Article

Assessing the Influence of Multimodal Feedback in Mobile-Based Musical Task Performance

Alexandre Clément [1,*] and Gilberto Bernardes [1,2]

1   Faculdade de Engenharia (FEUP), Universidade do Porto, 4200-465 Porto, Portugal
2   Instituto de Engenharia de Sistemas e Computadores, Tecnologia e Ciência (Inesc-Tec),
    4200-465 Porto, Portugal
*   Correspondence: a.clement@fe.up.pt

Abstract: Digital musical instruments have become increasingly prevalent in musical creation and production. Optimizing their usability and, particularly, their expressiveness, has become essential to their study and practice. The absence of multimodal feedback, present in traditional acoustic instruments, has been identified as an obstacle to complete performer–instrument interaction in particular due to the lack of embodied control. Mobile-based digital musical instruments present a particular case by natively providing the possibility of enriching basic auditory feedback with additional multimodal feedback. In the experiment presented in this article, we focused on using visual and haptic feedback to support and enrich auditory content to evaluate the impact on basic musical tasks (i.e., note pitch tuning accuracy and time). The experiment implemented a protocol based on presenting several musical note examples to participants and asking them to reproduce them, with their performance being compared between different multimodal feedback combinations. Collected results show that additional visual feedback was found to reduce user hesitation in pitch tuning, allowing users to reach the proximity of desired notes in less time. Nonetheless, neither visual nor haptic feedback was found to significantly impact pitch tuning time and accuracy compared to auditory-only feedback.

Keywords: digital musical instruments; interaction; instrument control; musical gesture; multimodal feedback; haptics; smartphone; human computer interaction

1. Introduction

Multimodal feedback is fundamental to tasks requiring physical control. From an early age, humans use touch as one of the main learning processes [1,2] and explore how the physical world reacts to our actions. Sight and hearing are senses we learn, which do not require physical contact to operate. What we perceive through them often does not directly correlate to our interaction with the world. However, auditory and visual stimuli are frequently triggered due to tactile interaction. We expect a response from physical objects we manipulate and learn how to interpret feedback to master physical tasks. Whenever a system is manipulated, we need a feedback response to guide our actions—we need to gauge how our control influences or triggers a response (most commonly via visual feedback). However, when physically manipulating an object to exert control over any system, we expect a response we can physically feel.

By design, traditional musical instruments are physical systems governed by physical laws. While most people associate musical instruments with their auditory feedback only, musical instruments are multimodal, i.e., they operate in different representational modes. Sound production is probably the most important, but other modalities are intrinsic to their operation. A guitar produces sound by vibrating its strings, and this vibration can be heard, observed, and felt. Similarly, most traditional acoustic instruments encompass auditory, visual, and haptic feedback while being played. Their physical nature, moving parts, and imprinting of movement on the encompassing space provoke multimodal responses.

On the other hand, digital musical instruments (DMIs) do not rely on acoustical processes for sound production, which would result in intrinsic physical response. Therefore, they do not embed multimodal representations by design, possibly fostering a disconnect between the performer and the DMI. Beyond the fundamental auditory response, which needs to be translated from a discrete digital representation to continuous acoustic energy, additional multimodal feedback potentially increases the *playability* (playability of a piece of music or an instrument: the quality of not being too difficult for someone to play [3]). This is reflected not only in the ease of learning said instrument but also in the increased control over it.

Today, the multitude of devices that can be used to create and run musical and auditory processes, together with the complexity of finding proper approaches beyond traditional instruments, open several avenues for research and musical creativity. A subset of DMIs that run on mobile handheld devices (MHDs) such as smartphones, tablets, and similar devices is of interest here. These devices are somewhat distinct from most DMIs, as they have become pervasive within our society. Traditional DMIs are typically designed from the ground up, with all elements of their system (i.e., control mechanisms, processing, and feedback methods) being created with a specific goal in mind. Mobile phones and smartphones are ubiquitous, and their use is becoming second nature to us [4]. Their increasingly powerful capabilities, portability, and their wide array of hardware components suited for feedback generation make them perfect for developing musical tools—allowing the creation of universally available DMIs as self-contained instruments running on readily available devices. Multimodal feedback in DMIs has an increasing body of literature, but much remains to be explored and studied in the field of MHD-based DMIs. Our work addresses this limitation by studying how the general population appropriates an MHD in a musical context, i.e., how one would use an MHD and operate it as a musical instrument, and how to best design said instrument in terms of control and feedback methods. The study presented in this article follows a previous experiment [5], where we explored how users map musical actions (i.e., note onset, note pitch variation, note duration variation, note amplitude variation) as MHD operative gestures. The contribution resulted in guidelines, or best practices, for mapping operative gestures to musical actions: results have shown that all these actions were overwhelmingly mapped as touch gestures operated using these devices' touchscreen.

