fig. 12.2, right.

Between Experiments 1 and 2, some interesting differences are observed between materials. In Experiment 1, metal was identified from auditory cues almost perfectly (difference between both plastic and wood was significant in multiple comparisons following a significant Friedman test: AuditoryWood-AuditoryMetal: $Z = 4.3$, Bonferroni-corrected $p < 0.01$; AuditoryPlastic-AuditoryMetal: $Z = 3.4$, $p < .01$). In contrast, in Experiment 2, the identification of metal was the poorest of the three materials. In a two-way repeated-measures ANOVA with Greenhouse-Geisser correction for insphericity, a significant main effect of Material was detected $(F(1.61,41.9) = 16.3$, $p \leq 0.001)$. The 95% confidence intervals of the three materials result in a partial overlap between Plastic (0.51–0.64) and Metal (0.42–0.57), whereas the 95% CI for Wood is entirely above their combined range (0.65–0.78). As the main difference between the stimuli in Experiments 1 and 2 was the length of the decay, it seems that the longer decay in Experiment 1 was an important identification cue, especially for metal.

Importantly for this chapter, the ability of our subjects to maximize their identification accuracy when using sounds and vibrations together suggests that audio-tactile summation may work in all individuals as soon as they have acquired a solid knowledge about a multisensory event belonging to the everyday experience, and not only if they have accumulated peculiar audio-tactile skills, e.g., by practicing for a long time with a musical instrument. This conclusion was reinforced by a further test, part of the same research, where *incongruent* bimodal stimuli were prepared by assembling sounds and vibrations reporting respectively on two different materials. This test in fact suggested that tactile feedback, in its limited possibility to convey timbre, became progressively more relevant as the auditory channel, in front of incongruent materials, left its leading role while remaining supportive of cross-modal perception.

12.3   Reproduction of Target Pressing Forces

An effect of haptic feedback on the control of finger-pressing force has been shown in the literature (e.g., [1, 28]). The present setup [27] approaches a musical task in that it measures memorized force targets in the presence of both auditory and vibrotactile feedback. The experiment was carried out by means of a tabletop device capable of measuring normal force while displaying vibrotactile feedback at its top panel (Fig. 12.4).

To simulate the haptic exchange taking place when playing acoustic or electroacoustic instruments—where musicians would learn the response of the instrument and would then perform by relying on kinesthetic memory [38]—participants first learned three target forces during a training phase, without additional feedback. Those targets were chosen empirically according to low, medium, and high pressing forces, within the data resolution of the interface (10-bit, corresponding to the 0–1023 range) and without anchoring them to corresponding values in Newton: the low target was set to 400, the medium one to 650, and the high target to 850. A double-sided window of 50 units was considered around each target as the acceptance range. The task was

Fig. 12.4 The interface used in the experiment for recording finger-pressing force and providing vibrotactile feedback
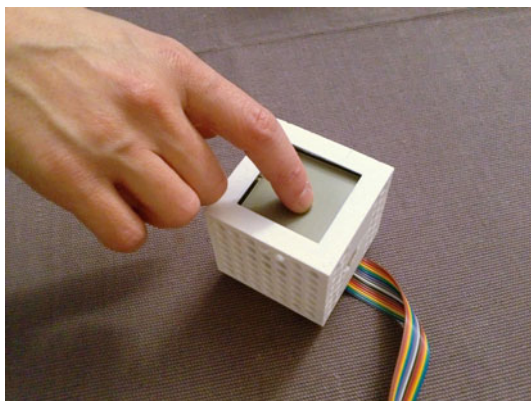
then to reproduce such forces "out of memory" under four feedback conditions: no feedback (N), auditory only (A), vibrotactile only (T), and auditory and vibrotactile together (AT). When participants believed they had reached the asked target they had to press an "OK" button with their free hand, while maintaining the pressing force on the touch panel.

For the sake of simplicity, a sinusoidal signal was chosen for rendering both auditory and tactile feedback, whose amplitude varied proportionally to the applied pressing force—thus implementing a gesture mapping commonly found in musical practice. The maximum intensity of vibrotactile stimuli was empirically set to the highest level that could be reproduced without perceivable distortion. The frequency of the sine wave was set to 200 Hz so as to maximize the produced vibrotactile sensation [55].

The test followed a 2-factor within-subjects design, where each participant was tested under each combination of conditions (12). All combinations were repeated 10 times, resulting in 120 trials that were presented in randomized order. Fourteen people (average age 33) participated in the experiment: five of them were pianists, five other musicians, and four non-musicians.[1]

Data analysis[2] showed a significant main effect of feedback factor ($F(3,143) = 16$, $p < 0.0001$). The effect of target force level was not significant ($F(2,143) = 0.7$, $p = 0.52$); however, the interaction "feedback × target level" was significant ($F(6,143) = 6.0$, $p < 0.0001$).

The interaction plots in Fig. 12.5 show that, for the low target force, mean errors are much smaller in the presence of auditory (A) or audio-tactile (AT) feedback, and somewhat smaller with tactile-only feedback (T) than with no-feedback (N). For the

---

[1] The relatively low participation number of musicians as well as non-musicians reflects the exploratory character of tactile experiments with pianists as far as one decade ago. Later, they have consolidated into more robust methodologies, including the participation of more musicians when necessary—see, e.g., Sect. 12.6.

[2] performed by *aligned rank transform*, the nonparametric equivalent to factorial within-subjects analysis of variance
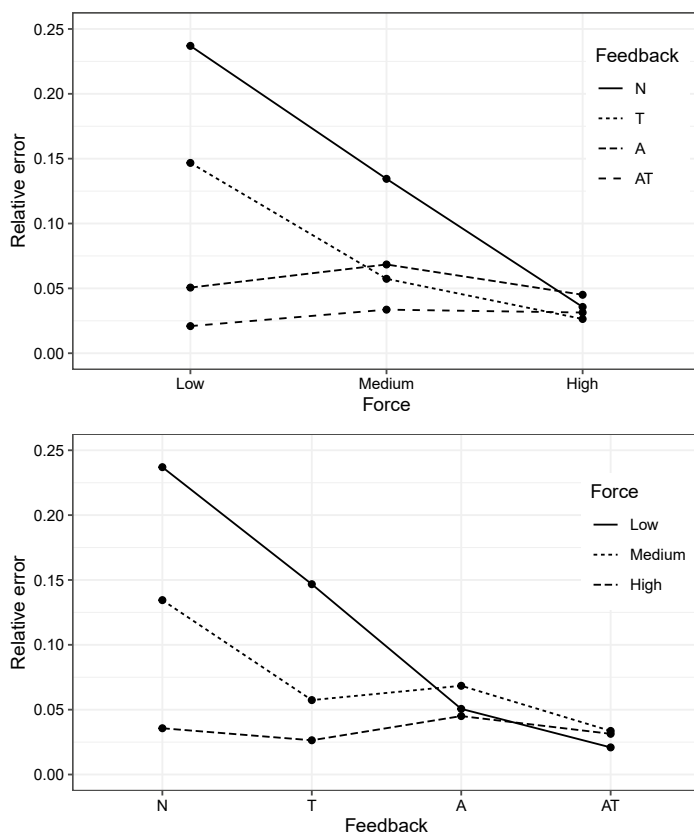
Fig. 12.5 Interaction plots. Top panel: mean relative errors at the three target forces, presented for each feedback condition. Bottom panel: mean relative errors at the four feedback conditions, presented for each target force level

medium target force, mean errors decrease in case of no-feedback (showing that the task becomes increasingly easier for higher forces) and with tactile-only feedback (T), whereas with auditory or audio-tactile feedback (A, AT) they did not change much from the low target force. For the high target force, however, the results are almost equivalent at all feedback conditions.