This first step towards defining guidelines for creating MHD-based DMIs allowed us to look into the next logical step in this process: by knowing the most representative modes of operations, how can we best inform them about how the system is behaving? Going back to the first paragraphs of this section, we discussed how traditional musical instruments are multimodal, so, how much of that behavior should be translated into these new instruments? MHDs have specific hardware components that easily recreate these additional representations or feedback modes. These components, i.e., the device actuators, allow us to translate digital information into physical responses: auditory (the device's speaker), visual (the device's screen), and physical (the device's vibration motor). We designed an experiment to ascertain the influence of these additional feedback modes in reinforcing traditional auditory feedback in a given musical context. Specifically, we wanted to determine if having additional layers of visual and physical feedback influenced the ability of users to execute note pitch tuning.

This article is structured as follows. Section 2 contextualizes the field of work by going over important points in the body of research and important concepts and premises behind the present experiment. Section 3 details the design of an experimental procedure followed and the data collection and analyses. Section 4 presents and discusses the results stemming from the experiment. Finally, Section 5, summarizes our study's main conclusions and contributions, as well as future work and implications drawn from these results.

2. Related Work and Contextualization

> "Playing music requires the control of a multimodal interface, namely, the music instrument, that mediates the transformation of bio-mechanical energy to sound energy, using feedback loops based on different sensing channels, such as auditory, visual, haptic, and tactile channels." [6]

This definition provides a good starting point for defining how a musical instrument (digital or otherwise) could be viewed: a system controlled by a multimodal interface, converting input information (energy in the case of acoustical instruments) into a diverse array of auditory, visual, haptic, and tactile feedback signals. However, how do we define and determine what the input control and output feedback are? Moreover, what exactly is feedback? Let us take a side-step out of our field of focus and into the field of education. We can find a good starting point to try and define this:

> "Feedback is conceptualised as information provided by an agent (e.g., teacher, peer, book, parent, self, experience) regarding aspects of one's performance or understanding." [7]

Education is a field where feedback is paramount in the process, considering it is the main way by which students assess their errors and correct their results—"Feedback thus is a "consequence" of performance." [7]. This feedback can be of various different natures/types—auditory (e.g., speech, warning sounds), visual (e.g., color codes, symbols), tactile (e.g., movement or resistance), for example—and can be multimodal when the information is conveyed via a combination of different feedback types. Multimodality has consistently been found to provide a richer and more accurate way for recipients to assimilate and react to their performance. This has become especially important for digital tools and multimedia, where operation is usually not reliant on physicality, and the tools' behavior and user feedback must be defined.

Multimodal feedback has a solid base of research and evidence supporting its use and application as a positive influence on task performance in the context of mobile experiences. It is important to note, nonetheless, that the overwhelming majority of studies focus on multimodality anchored around visual feedback—all other feedback types serving as additional layers of information supporting that specific mode. In particular, feedback multimodality in mobile device operation improves usability and expressiveness [8]. It also provides a more flexible way for users to better perceive and avoid errors [9]. Research has shown that multimodal feedback positively impacts performance, perceived effort, error correction/avoidance, and error reaction time [10–13]. For example, Ref. [14] studied the impact of multimodal feedback on mobile gaming as a way of providing additional layers of information to the game experience, namely by testing and comparing using tilt feedback (i.e., physically moving the device) and using a combination of auditory, haptic, and visual feedback. Their research found that the second approach had a noticeable positive impact whenever at least one of the types of feedback was present—especially with the inclusion of auditory feedback. They also confirmed that users preferred the presence of additional feedback due to the added layer of interaction introduced to the experience. Similarly, Ref. [15] found that multimodal feedback (i.e., audio-visual, haptic-visual and audio-haptic-visual) led to better performance when compared to strictly visual feedback, with a particular favoring towards haptic-visual feedback. Haptics and sound are already frequently used in conjunction and can be found in interactions that are becoming a staple in mobile device use (e.g., the "Home" button click sound/vibration, simulating a physical button press). The combination of both auditory and haptic feedback was found to elicit a considerably different performance impact when compared to either type added by itself to the original visual feedback. Ref. [16] found that haptic-visual feedback was favored by users when compared to multimodal audio-visual feedback, noting the importance of these additional feedback types when considering (for example) usage by visually-impaired users. A comprehensive meta-analysis by Ref. [17] goes over 45 studies and highlights the potential of haptic (vibrotactile, as referred to by the authors) feedback as supporting

visual information. This feedback multimodality has also been researched and applied as a way of enriching the musical experience, and its perception by sensory-impaired persons [18,19]—suggesting that the additional layers of information are important in reinforcing the perception of the primary feedback mode.

Research surrounding multimodal feedback for digital musical interfaces and systems also has a growing body of research, which consistently finds support in validating the positive impact of introducing additional feedback types as support to auditory feedback [20–22]. Our research, however, focuses on a specific control interface and musical system: mobile-based musical instruments. This is an approach with very little research. However, some studies have addressed both, enabling easy implementation and assessing the impact of feedback multimodality when supporting audio-based tasks [23,24], illustrating the interest in this subject.

This study focused on three distinct feedback types: auditory, visual, and haptic. These are the most common feedback methods in related literature, with visual feedback being the most predominant. While visual and auditory feedback are prevalent throughout the digital field, commonly used on both physical and digital tools, haptic is somewhat less common and thus less familiar to users. From the Greek haptikos—"to grasp or perceive"—haptics has come to refer to what one might call the science of touch. In a context-specific definition, haptics refers to physical feedback, able to be perceived by tactile means, most commonly employing vibrotactile stimulus (i.e., vibration force). Haptics has become an important aspect in the design of digital musical instruments and has been the focus of research, and dedicated publication [25]. This feedback type is important in fostering the connection between performers and Digital Musical Instruments (DMIs) and Digital Music Control Interfaces (DMCIs) by providing further embodied interaction via a physically perceivable layer of information concerning the auditory content. This application of haptics, however, mostly relies on using external devices and wearable contraptions in order to generate the vibrotactile feedback [26–30]. If, in the case of DMCIs, these can be integrated with the physical controller without altering its footprint or increasing its unwieldiness, in the case of software-based DMIs, they represent an additional external element, which might be disruptive to the interaction process. Mobile devices present a unique case wherein the device running the DMI is also equipped with the necessary tools for both control input and multimodal feedback generation.

2.1. Specificities of Mobile Devices

DMIs rely on physical actuators to produce feedback for the player, either acoustical, mechanical, or optical. In the context of mobile phones, or mobile handheld devices (MHDs), there are a plethora of actuators bundled with the device. Some projects and prototypes aim to further expand their possibilities with external add-ons. However, to design a self-contained digital instrument to be used in a commonly available MHD, one must conform to the reality of the devices and their available sensors and actuators. This can be seen at the same time as a limitation and possibility. On the one hand, these actuators are available for use in most devices as-is, but on the other, one is limited to the ones most manufacturers adopt. As it stands, most MHDs are equipped with (at least) an acoustical actuator (speaker), a mechanical actuator (vibration motor), as well as several optical actuators (e.g., status LEDs, edge/rim LEDs, flashlight), in addition to the ubiquitous touchscreen.

2.1.1. Input (Sensors)

Haptics are something to consider both in terms of output feedback and input control, with pressure being the most immediate choice. The usage of this as data input for MHD DMIs' control is tightly linked to physical actuators available, which are (to this day) still widely unavailable in most devices—although some simulations making use of other more common actuators have been proposed. For example, Ref. [31] proposed an emulation of pressure-sensing technology, making use of the MHD's speaker and microphone by using the microphone to record a continuous sound frequency generated by the speaker.

They extrapolated a degree of pressure applied to the device's body by comparing both signals and calculating physical signal damping. Ref. [32] proposed a pressure-sensing technology using the same sensor and actuator by generating a predefined sound, recording it with the device's microphone, and comparing both to measure the physical bend in the MHD's body and thus extrapolate touch pressure. One can easily see how both approaches will greatly depend on both devices (microphone and speaker) capabilities: the speaker needs to generate an inaudible sound at constant frequency and amplitude, while the microphone has to be able to record said frequency properly. This brings into play variables as mentioned above like distortion sensibility, frequency range, and dynamic range, which greatly vary with the quality and build of the devices. Speakers have a wide array of frequency and dynamic range, with massive changes in capabilities from the lower-end to the higher-end devices, which result in a limitation in developing a solution that will fit most devices. In addition to these hardware constraints, this algorithm would be nigh impossible to extrapolate to a wide range of devices, considering their physical specificities in terms of materials and format.

2.1.2. Output (Actuators)

Vibration motors vary greatly in terms of capabilities and, consequently, feedback possibilities. While some, for example, allow for vibration intensity variations, others only allow system-wide fixed intensity. The vibration intensity range also greatly varies from device to device. Considering the hardware diversity of available actuators, it is difficult to design an exact and dynamic user-feedback haptic model making use of them.

The most universally available and consistent haptic actuator can be considered the touchscreen—by definition, the devices we focus on depend on having a touchscreen to be considered as such. These devices' quality (e.g., pixel density, contrast, color quality, refresh rate, touch sensibility, points of contact) greatly varies from the low to high-end range of MHDs. Nonetheless, their capabilities in visual cue production can be seen as universal, at least on the level one can expect in the context of this work.

3. Materials and Methods

We conducted an experiment targeting the general population, regardless of musical experience, to evaluate the impact of visual and haptic feedback on the effectiveness of musical note reproduction on mobile-based DMIs. More specifically, the experiment aimed to assess whether visual and haptic feedback methods impacted note pitch tuning time and/or note pitch tuning accuracy when used as a complement to auditory feedback. Furthermore, we designed the experiment to ascertain whether familiarity was exhibited or noticeable with practice. In greater detail, the experiment was designed to answer the following questions in the context of mobile-based DMIs:

1. Does visual feedback impact pitch tuning time and accuracy?
    We wanted to assess if visual feedback of note pitch height, in addition to an auditory response, impacts the time and pitch accuracy participants took to reach a target pitch;
2. Does haptic feedback impact pitch tuning time and accuracy?
    We wanted to assess if haptic feedback of note pitch, in addition to the auditory response, impacts the time and pitch accuracy participants took to reach a target pitch;
3. Do these additional visual and haptic feedback methods impact macro and micro note tuning?
    We wanted to assess whether visual and haptic feedback impacts the time to reach a target note in pitch tuning. In the scope of this experiment, we considered micro tuning the pitch tuning adjustments within a semitone distance to the target and macro tuning as tuning up until that point (please refer to Section 3.3 for a detailed description of macro and micro tuning);
4. Is there a perceivable degree of familiarity for visual and haptic feedback methods in note reproduction on mobile-based DMIs?
    We wanted to study the degree of improvement of note tuning using visual and

haptic feedback in addition to an auditory response, which can indicate the impact of familiarity in the successful completion of the note tuning action.