The results generally show that the addition of vibrations to auditory feedback may improve performance in musical finger-pressing tasks, enabling subjects to achieve memorized target forces with higher accuracy.

12.4 Vibrotactile Recognition of Traditional Musical Scales

The harmonium, visible in Fig. 12.6 (left), is played in both Western and Oriental music using scales that belong to the respective tradition. Musicians and also listeners with a normal understanding of music immediately recognize the ethnicity of a scale. In fact, the human ear is especially accurate in assessing the intervals existing between the fundamental frequencies of musical notes.

Does a haptic counterpart of scale recognition ability exist, result of a tactile frequency identification process musicians have internalized as part of their practice on an instrument? And, if recognition does not occur, would they be able to at least discriminate between different ethnicities? If either answer was positive, then musical vibrations would prove to be active carriers of spectral information capable of supporting, or even substituting, an especially important component of the musical message coming from an instrument.

Western and Indian notes have fundamental frequencies that in general do not match; furthermore, such intervals between notes differ depending on the scale. As a result, clearly audible discrepancies exist between Western and Indian musical scales, and then between different scales belonging to the same ethnicity.

The stimuli for the experiment [48] consisted of two Western (*C natural* and *A minor*) and two Indian (*Raag Bhairav* and *Raag Yaman-Kalyan*) scales played on the harmonium in the setup of Fig. 12.6 (right) by an Indian performer living in Europe. After listening to the four scales without touching the instrument during a training session, participants in a tactile recognition test were sitting on the left side of the same setup with their hands on the harmonium. At every trial, they were exposed to a train of vibrations corresponding to the sequence of notes belonging to a scale played by the performer. At the end of it, they had to decide whether the vibration was reporting about a Western or Indian scale, and to which one of the two.
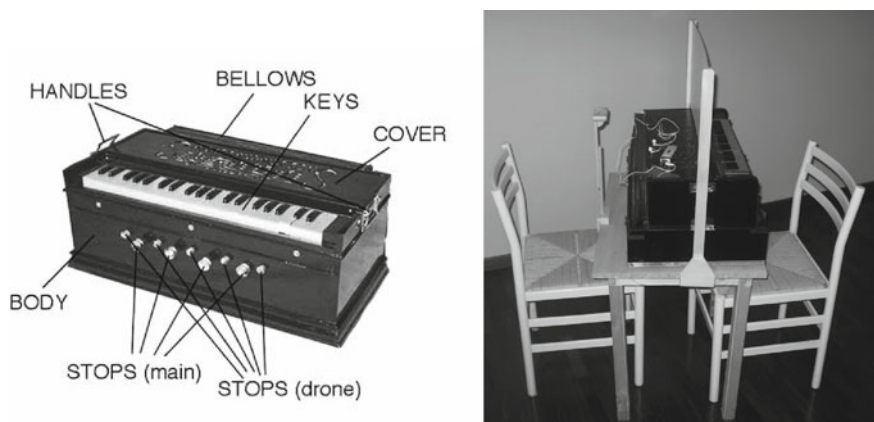

Fig. 12.6 Left: the harmonium. Right: experimental setup

Table 12.2 Individual subjective performance

Subject typology	Western participants		Indian participants	
	Recognition of tradition	Recognition of scale	Recognition of tradition	Recognition of scale
A—teacher of music	7/16	4/16	12/16	12/16
B—teacher of music	11/16	9/16	13/16	11/16
C—amateur musician	15/16	12/16	13/16	8/16
D—professional musician	16/16	12/16	10/16	7/16
E—professional musician	13/16	11/16	10/16	5/16
Overall recognition	62/80	48/80	58/80	43/80
Overall percentage	77.5%	60%	72.5%	53.75%
Chance percentage	50%	25%	50%	25%

During the test, they neither wore headphones emitting masking noise nor could they observe the playing action, thanks to a panel standing amid the harmonium body, avoiding the performer and participant from seeing each other.

The test was performed by a native group of Italians and then repeated in India. The two groups of participants, identical in number, were selected so as to have comparable levels of musical knowledge and performing skills. Results are listed in Table 12.2: They reveal the ability of both groups to recognize the ethnic origin with no significant differences between groups. Limited to specific subgroups, i.e., Western performers and Indian music teachers, the specific scale was recognized as well. The surprisingly high performance shown by our participants suggests the existence of a well-developed tactile memory for tones and/or note scales in musicians, a possible result of musical instrument training. However, the support during the task of nearly masked auditory cues of pitch bypassing the headphone insulators, or traveling from the hands to the cochlea through bone conduction, in principle could not be excluded. Similarly, scale-dependent temporal nuances biasing the recognition of the stimuli might have been unconsciously introduced by the performer during playing. In spite of its limited control, this experiment nevertheless represented an interesting starting point for the study of the role of touch in musical scale recognition.

12.5 Perception of Plucked Strings

Keytar is a plucked-string instrument interface [17]. Its software was developed within the Unity3D development engine. While running on a PC, Keytar provides real-time auditory, visual, and haptic feedback to the player who controls a virtual plectrum through a Phantom Omni robotic arm with one hand, while selecting notes and chords with the other hand (see Fig. 12.7, left). An accurate haptic rendering of the interaction point was made possible by modeling each string as a queue of
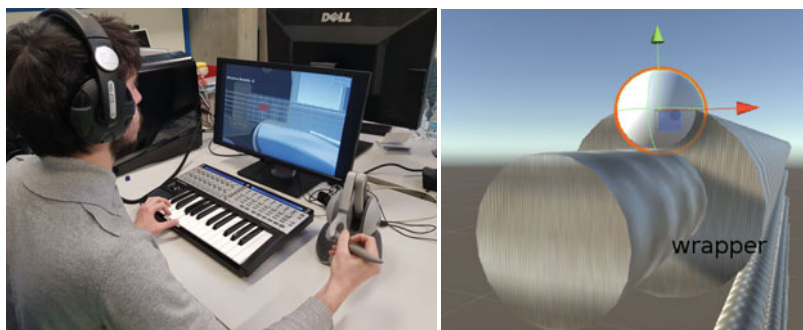
Fig. 12.7 Left: Keytar. Right: particular of the plectrum-string interaction point

short cylinders with alternating radius, and then by characterizing the contact of the plectrum using physical parameters which, due to the elastic behavior of the string, fall within the operating range of the Phantom Omni (see Fig. 12.7, right). This way, the robotic arm not only reproduces the elastic response of the plucked strings, but also some fine-grained dynamic textures arising between the colliding plectrum and the vibrating string. The sensation of rubbing the string during plucking is further enhanced by a realistic noise of frictional contacts coming from the servo-mechanisms of the robotic arm, while they are continuously switched on and off by the collision detection software module. The overall virtual environment defined an especially convincing reproduction of string plucking [45].

In a virtual reality experiment [44], twenty-nine participants on average having 8.2 years (SD = 8.3) of regular practice on a music instrument were asked to first pluck the strings of a real guitar, and then to wear an Oculus Rift CV1 helmet displaying an electric guitar and a plectrum in a nondescript virtual room. Twenty-one such participants in particular reported being able to play one or more stringed instruments. Interaction with the plectrum was made possible using the robotic arm controlled by Keytar, furthermore, the collision detection module controlled also a vibro-tactile actuator standing below Phantom Omni. This active stand was used to produce additional vibrations independently of the kinesthetic feedback. On such a setup, a within-subjects study compared four different haptic conditions during plucking: no feedback (N), force only (F), vibration only (V), and force and vibration together (FV).

Each participant was exposed to every condition, in randomized order, for approximately 20 minutes each. On every condition, first all six strings were plucked twice in a randomized order by the guidance of a visual marker emphasizing the string to pluck; then, participants were encouraged to freely interact by both plucking each string individually and strumming the entire string set. When one condition was completely tested, each participant evaluated four metrics on a Likert scale (see Fig. 12.8): overall perceptual similarity with the real instrument (from completely different to identical); stiffness similarity between virtual and real strings (from much lower to much higher); overall realism of the virtual instrument (from strong disagreement to
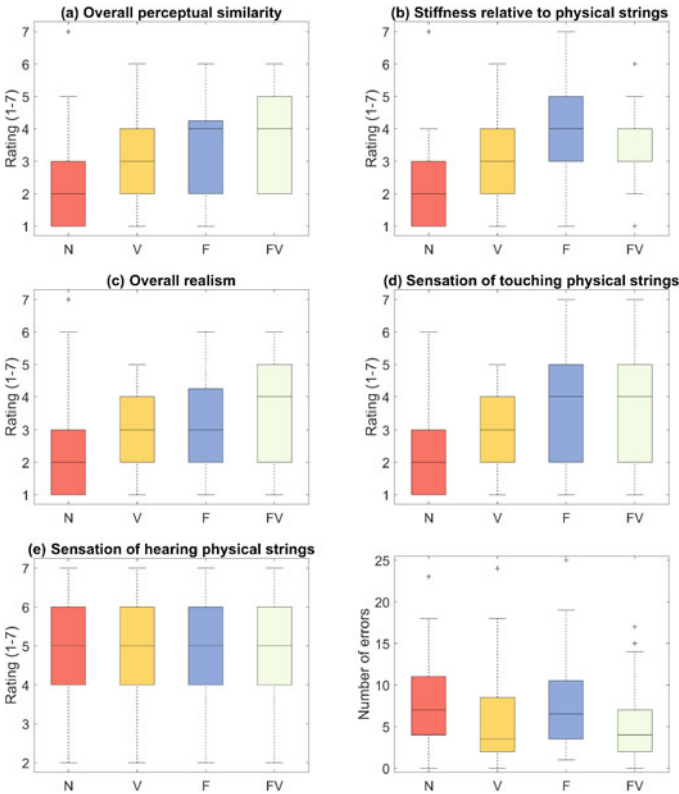
Fig. 12.8 Keytar: experimental results

strong agreement); touch realism of the virtual strings (from strong disagreement to strong agreement); effects of haptic cues on sound realism. At the end of the test each participant was additionally asked to choose his/her preferred condition. Finally, the errors made on plucking a wrong instead of a visually marked string during the part of the test involving individual strings were logged.

Results suggest the existence of significant effects of haptic feedback on the perceived realism of the strings. Further considerations can be drawn from the specific histograms [46]. By contrast, as can be seen from the left histogram below in Fig. 12.8, no effects on sound realism were measured. The lesson to take home from this experiment, hence, is that increasing the haptic realism of a virtual musical instrument in principle has no effects on its perceived auditory quality.

12.6 Piano Playing

A different lesson was instead learnt from an experiment in which the realism of the interaction with the musical instrument, in this case a piano, was pushed to its limit [18]. The piano keyboard in fact offers a controlled experimental setting, as the performer can only hit and then release one or more keys with one or more fingers while the rest of their body is disconnected from the instrument. This setting permitted to design a task in which auditory and haptic feedback could be delivered separately and independently. Furthermore, the intensity of both feedback channels is a reliable function of the key velocity which, in turn, is driven by the pianist's finger. Under these experimental premises, Yamaha's Disklavier pianos in particular offer two specific advantages: first, they can both record and mechanically reproduce the action of a pianist on all keys; secondly, they can be automatically switched between normal operation and a *silent* mode. When this mode is set, all strings are decoupled from the respective key hammers in ways that the instrument produces no sound, meanwhile conveying the same haptic feedback as to when the performer also hears the instrument.

The group of participants was split into two independent subgroups. Either subgroup performed on a grand Disklavier model DC3 M4 (in Padova, Italy) or on an upright model DU1A (in Zurich, Switzerland). During the tasks, the acoustic and silent modes were randomly switched across trials, letting the participants receive either natural or no steady vibrations from the keys after the initial percussive event. In both configurations participants via insulated headphones received the same auditory feedback, consisting of piano sounds synthesized by Modartt Pianoteq 4.5 digital piano software which was set to simulate a grand or an upright piano, and was driven in real time by the respective Disklavier's Musical Instrument Digital Interface (MIDI). The synthetic sounds were equalized so as to match those of the corresponding piano, by positioning a KEMAR mannequin visible in Fig. 12.9 (left), where the setup is shown during the calibration procedure. Figure 12.9 (right) shows a typical train of vibrations reaching the pianist's finger when the piano was operating in acoustic mode: the initial percussion event preceding the vibrations coming from the strings is evident in this figure.

Participants performed first a playing task and then a rating task. The former is relevant for this chapter. Three note ranges were considered separately across the keyboard, labeled low (keys below D3), mid (keys between D3 and A5), and high (keys above A5). Participants could play freely, within one range at a time, to compare the quality of the instrument in the presence and absence of string vibrations following the initial percussive events. Twenty-five professional pianists, mostly classical and a few jazz, took part in the tests: 15 on the upright and 10 on the grand piano (the slight imbalance in group sizes was due to varying easiness of recruitment in the two locations). Their average age was 27 years and their average piano experience was 15 years. Using a manual control, they could switch at their convenience between two setups, X and Y, associated with the silent and acoustic modes of the Disklavier. The difference between the two setups was not explained to them.
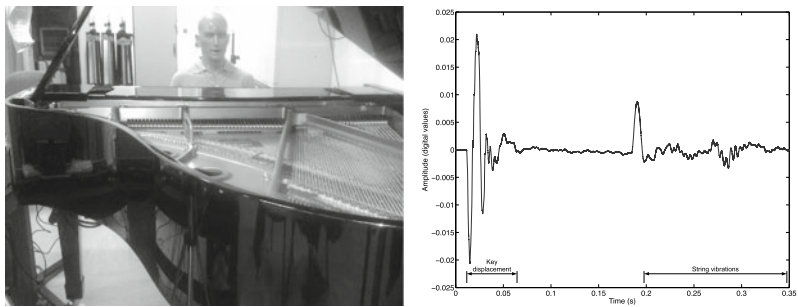
Fig. 12.9 Left: setup calibration. Right: acceleration signal measured on the key surface (note A2; MIDI velocity equal to 12; grand piano)


Fig. 12.10 Results with errorbars ±SE. Positive values signify preference for the vibrating mode. X-axis presents ratings for dynamic range, loudness, richness, and naturalness at low (A0-D3), mid (D3-A5), and high ranges (A5-C7) (l, m, and h, respectively). Preference was rated in full range only
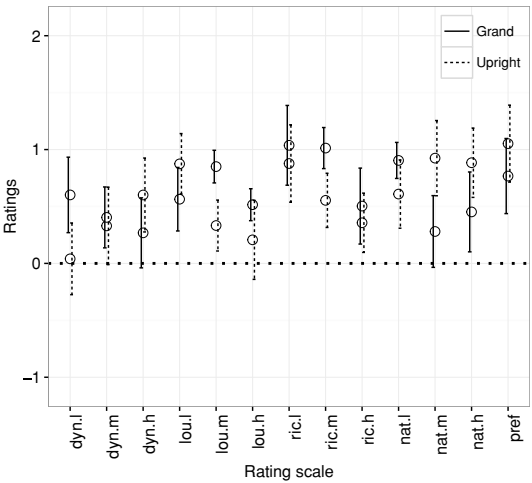
The task was to compare the setups on a Likert scale (from "X much better than Y" until "Y much better than X") with respect to the following attributes: dynamic range, loudness, richness, naturalness, and preference. The first four were rated separately in the low, mid, and high ranges, while the preference rating was given considering the entire keyboard. Participants were given definitions of the attributes and informed that dynamic range, loudness, and richness were mainly related to sound, whereas naturalness and preference could also be related to touch. A laptop finding place next to the piano displayed a set of sliders that were accessible at any moment to pianists for rating such attributes.