In order to provide evidence that can shed light on the above questions, the experiment implemented a protocol where participants were asked to reproduce 12 random musical notes ranging from C4 to C7. Their performance was captured and analyzed based on these parameters: (1) total time executing a target note, (2) distance to target pitch, (3) number of tuning actions, (4) micro-tuning time and micro-tuning actions count. The parameters definition and computation are detailed in Section 3.3. Participants controlled note triggers and pitch using a single-touch Y-axis position on the touchscreen to reproduce the target notes. This mapping follows the operative guidelines validated in a previous experiment [5]. Both auditory examples and participant-generated notes had a constant amplitude—no mechanism was provided for participants to manipulate note amplitude. The device control was planned for single-hand device manipulation.

An application entitled NemmiProtoDue (https://paginas.fe.up.pt/~up200005343, accessed on 29 June 2022) was developed to conduct the experiment. It adopts Java and Pure Data using libPD [33] for the Android operating system. NemmiProtoDue was able to generate the required auditory, visual, and haptic feedback and was made publicly available through the Google Play Store (https://play.google.com/store, accessed on 29 June 2022) to anyone willing to participate in the experiment. Additionally, the application creates a log file per experiment trial, storing device actuator data (i.e., timestamped touch positions and movement, target pitch, and reached pitch per test case). The logged data was used to analyze the parameters under evaluation for different visual and haptic feedback methods. This experiment was conducted remotely, with participants using their personal devices to run it and submitting the logged data via email at the end. Logged data was not influenced by device specificity, as the only sensor-bound data (touchscreen position coordinates) that might present variability between devices was converted to percentage values (0–100% left-to-right and down-to-up), making it device-agnostic.

Prior to the experiment, participants were first introduced to NemmiProtoDue's interface by conducting a trial phase, which aimed to introduce the note triggering and pitch tuning operations. After the successful completion of the experiment, participants were asked to submit a log file capturing the performed actions via an automatically generated email. In addition to the in-app instructions, a website (https://paginas.fe.up.pt/~up200005343/instructions.php, accessed on 29 June 2022) was created to guide participants in the experiment procedure via video and text descriptions. The experiment required participants to wear wired headphones to minimize external interference (e.g., noise or any other source of sound that might interfere with the listening experience) and ensure audio examples were reproduced correctly by all devices regardless of their hardware. The choice of wired over wireless headphones remained with the need to avoid the loss of audio signal, which is not guaranteed in wireless headphones due to possible network connection instability.

The architecture of the NemmiProtoDue application is shown in Figure 1. It includes two operative modules: the host application and the embedded Pure Data patch. While the Pure Data patch was responsible for all audio-related processes (e.g., note reproduction, triggering the onset and offset), the host was responsible for general test management, sensor and actuator control and data logging. Communication between both the cores was accomplished through Pure Data send and receive objects.
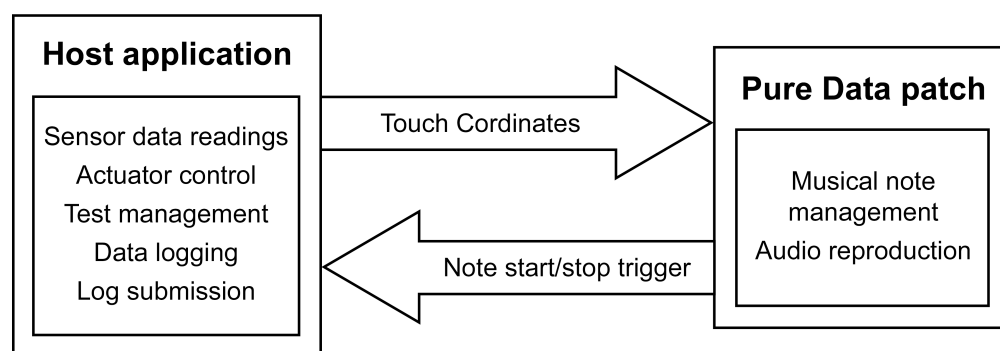
Figure 1. The architecture and data flow of NemmiProtoDue, an application for the Android operating system we designed to run our study experiment.

3.1. Stimuli

The experiment included four test cases, which assessed the following four feedback method combinations: (1) auditory feedback, (2) auditory and visual feedback, (3) auditory and haptic feedback, and (4) auditory, visual and haptic feedback (see Figure 2). Each test case consisted of a participant prompt/action alternation. The application generated an example stimulus the participant was asked to reproduce using the touchscreen's *Y*-axis to find the correct tuning. The application's behavior would sequentially change, bringing into its operation the specific feedback methods.
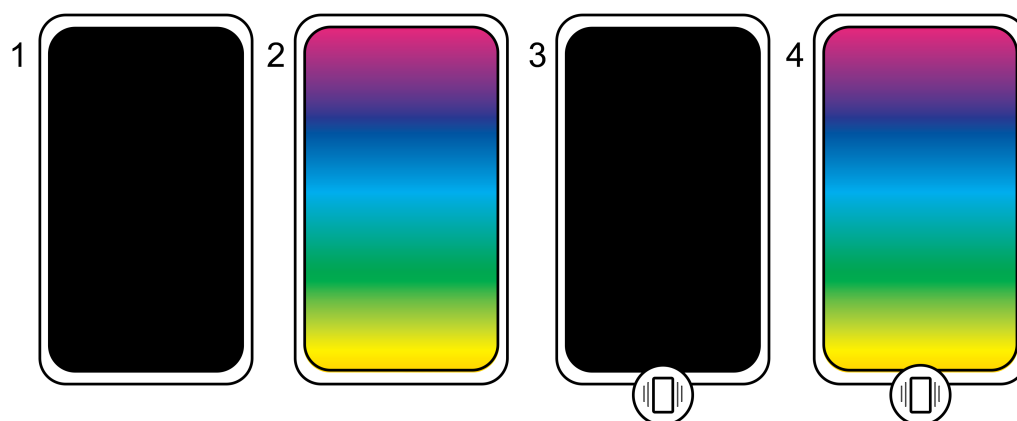

Figure 2. Different test cases featured in the experiment: (**1**) auditory feedback, (**2**) auditory and visual feedback, (**3**) auditory and haptic feedback, (**4**) auditory, visual and haptic feedback.

For each test case, three different example stimuli were presented to the participant, each test case effectively being repeated three times. Each of these test case repetitions is henceforth referred to as an iteration of the experiment. The experiment included 12 iterations in total, resulting from the 3 repetitions of each of the 4 test cases. This threefold repetition of the same test case aims to inspect the potential impact of familiarity in the context of the analyzed parameters.

3.2. Stimuli Feedback Methods

1. Auditory feedback—mapping the user touch's Y-axis position to the generated note's pitch;
2. Visual feedback—mapping the generated note's pitch to the fullscreen application background color);
3. Haptic feedback—mapping the generated note's pitch to the device's vibration intensity).

3.2.1. Auditory Feedback

For the participant's generated notes, musical parameter mappings were taken from the aforementioned previously studied guidelines [5]. Note trigger (onset and offset) and duration were linked to touch gestures. Notes started reproduction when touch started, and was sustained for as long as touch was kept. Notes stopped playing whenever touch ended. Note pitch was mapped to the touchscreen's *Y*-axis with notes ranging from C4 to C7—the lowest vertical position corresponding to the lowest end of the range and the highest vertical position corresponding to the highest end of the range.

The auditory response was synthesized using the Pure Data digital signal processing engine. Notes consisted of a simple sine oscillator whose amplitude was controlled by an audio envelope. The example notes adopted an Attack-Sustain-Release envelope with durations of (respectively) 30 ms, 3000 ms, and 50 ms. Participant-generated notes adopted individual triggering of the Attack and Release stages of the envelope (30 and 50 ms, respectively), with note sustain tied to the participant's touch.

3.2.2. Visual Feedback

The pitch-to-color correspondence was achieved by mapping the note pitch to the color's Hue, following an HSL (Hue, Saturation, Lightness) color format (C4–C7 pitch to 60–360 degrees Hue). Figure 2 shows an approximate color correspondence relative to the screen's *Y*-axis. Studies have consistently found that visual memory and cues outperform and take precedence over the auditory counterpart [34,35]. By using a scale-based visual representation of pitch (arguably the most immediate visual representation for this musical parameter), there would be an immediate visualization of the touch position. This would likely lead the participant to fall back on that as the primary reference source over auditory memory. Our approach provides a visual reference for the participant while preventing the immediate disclosure of a target position.

3.2.3. Haptic Feedback

Considering the lack of a universally available vibration control across all Android devices, we opted to control the vibration intensity utilizing an approach similar to Pulse Width Modulation (PWM). We leveraged the Android SDK's capability of designing a vibration pattern for this actuator, alternating fast On and Off times that would result in a variation in perceived vibration intensity. Contrary to a typical PWM, which uniquely changes the duty cycle time, the generated vibration pattern changed the total cycle time and the duty cycle simultaneously, resulting in a repeatable pattern that provided the desired perceived feedback intensity. This approach was inspired by an article on the TipsForDev website (https://tipsfordev.com/algorithm-for-generating-vibration-patterns-ranging-in-intensity-in-android, accessed on 29 June 2022), which we modified according to participant feedback collected in the pilot experiment. In detail, intensity values below 25% and above 65% were reported as being barely noticeable.