Results are shown in Figs. 12.10 and 12.11, suggesting a general preference for the vibrating mode. Since this preference was not explicitly linked to a specific attribute, two principal components, PC1 and PC2, were discovered to account for 80% of the variance. PC1 had the highest positive correlations with richness, naturalness, and preference; PC2, less powerful, was associated with dynamic range and loudness, which conversely decrease as naturalness and preference increase.

Fig. 12.11 Quality rating profiles projected onto the first two principal components. Subjects were segmented a posteriori according to positive/negative rating on preference. Ellipses enclose 68% of subjects in each group
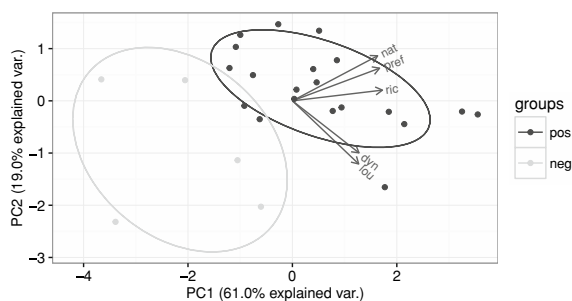
Analysis of Lin concordance correlation coefficients revealed a subgroup of seven subjects whose inter-individual consistency was negative. It was observed that most of them belonged to the group of five subjects who gave a negative preference rating. Therefore, participants were segmented a posteriori based on a positive versus negative preference rating. As seen in Fig. 12.11, the negative group differs from the majority of participants in that their ratings are negative on both principal components; in fact, while both groups gave rather similar ratings for dynamic range and loudness, their mean ratings for richness, naturalness, and preference are nearly opposite to each other. The conclusion was that approximately 80% of the participants preferred the vibrating setup and perceived higher naturalness and richness from it. Why the remaining 20% did not perceive any benefits from vibrations could not be thoroughly explained; however, in that group were two subjects who performed significantly under average in a vibration detection experiment related to this study. Notably, the negative group also included some jazz pianists. They reported performing frequently in small ensembles where digital stage pianos are used, which lack the natural vibrotactile feedback found on acoustic pianos.

At any rate, after completing the test in Zurich, the experimenter asked each participant what may have caused the difference between the setups: Interestingly, only 1 out of 15 participants could pinpoint vibrations. Thus, while the participants generally preferred the vibrating setup, they were not actively aware of vibrations. Their unawareness testifies to the especially high level of cross-modal integration that piano sounds and vibrations achieve in a real instrument.

12.7 Digital Piano Playing

An effect related to what was observed on acoustic pianos was discovered to play a role with digital pianos [19]. Since electronic instruments do not vibrate except for possible mechanical perturbations coming from the internal speakers, potential additional effects of artificial vibratory feedback to perceived instrument quality, precision in timing, and dynamic performance were investigated. The setup definition required to disassemble a digital piano keyboard, and then attach two vibrotactile
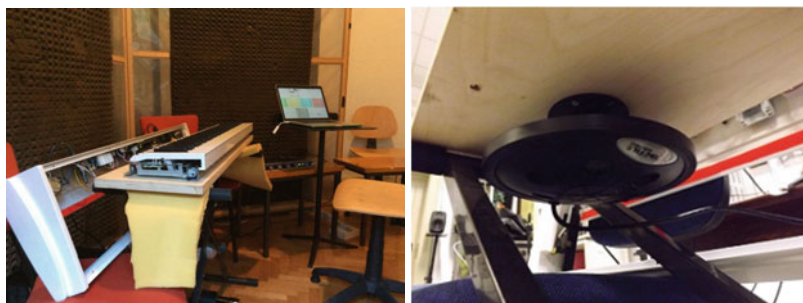
**Fig. 12.12** Left: experimental setup. Right: transducer conveying vibrations to the keyboard

actuators (Fig. 12.12, right) on a stiff wooden panel which was firmly screwed below its keybed (Fig. 12.12, left).

These actuators conveyed stimuli that had previously been acquired from an acoustic piano. In parallel, binaurally recorded tones were reproduced using headphones. Such tones and vibrations had previously been calibrated to have an intensity equal to that measured on the finger and ears of a pianist performing on a Disklavier grand piano, in the same fashion as the experiment in Sect. 12.6. In particular, calibration is required to equalize the vibration signals in order to avoid unrealistic resonance peaks on the digital keyboard for certain played notes.

Eleven pianists, five females and six males, participated in the experiment. Their average age was 26 years, and their average piano playing experience was 8 years after reaching the conservatory level. Two participants were jazz pianists. Audiotactile stimuli were produced at runtime: the digital keyboard in fact sent MIDI messages to a computer running Modartt Pianoteq 4.5 piano synthesizer and, in parallel, Native Instruments Kontakt 5 sampler in series with MeldaProduction MEqualizer parametric equalizer for playing back the corresponding vibration samples.

Perceived instrument quality was assessed by feeding the digital keyboard respectively with (A) no vibrations, (B) grand piano vibrations, (C) grand piano vibrations with 9 dB boost, and (D) synthetic vibrations. By contrast, the sound synthesis parameters were kept constant throughout the experiment. Pianists were asked to play freely while assessing the experience on five attribute rating scales: Dynamic control, Richness, Engagement, Naturalness, and General preference. During playing, at their convenience they could switch among two unknown setups, α and β: the former was always made to correspond to A, whereas the latter could randomly correspond to B, C, or D. The assessment was conducted by rating β relatively to α during 10 minutes of piano performance, for a session that hence lasted half an hour. During each assessment, participants at any time could rate every attribute by pointing to the respective virtual slider and setting a level by clicking with the mouse on a graphical user interface that was displayed by a laptop computer at hand reach. Each slider exposed a continuous Comparison Category Rating scale ranging from –3 ("β much better than α") to +3 ("β much worse than α"). Once the quality rating

Fig. 12.13 Results of the quality experiment. Boxplot presenting median and quartile for each attribute scale and vibration condition
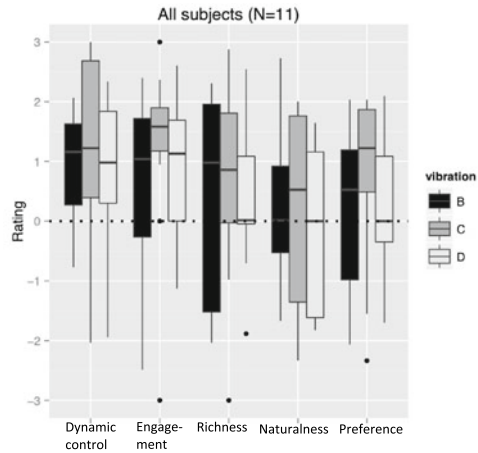
of the keyboard was over, another half an hour was spent by each participant to participate in the remaining two tests, assessing precision in timing as well as dynamic performance.

Results show that the augmented setups were generally preferred, with an emphasis on boosted vibrations (Fig. 12.13). Again, heterogeneity was observed in the data, as might be expected due to the high degree of variability in the inter-individual agreement scores. A k-means clustering algorithm was used to segment the subjects a posteriori into two classes, according to their opinion on General preference. Eight subjects were classified into a "positive" group and the remaining three into a "negative" group. The results of the respective groups are presented in Fig. 12.14. A difference of opinion is evident: The median ratings for the preferred setup C are nearly +2 in the positive group and –1.5 in the negative group for General preference. In the positive group, the median was positive in all cases except for Naturalness in D, whereas in the negative group, the median was positive only for Dynamic control in B.

Similar to what was observed in Sect. 12.6 while experimenting with the acoustic piano, low concordance between pianists exposed to vibration suggests that intra- and inter-individual consistency is an issue also while playing a digital piano. By contrast, no effect was observed on timing or dynamics accuracy in the performance tests. Taken together, these considerations point to conclude that vibrations do unconsciously influence the perceived keyboard instrument quality, however, along a direction which depends on the performer's previous multisensory experience of a specific instrument. Hence, augmenting a digital piano with the vibrations of an acoustic piano might not increase sense of quality if the performer played a digital (i.e., non vibrating) keyboard for most of the time. In parallel, haptic augmentation neither improves nor disrupts key aspects of piano performance such as timing and dynamic control.
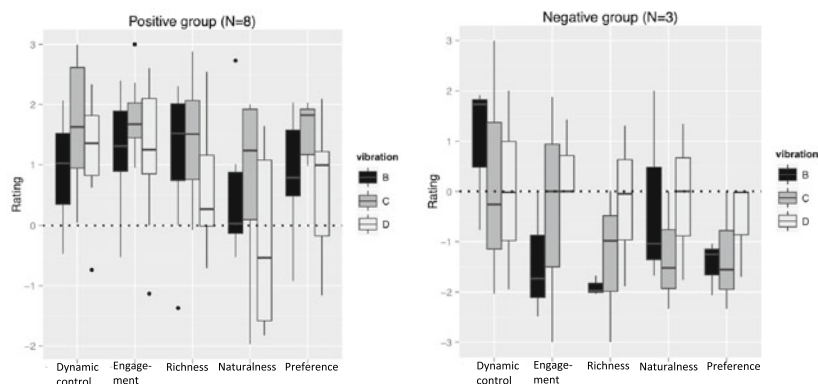
Fig. 12.14 Differences in quality ratings between the positive (left) and negative (right) groups formed by a posteriori segmentation. Boxplot presenting median and quartile for each attribute scale and vibration condition

12.8 Playing Experience on a Haptic Surface for Musical Expression

A multi-touch force-sensitive surface for musical expression was equipped with multi-point localized vibrotactile feedback, resulting in the HSoundplane haptic interface [43] shown in Fig. 12.15. A subjective assessment was conducted using the HSoundplane, which measured how the presence and type of vibration affect the perceived quality of the device, as well as various attributes related to the playing experience [41].

Fig. 12.15 The experimental setting for the HSoundplane experiment

12.8.1 Design

Two clearly distinct sound presets were tested, each with three vibrotactile feedback strategies.

The pitch of the audio feedback—ranging from A2 ($f_0 = 110\,\text{Hz}$) to D5 ($f_0 = 587.33\,\text{Hz}$)—was controlled along the x-axis. The two offered sound presets were

Sound 1—A sawtooth wave filtered by a resonant low-pass and modulated by a vibrato effect (i.e., amplitude and pitch modulation). A markedly expressive setting, responding to subtleties and nuances in the performer's gesture.

y-axis control: Vibrato intensity is controlled along the y-axis, from no-vibrato (bottom) to strong vibrato (top).

z-axis control: The filter cutoff frequency is controlled by the applied pressing force (i.e., higher force maps to brighter sound), and so is the sound level (i.e., higher force maps to louder sound).

Sound 2—A simple sine wave is added with noise depending on the location on the y-axis. A setting offering a rather limited sonic palette and no amplitude dynamics.

y-axis control: Moving upwards adds white noise of increasing amplitude, filtered by a resonant band-pass. The filter's center frequency follows the pitch of the respective tone.

z-axis control: Pressing force data are ignored, resulting in fixed intensity.

The different degrees of variability and expressive potential of the two sound settings allowed us to investigate whether the possible effect depends on audio feedback characteristics. All sounds were processed by a reverb effect so as to make the playing experience more acoustic-like. Sound was provided to the participants by means of closed-back headphones (Beyerdynamic DT 770 Pro). Audio examples of the two sound types are made available online,[3] demonstrating C3, C4, and C5 tones modulated along the y- and z-axes.

Before being routed to the actuators layer, vibration signals were filtered in the $10-500\,\text{Hz}$ range by a 10th-order band-pass, so as to optimize the actuators' efficiency and consequently the vibratory response of the device, as well as to minimize sound leakage. Any residual sound spillage produced by the actuators was taken care of by the closed-back headphones carrying auditory feedback. Three vibrotactile strategies were implemented:

Sine—Pure sinusoidal signals, whose pitch follows the fundamental of the played tones (f_0 within $110-587.33\,\text{Hz}$), and whose amplitude is controlled by the intensity of the pressing forces. By focusing vibratory energy at a single frequency component, this setting aimed at producing sharp vibrotactile feedback.

Audio—The same sounds generated by the HSoundplane used to render vibration: the audio signals are also routed to the actuators layer. Vibration signals thus share the same spectrum (within the $10-500\,\text{Hz}$ pass-band) and dynamics of the related sound. This approach ensured the highest coherence between musical output and

---

[3] https://tinyurl.com/HS-sounds.

tactile feedback, mimicking what occurs on acoustic musical instruments, where the source of vibration coincides with that of sound.

Noise—A white noise signal of fixed amplitude. This setting produced vibrotactile feedback generally uncorrelated with the auditory one, ignoring any spectral and amplitude cues possibly conveyed by it. The only exception is with Sound 2 and high y-axis values, which resulted in a similar noisy signal.

The designed vibration types offered different spectral and dynamics cues resulting in varying degrees of similarity with the audio feedback, thus enabling to determine the importance of the match between sound and vibration. The intensity of vibration feedback was set by the authors in a pilot phase, aiming at two main goals: (i) sound and vibration intensities had to feel reciprocally consistent; (ii) while levels had to be overall comfortable for prolonged use, vibration had to be clearly perceivable even at low force-pressing values [42].

At each trial, the task was to play freely while comparing two related setups: they were labeled A/B in a balanced way, and differed only in the presence/absence of vibration (i.e., they shared the same sound setting). Participants could switch at any time between A and B and had to provide ratings for four attributes: *Preference*, *Control and responsiveness* (referred to as *Control*), *Expressive potential* (referred to as *Expression*), and *Enjoyment*. Ratings were given by adjusting a respective slider on a continuous visual analog scale ranging from A (left) to B (right) to reflect the degree of preference in terms of the given attribute. In case of perceived equality between A and B, the slider would be set to the midpoint. All 4 (attributes) × 3 (vibration types) × 2 (sound types) factor combinations were evaluated twice.