3.3. Experiment Design

Each experiment iteration followed a sequence of five events:

1. The application showed a call-to-action message over neutral black background prompting the participant to listen to an example stimulus;
2. The application played a stimulus note with random pitch and the corresponding feedback methods;
3. The application showed a call-to-action message over neutral black background prompting the user to find the same pitch once the message disappeared;
4. The application started playing audio when participating initiated a touch. At the same time (as applicable), it changed the background color and vibration intensity following patterns described in Section 3.2. Values for audio, visual, and haptic feedback were mapped from the participant touch's *Y*-axis position, and reproduction of all feedback stopped once the touch was released;

5.   A new test case of random type (1–4) would start, going through the total 12 iterations.

The order of the total 12 iterations was randomized (using a uniform random distribution) for each participant to minimize order effects across feedback methods. To collect the participants' performance in note reproduction on mobile-based DMIs, we defined the following five metrics that were collected on an interaction basis:

- Time reaching the end note: time participant took from first starting touch to releasing;
- Final note distance: absolute distance from the end note to the target note;
- Number of tuning actions: total touch position changes (moving the touch corresponds to note tuning as touch position controls note pitch)
- Micro-tuning time: percentage of total iteration time spent in the micro-tuning stage; Micro-tuning actions: percentage of total iteration actions done in the micro-tuning stage.

End note refers to the final note reached by the participant. Target note refers to the example provided, which the participant is attempting to reproduce. We considered micro-tuning as the stage in which participants were executing precise note tuning below a semitone distance to the end note. Any tuning actions occurring sequentially below this threshold at the final stage of text were considered micro-tuning. Although research has shown that time needed to perceive off notes seems to lie under 50 ms [36], this perception ties into the note duration, with longer notes being easier to gauge. Micro-tuning thus becomes noticeably important on longer notes, where pitch deviations become more apparent. We adopted a semitone as the threshold for micro-tuning as it is the smallest interval within Western musical traditions.

To reduce participant's error, we implemented the following two fail-safe conditions during the experiment data collection:

- Time < 200 ms AND number tuning actions < 15;
- Final note distance > 4 semitones AND number tuning actions < 15.

If any experiment iteration verified either of these two conditions, it was considered invalid and silently repeated (i.e., with the participant unaware of this repetition). Each iteration was allowed one single repetition—if this repeat was once again under the specific thresholds, it was not repeated. This fail-safe procedure avoiding "false positive" notes was revealed in a pilot experiment we ran to validate the protocol and resources. Participants would accidentally touch the screen and release the note by accident, resulting in far-off notes from the example's target pitch. Thresholds and conditions were based on the logs gathered from the pilot experiment.

In addition to these in-test measures, additional cutoff measures were applied while processing the collected data from the experiment. Considering the possibility of having invalid iterations logged, even with the silent repeats, a second stage of result selection was applied before statistical analyses. Invalid iterations were counted, and any participant with over 4 logged repeats or a total of repeats and invalid iterations over 4 was considered ineligible for consideration.

3.4. Data Analysis

The statistical analysis of the collected data was conducted in SPSS version 26 (IBM, 2020). Results were analyzed both on an inter and intra-test case basis:

- Inter-test case analysis: this analysis compared all pairwise test case combinations per considered parameter (i.e., time, tuning actions, note distance, micro tuning time, and micro tuning actions). These tests aimed to ascertain any trends concerning the impact of the different feedback methods on the aforementioned parameters. In the inter-test case analysis, we adopt the data resulting from the third test case repetition only, considering that this would represent the iteration where the participants would be most familiar with that specific test case operation;

- Intra-test case analysis: compared the three test case repetitions per feedback method for all parameters under analysis to ascertain any noticeable trends leading towards an impact of familiarity within particular test case operative conditions.

Considering the non-parametric nature of the data and our aim to compare and rank the impact of different feedback methods on the considered metrics, we used the Friedman test to analyze the collected experiment data, which provided us with rankings for all compared datasets. These rankings are generated by comparing individual results for each test case and ranking the data from lowest (rank 4) to highest (rank 1). The individual counts are then organized in a table/graph with the total frequency count for each ranking. For example, if we consider the inter-test case analysis for test time, each participant will have a ranking between test cases, with one taking the least time (rank 4) and one taking the most (rank 1). After creating the rank for each participant, these can be organized in overall frequency graphs, showing the frequency in each rank for each of the compared test cases.

The Friedman test then calculates statistical significance related to the similarity (or difference) of the ranking distribution between the analyzed test cases. A statistically significant result indicates a rejection of the null hypothesis (that the distribution for each of the analyzed sets is the same), meaning there is a significantly different distribution among the analyzed sets.

For statistical analyses, we considered the following significance level convention: Significant result: $p < 0.05$; non-significant result: $p \geq 0.05$.

4. Results

Participants were recruited from several academic mailing lists, social network groups, and word-of-mouth sharing. Participation was voluntary and not subject to selection. One participant was excluded via pre-determined performance cutoffs, as described in Section 3.3, resulting in 60 eligible participants.