All 29 participants—7 males and 22 females, aged 18–48 years (M = 25.4, SD = 7.1)—were professional musicians or music students. Their main instrument was either a keyboard or a string instrument, on which they had on average 17 years of experience. Roughly one-third of the participants had significant experience with electronic musical instruments, mostly synthesizers, or digital musical interfaces.

12.8.2 Results

The continuous slider scale ratings were mapped to the closed interval [0, 1], where 1 indicates a maximal preference for the vibrating setup and 0 maximal preference for the non-vibrating setup, and 0.5 is the point of perceived equality. Statistical analysis was carried out by fitting a zero-one-inflated beta (ZOIB) model, whose parameters were estimated with Bayesian methods [9, 32]. Four parameters describe the ZOIB distribution: the mean (μ) and precision (ϕ) of the beta distribution, the probability of a binary {0, 1} outcome (zoi), and the conditional probability of outcome {1} (coi). The mean of the beta distribution was modeled by sound, vibration type, their interaction, and attribute. The models for the precision (ϕ) and zero-one-inflation parameters (zoi, coi) were set to depend on vibration type, sound, and attribute without interactions.

Fig. 12.16 Marginal effects; estimated μ parameters with 95% Credible Intervals (N = 29). **a** Interaction between vibration and sound type; **b** Effect of vibration on the evaluated attributes. 0.50 = point of perceived equality; higher values indicate preference for vibrating over non-vibrating setup
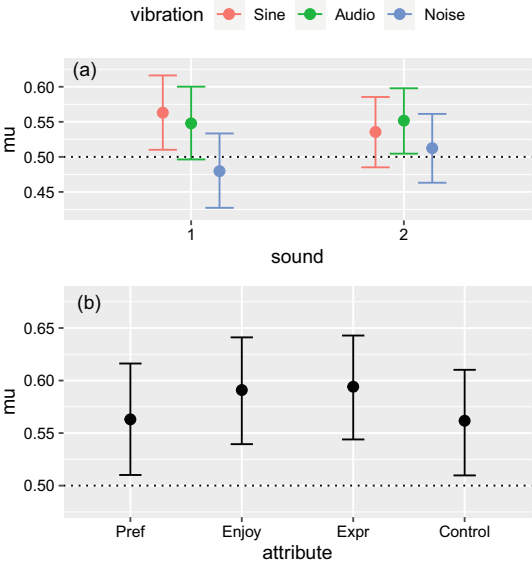
Estimates for the beta distribution means and their corresponding 95% Credible Intervals are presented in Fig. 12.16. On average, the vibrating setups were preferred to their non-vibrating versions: all mean estimates but one are above 0.50 (the point of perceived equality) as well as most of the respective credible intervals.

The model output showed the following effects.[4] The mean parameter for Audio vibration was not credibly different from Sine vibration, while Noise vibration was rated credibly lower. Sound type had a credible effect on the mean parameter (μ) only in combination with Noise vibration. *Expression* and *Enjoyment* both had a rather credible positive effect, although slightly short of 95%, on the mean parameter relative to *Preference* and *Control*. However, many of the manipulated factors had credible effects on the precision parameter (ϕ) and on the zero-inflation parameter (zoi), suggesting that even if the means are not credibly different, the shapes of the respective distributions may differ.

The main findings of this study may be summarized as follows: i) although not large, the measured effect of Sine or Audio vibration was appreciably positive. ii) Noise vibration did not credibly enhance the subjective quality of the interface as compared to the non-vibrating condition. iii) Vibrotactile feedback especially increased the perceived expressiveness of the interface and the enjoyment of playing. As appears from Fig. 12.16 (a), a more marked effect was found when vibration was more similar to the sonic feedback and consistent with the user's gesture: Indeed, Sine and Audio vibration follow the pitch of the produced sound and their intensity can be controlled by pressure. Conversely, Noise vibration—offering fixed ampli-

---

[4] Note that unlike the other studies reported in this chapter, these data were analyzed using Bayesian inference; therefore, we use the term "credible" instead of "significant" of effects.

tude, independent of the input gesture, and flat spectrum—was rated lowest among the vibrating setups. Noise vibration resulted in slightly better ratings when Sound 2 was used as compared to Sound 1: Again, that was likely because vibrotactile feedback is consistent, at least partially, with the noise-like sonic feedback produced for high y-axis values. Interestingly, no credible difference in the globally positive effect was found between Sine and Audio vibration. This may be at least partially explained by a masking effect taking place in the tactile domain toward higher frequencies, thus impairing waveform discrimination [5]. However, such phenomenon seems not to apply to markedly different signals [49]. In this regard, our informal testing revealed that Sine and Audio vibration were virtually indistinguishable, especially when Sound 1 (modulated sawtooth waveform) was selected.

Response consistency across repetitions was evaluated by modeling participants' first- and second-round responses by linear regression. Pooled over participants and factor combinations, the regression coefficient ($\beta = 0.32$, p < 0.001) indicated a general overall consistency (i.e., participants preferred the same vibrating or non-vibrating setup twice across repetitions). However, ten participants frequently preferred once the vibrating and once the non-vibrating setup in the same factor combination, resulting in regression coefficients ≤ 0 (mean coefficient over the N $= 10$ subjects was $\beta = -0.19$). The remaining subjects (N $= 19$) instead gave consistent ratings ($\beta = 0.53$). Interestingly, the inconsistent group (N $= 10$) spent noticeably less time with the tasks than the reliable group (N $= 19$): the median length of their gestural data logs was only 62% of that of the consistent group. In order to estimate the effect of the inconsistent participants, we re-run the ZOIB model including only the N $= 19$ consistent subjects and finding that the main result was similar to the full dataset: only vibration type had a clearly credible effect on the estimated mean parameter. However, this way the effect is somewhat larger, as the mean estimates for vibration types Sine and Audio (with Sound 1) slightly increase, while that for Noise decreases (see Table 12.3). Also in this case, *Expression* is the highest rated attribute; its marginal mean estimate increases from 0.59 to 0.64 (see Table 12.4).

As the participants were highly skilled musicians, we believe that the recorded inconsistent responses were not due to the task being too difficult. However, as

Table 12.3 Estimated μ parameters from the ZOIB fit (on original response scale) for the marginal effects of sound and vibration (attribute $=$ *Preference*). N $= 29$: all subjects; N $= 19$: consistent subjects

Sound	Vibration	Estimate (N $= 29$)	Estimate (N $= 19$)
1	Sine	0.563	0.604
1	Audio	0.548	0.576
1	Noise	0.480	0.466
2	Sine	0.536	0.558
2	Audio	0.552	0.550
2	Noise	0.512	0.493

Table 12.4 Estimated μ parameters from the ZOIB fit (on original response scale) for the marginal effects of Attribute (sound = Sound 1, vibration = Sine). N = 29: all subjects; N = 19: consistent subjects

Attribute	Estimate N = 29	Estimate N = 19
Preference	0.563	0.604
Control	0.562	0.594
Expression	0.594	0.645
Enjoy	0.591	0.628

they were not screened for individual vibrotactile sensitivity, it is possible that they did not feel vibrations equally. On top of that, we argue that rating inconsistency may be linked to the varying perceived vibration strength and audio-tactile congruence, depending on where and how the participants were playing over the interface's surface. Indeed, vibrotactile intensity perception is affected by vibration amplitude (obviously), spectral content (with a peak in the $200-300$ Hz range [55]), and the exerted pressing force [42]; also, varying degrees of spectral and temporal similarity between auditory and vibratory feedback may result either in cross-modal perceptual integration or interference [60]. However, we specifically chose a free playing task in order to measure the effect of vibrotactile feedback on various aspects of the playing experience.

With regard to the coherence of specific audio-tactile combinations, although Noise vibration resulted in very uniform ratings when associated with Sound 1 ($\beta = 0.56$, p < 0.001), it produced the lowest rating consistency with Sound 2 ($\beta = 0.16$, p < 0.05). While this was obviously affected by the general tendency of ten participants toward inconsistent ratings, one may also consider the varying degree of similarity between Sound 2 and Noise vibration: at the upper range of the y coordinate Sound 2 was noise-like, while for lower y values it was increasingly sinusoidal; inconsistency might follow from having played once mostly at high y and once mostly at low y. Conversely, Sound 1 retained the same degree of (dis)similarity with Noise vibration, independent of the playing position/style. Overall, the noticed inconsistency of responses sets a future challenge for screening the participants and controlling the playing task.

12.9   Conclusions

Based on the reported results, we suggest that the design of future multisensory interface technologies, especially if applicable to music performance, should take into consideration the addition of advanced vibrotactile feedback. This would enable the re-establishment of a consistent physical exchange between users and their digital devices—similar to the natural relationship that musicians establish with their instrument, where the source of sound and vibration coincides—with the demon-

strated potential to enhance the experience and the perceived quality of the interface. Indeed, several participants in the reported musical studies were impressed with the novelty and "aliveness" of haptic interfaces, as opposed to their experience with existing digital musical devices.

Ultimately, it is yet to be seen if and how such subjective enhancements may be reflected in the quality of playing, and musical performance altogether. Making objective measurements of these aesthetic aspects however poses a major research challenge, and the present work only scratched the surface in this direction. Instead, this will be the main object of a follow-up experiment currently in the works.

References

1. Ahmaniemi, T.: Effect of Dynamic Vibrotactile Feedback on the Control of Isometric Finger Force. IEEE Trans. on Haptics (2012).
2. Altinsoy, M. E. in Lect. Notes Comput. Sci. (eds Nordahl, R., Serafin, S., Fontana, F., Brewster, S.) 20–25 (Springer Berlin / Heidelberg, 2010).
3. Altinsoy, M. E.: Quality of auditory-tactile virtual environments. J. Audio Eng. Soc. **60**, 38–46 (2012).
4. Banu, A., Praliyev, N., Evagoras, X.: Effect of Frequency Level on Vibrotactile Sound Detection in Proc. of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications **2** (SciTePress, Prague, Czech Republic, July 2019), 97–102.
5. Bensmaïa, S. J., Hollins, M.: Complex tactile waveform discrimination. J. Acoust. Soc. of Am. **108**, 1236–1245 (2000).
6. Brungart, D. S.: Auditory localization of nearby sources. III. Stimulus effects. J. Acoust. Soc. Am. **106**, 3589–3602 (Dec. 1999).
7. Brungart, D. S., Durlach, N. I., Rabinowitz, W. M.: Auditory localization of nearby sources. II. Localization of a broadband source. J. Acoust. Soc. Am. **106**, 1956–1968 (Oct. 1999).
8. Brungart, D. S., Rabinowitz, W. M.: Auditory localization of nearby sources. Head-related transfer functions. J. Acoust. Soc. Am. **106**, 1465–1479 (Sept. 1999).
9. Bürkner, P.-C.: brms: An R Package for Bayesian Multilevel Models Using Stan. J. Stat. Softw. **80** (2017).
10. Caclin, A., Soto-Faraco, S., Kingstone, A., Spence, C.: Tactile "capture" of audition. Perception & Psychophysics **64**, 616–630 (2002).
11. Caetano, G., Jousmäki, V.: Evidence of vibrotactile input to human auditory cortex. Neuroimage (2006).
12. Dahl, S., Bresin, R.: Is the player more influenced by the auditory than the tactile feedback from the instrument? in Proc. Int. Conf. on Digital Audio Effects (DAFx) (Limerick, Ireland, Dec. 2001), 194–197.
13. De Pra, Y., Fontana, F., Järveläinen, H., Papetti, S., Simonato, M.: Does it ping or pong? Auditory and tactile classification of materials by bouncing events. ACM Trans. Applied Perception (TAP) **17**, 1–17 (2020).
14. Ernst, M. O., Bülthoff, H. H.: Merging the senses into a robust percept. Trends Cogn. Sci. **8**, 162-9 (2004).
15. Falchier, A., Cappe, C., Barone, P., Schroeder, C. E. in The New Handbook of Multisensory Processing (ed Stein, B. E.) chap. 4 (MIT Press, 2012).
16. Fiebelkorn, I. C., Foxe, J. F., Molholm, S. in The New Handbook of Multisensory Processing (ed Stein, B. E.) chap. 21 (MIT Press, 2012).
17. Fontana, F. et al.: Rendering and subjective evaluation of real vs. synthetic vibrotactile cues on a digital piano keyboard in Proc. Int. Conf. on Sound and Music Computing (SMC) (Maynooth, Ireland, July 2015), 161–167.