Each iteration in the results section, in Figures 3–10, is identified by a string that concatenates the following information: *TestCaseID_RepetitionNumber_ParameterID*. Both the test case ID and repetition number start counting from 0 (i.e., 0–3 for test cases; 0–2 for repetitions). Table 1 details the parameter IDs used in these results and the units in which they are measured.

Table 1. Parameter specifications used in results; percentages are expressed as a number between 0 and 1.

Parameter	Iteration Time	Tuning Actions	Final Note Distance	Micro Tuning Time	Micro Tuning Actions
ID	Time	Actions	Distance	Micropoint	Microactions
Unit	Ms	Count	Semitones	Percentage	Percentage

For example, the first test case using auditory feedback has ID T0; the code for the second repetition of this test case would be T0_1; the full code string for the test time parameter of this iteration would be: T0_1_time.

Due to the underlying requirement of the Friedman test to analyze repeated measures of variance by ranks, multiple frequency charts for the ranking of each set of iteration data will be presented. Each of the individual compared iterations was ranked from the lowest (which always corresponds to the total number of compared iterations) to the highest (always 1) value. Different parameters have different rank interpretations. For the time parameter, the lowest rank represents less iteration time, which is the best result. For tuning actions count, the lowest rank is the test case with the smallest count. For distance, the lowest rank will similarly represent the test case where the participants' end note was closer to the target note's pitch. In the case of micro tuning time and micro

tuning actions, there is less of a positive or negative representation for the ranks. These parameters represent, respectively, a percentage of total iteration time and total tuning action count. This means that a lower ranking for micro tuning time represents less time spent in the micro tuning phase (and consequently greater time spent in macro tuning). In contrast, a lower ranking for micro tuning actions represents a majority of tuning actions being made in the macro tuning phase (less action count percentage in micro tuning). This will become more apparent as we further encounter the graphic representations of each and discuss the results.

4.1. Inter-Test Case Analyses

As previously described in Section 3.4, we considered the last repetition of each test case only and compared rankings between each test case for the defined parameters (i.e., T0_2 for each parameter).

Table 2 shows the statistical significance of the Friedman test results for the inter-test case analysis of the considered parameters. Statistically significant results are shown in bold. We can conclude that the iteration time, final note distance, and the number of tuning actions have no statistically significant results. Therefore, their pairwise comparison ranking distribution per test case does not significantly differ. In other words, no feedback method impacts either of these parameters noticeably. Micro tuning time and micro tuning actions have statistically significant ranking differences. Figures 3 and 4 show the data frequency distribution for the two statistically significant results.

Table 2. Friedman testing significance results for inter-test case analysis; bold: significant result.

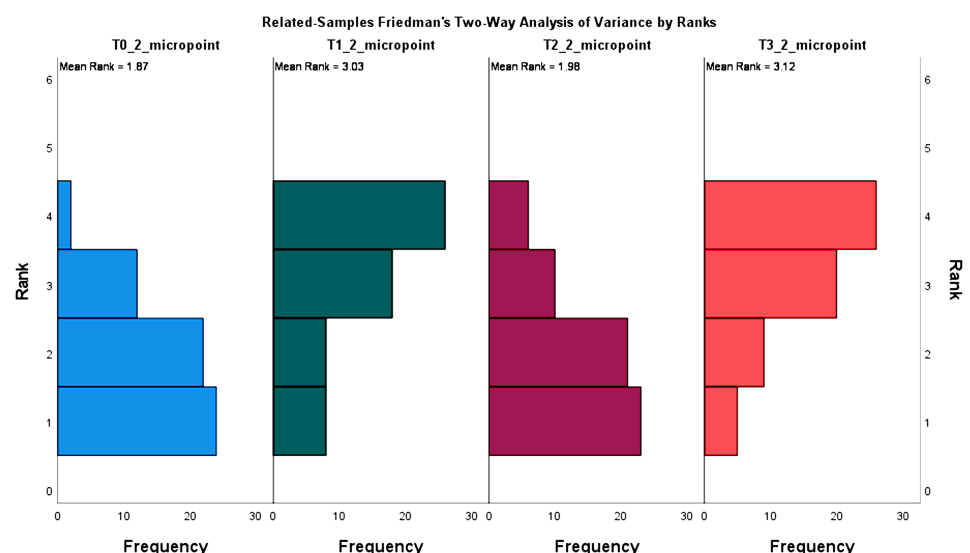| Parameter	Time	Distance	Actions	Micro Tuning Time	Micro Tuning Actions
Significance (p)	0.134	0.460	0.819	**<0.001**	**<0.001**

Figure 3. Inter-test case Friedman test rankings for test micro tuning time.