18. Fontana, F., Paisa, R., Ranon, R., Serafin, S.: Multisensory plucked instrument modeling in Unity3D: From Keytar to accurate string prototyping. App. Sci. **10**, 1452 (2020).
19. Fontana, F., Papetti, S., Järveläinen, H., Avanzini, F.: Detection of keyboard vibrations and effects on perceived piano quality. J. of the Acoust. Soc. Of Am. **142**, 2953–2967 (2017).
20. Fujisaki, W., Nishida, S.: Audio-tactile superiority over visuo-tactile and audio-visual combinations in the temporal resolution of synchrony perception. Exp. Brain Res. **198**, 245–259 (2009).
21. Galembo, A., Askenfelt, A.: Quality assessment of musical instruments - Effects of multimodality in Proc. of the 5th triennial conference of the European Society for the Cognitive Sciences of Music (ESCOM) (Hanover, Germany, Sept. 2003), 441–444.
22. Glennie, E.: Hearing Essay 1993.
23. Goble, A. K., Hollins, M.: Vibrotactile adaptation enhances frequency discrimination. J. of the Acoust. Soc. of Am. **96**, 771–780 (1994).
24. Hopkins, C., Maté-Cid, S., Seiffert, G., Fulford, R., Ginsborg, J.: Inherent and learnt abilities for relative pitch in the vibrotactile domain using the fingertip in 20th Int. Congr. Sound Vib. (ICSV 2013) (Bankok, 2013), 3207–3214.
25. Huang, J., Gamble, D., Sarnlertsophon, K., Wang, X., Hsiao, S.: Feeling music: integration of auditory and tactile inputs in musical meter perception. PLoS One **7** (Jan. 2012).
26. Jack, R. H., Mehrabi, A., Stockman, T., Mcpherson, A.: Action-sound latency and the perceived quality of digital musical instruments: Comparing professional percussionists and amateur musicians. Music Percept. An Interdisc. J. **36**, 109–128 (2018).
27. Järveläinen, H., Papetti, S., Schiesser, S., Grosshauser, T.: Audio-Tactile Feedback in Musical Gesture Primitives: Finger Pressing in Int. Conf. on Sound and Music Computing (SMC) (2013), 109–114.
28. Jiang, L., Cutkosky, M. R., Ruutiainen, J., Raisamo, R.: Improving finger force control with vibrational haptic feedback for Multiple Sclerosis in Proc. of the IASTED Int. Conf. on Telehealth/Assistive Technologies (ACTA Press, Baltimore, Maryland, 2008), 110–115.
29. Kaaresoja, T., Brewster, S., Lantz, V.: Towards the temporally perfect virtual button: touch-feedback simultaneity and perceived quality in mobile touchscreen press interactions. ACM Trans. on Applied Perception **11**, 1–25 (2014).
30. Kayser, C., Petkov, C. I., Augath, M., Logothetis, N. K.: Integration of Touch and Sound in Auditory Cortex. Neuron **48**, 373–384 (2005).
31. Kitagawa, N., Spence, C.: Audiotactile multisensory interactions in human information processing. Jpn. Psychol. Res. **48**, 158–173 (2006).
32. Kruschke, J. K.: Doing Bayesian data analysis - A tutorial with R, JAGS, and Stan 2nd (Academic Press, 2014).
33. Kuchenbuch, A., Paraskevopoulos, E., Herholz, S. C., Pantev, C.: Audiotactile integration and the influence of musical training. PLoS One **9**, e85743 (2014).
34. Landry, S. P., Champoux, F.: Musicians react faster and are better multisensory integrators. Brain Cogn. **111**, 156–162 (2017).
35. Levänen, S., Hamdorf, D.: Feeling vibrations: enhanced tactile sensitivity in congenitally deaf humans. Neuroscience letters **301**, 75–77 (2001).
36. Merchel, S., Schwendicke, A., Altinsoy, M. E.: Feeling the sound: audiotactile intensity perception in Proc. 2nd Polish-German Struct. Conf. Acoust. 58th Open Semin. Acoust. (2011).
37. Merchel, S., Leppin, A., Altinsoy, M. E.: Hearing with your Body: The Influence of Whole-Body Vibrations on Loudness Perception in Proc. 16th Int. Congr. on Sound and Vibration (ICSV16) (Kraków, Poland, July 2009), 5–9.
38. O'Modhrain, S., Gillespie, R. B. in Musical Haptics (eds Papetti, S., Saitis, C.) 11–27 (Springer International Publishing, Cham, 2018).
39. Okazaki, R., Kajimoto, H., Hayward, V.: Vibrotactile Stimulation Can Affect Auditory Loudness: A Pilot Study in Haptics: Perception, Devices, Mobility, and Communication (eds Isokoski, P., Springare, J.) (Springer Berlin Heidelberg, Berlin, Heidelberg, 2012), 103–108.
40. Papetti, S., Fröhlich, M., Schiesser, S.: The TouchBox: an open-source audiohaptic device for finger-based interaction in IEEE World Haptics Conf. (WHC) (Tokyo, Japan, July 2019), 491–496.

41. Papetti, S., Järveläinen, H., Schiesser, S.: Interactive vibrotactile feedback enhances the perceived quality of a surface for musical expression and the playing experience. IEEE Trans. on Haptics, 1–1 (2021).
42. Papetti, S., Jarvelainen, H., Giordano, B. L., Schiesser, S., Frohlich, M.: Vibrotactile Sensitivity in Active Touch: Effect of Pressing Force. IEEE Trans. on Haptics **10**, 113–122 (2017).
43. Papetti, S., Schiesser, S., Fröhlich, M.: Multi-point vibrotactile feedback for an expressive musical interface in Proc. New Interfaces for Musical Expression (NIME) (2015).
44. Passalenti, A. et al.: No Strings Attached: Force and Vibrotactile Feedback in a Virtual Guitar Simulation in Proc. 2019 IEEE Conf. on Virtual Reality and 3D User Interfaces (VR) (2019), 1116–1117.
45. Passalenti, A., Fontana, F.: Haptic interaction with guitar and bass virtual strings in Proc. Int. Conf. on Sound and Music Computing (SMC) (Limassol, Cyprus, July 2018), 427–432.
46. Passalenti, A. et al.: No Strings Attached: Force and Vibrotactile Feedback in a Guitar Simulation in Proc. Int. Conf. Sound and Music Computing (SMC 2019) (Málaga, Spain, 2019), 28–31.
47. Perrault Jr, T. J., Rowland, B. A. in The New Handbook of Multisensory Processing (ed Stein, B. E.) chap. 6 (MIT Press, 2012).
48. Romagnoli, M., Fontana, F., Sarkar, R.: Vibrotactile Recognition byWestern and Indian Population Groups of Traditional Musical Scales Played with the Harmonium in Haptic and Audio Interaction Design (eds Cooper, E. W., Kryssanov, V. V., Ogawa, H., Brewster, S.) (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011), 91–100.
49. Russo, F. A., Ammirante, P., Fels, D. I.: Vibrotactile discrimination of musical timbre. J. Exp. Psychol. Hum. Percept. Perform. **38** (Aug. 2012).
50. Sanabria, D., Soto-Faraco, S., Spence, C.: Spatiotemporal interactions between audition and touch depend on hand posture. Exp. Brain Res. **165**, 505–14 (2005).
51. Schürmann, M., Caetano, G., Hlushchuk, Y., Jousmäki, V., Hari, R.: Touch activates human auditory cortex. Neuroimage **30**, 1325–31 (2006).
52. Schürmann, M., Caetano, G., Jousmäki, V., Hari, R.: Hands help hearing: facilitatory audiotactile interaction at low sound-intensity levels. J. of the Acoust. Soc. of Am. **115**, 830–832 (2004).
53. Shore, S. E., Dehmel, S. in The New Handbook of Multisensory Processing (ed Stein, B. E.) chap. 1 (MIT Press, 2012).
54. Stanford, T. A. in The New Handbook of Multisensory Processing (ed Stein, B. E.) (MIT Press, 2012).
55. Verrillo, R. T.: Vibration sensation in humans. Music Perception **9**, 281–302 (1992).
56. Wilson, E. C., Reed, C. M., Braida, L. D.: Integration of auditory and vibrotactile stimuli: Effects of phase and stimulus-onset asynchrony. J. Acoust. Soc. Am. **126**, 1960–74 (2009).
57. Wilson, E. C., Reed, C. M., Braida, L. D.: Integration of auditory and vibrotactile stimuli: Effects of frequency. J. Acoust. Soc. Am. **127**, 3044–59 (2010).
58. Wollman, I., Fritz, C., Poitevineau, J., McAdams, S.: Investigating the role of auditory and tactile modalities in violin quality evaluation. PLoS One **9**, e112552 (2014).
59. Yau, J. M., Olenczak, J. B., Dammann, J. F., Bensmaïa, S. J.: Temporal frequency channels are linked across audition and touch. Curr. Biol. **19**, 561–6 (2009).
60. Yau, J. M., Weber, A. I., Bensmaïa, S. J.: Separate mechanisms for audiotactile pitch and loudness interactions. Frontiers in psychology **1**, 160 (2010).
61. Young, G. W., Murphy, D., Weeter, J.: Auditory Discrimination of Pure and Complex Waveforms Combined With Vibrotactile Feedback in Proc. New Interfaces for Musical Expression (NIME) (Baton Rouge, LA, 2015).
62. Young, G. W., Murphy, D., Weeter, J.: Haptics in music: The effects of vibrotactile stimulus in low frequency auditory difference detection tasks. IEEE Trans. on Haptics **1412**, 135–139 (2017).