In Figure 3, we see the ranking representations for the micro tuning time for each of the test cases. Test cases T1 (auditory and visual feedback) and T3 (auditory, visual, and haptic feedback) exhibit a similar distribution, with rank four emerging as the one with the higher count. This indicates that participants spent less time in micro tuning than on

the remaining test cases. The common point between these two test cases is the presence of visual feedback. The two remaining test cases (T0 and T2), where the visual feedback method is absent, show a distribution which appears to be inverted, i.e., participants spent a greater percentage of test time in the micro-tuning phase.
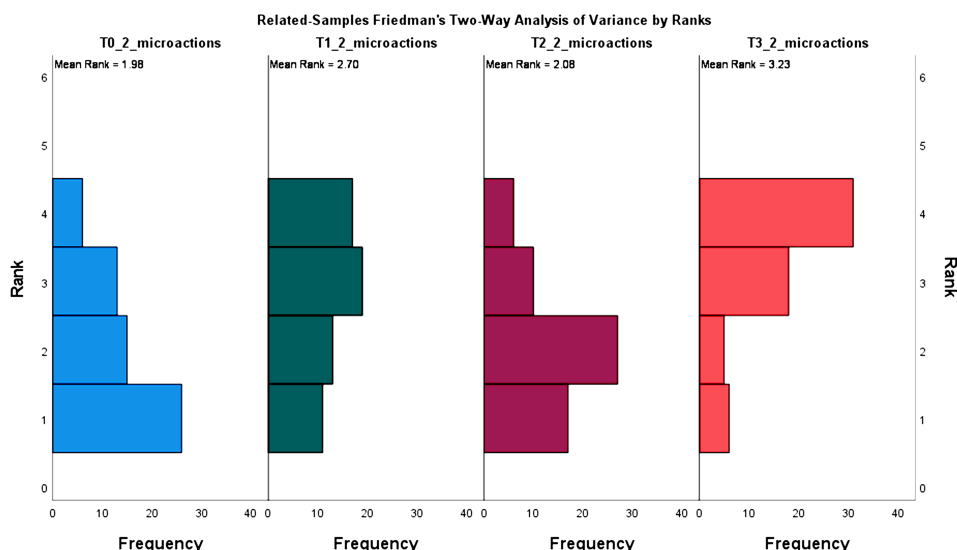

Figure 4. Inter-test case Friedman test rankings for test micro actions.

Figure 4 shows the ranking representations for micro tuning action for each test case. No clear tendencies across the different test cases are observed. However, the overall distribution between T0 (auditory-only) and T3 (auditory, visual and haptic feedback) is inverted. T0 indicates a greater number of total tuning actions performed during the micro tuning phase, conversely to T3, where this phase overwhelmingly is shown to be the one with the lowest percentage.

Our data suggest that none of the visual and haptic feedback methods significantly impact the iteration time, tuning action count, and final note distance when performing note reproduction on mobile-based DMIs. Therefore, these feedback methods do not improve (nor hinder) the tuning accuracy or the required time to reach a target pitch.

On the other hand, micro tuning is shown to be influenced by visual and haptic feedback methods. Results suggest that visual feedback has a pronounced impact on the tuning phases. Visual feedback results in a shorter time spent in the micro tuning phase over macro tuning. Although results show a less pronounced influence in the tuning actions performed during micro tuning, the tuning action count during the micro tuning phase is also reduced when adopting visual feedback. Therefore, visual feedback guides participants to reach the target pitch's vicinity quicker and with fewer subsequent tuning actions.

4.2. Intra-Test Case Analyses

Table 3 shows the results of intra-test case analyses that compared subsequent test case repetitions. The intra-test case analyses were conducted to identify trends in familiarity, i.e., whether the participant's performance showed a gradual amelioration in results over each subsequent repetition). Statistically significant results are shown in bold. Figures 5–10 show the ranking frequencies for the parameters found to be statistically significant.

Table 3. Friedman testing significance results for intra test case analysis; bold: significant result.

Feedback Type	Time	Distance	Actions	Micro Tuning Time	Micro Tuning Actions
Auditory	**0.049**	0.245	**0.016**	**<0.001**	**<0.001**
Auditory + visual	0.522	0.300	0.081	**<0.001**	**<0.001**
Auditory + haptic	0.522	0.218	0.835	**<0.001**	**<0.001**
Auditory + visual + haptic	**0.038**	0.415	**0.035**	**<0.001**	**<0.001**

Similarly to the previously discussed inter-test case results, we can observe that micro tuning time and micro tuning actions present significant results over all test cases. Additionally, statistically significant results on two of the three main parameters, namely the iteration time and the number of tuning actions, are observed in the auditory feedback and the combination of auditory, visuals, and haptic feedback.

Figures 5 and 6 show the ranking frequencies for the two test cases with statistically significant results for the iteration time and tuning action count parameters—test cases 1 (T0—auditory feedback only) and 4 (T3—auditory, visual and haptic feedback). In all four cases, both parameters have a visibly negative tendency at each subsequent repetition. The first repetition for both T0 and T3 exhibits a greater number of participants, ranking this as the fastest iteration (rank 3 = less time), while this distribution seems to become almost evenly distributed in the second repetition and inverted on the third one, with this repetition shown to be the longest one. This tendency is also observed concerning the total number of tuning actions, with the evolution following the same pattern, indicating a worsening result for each subsequent repetition.
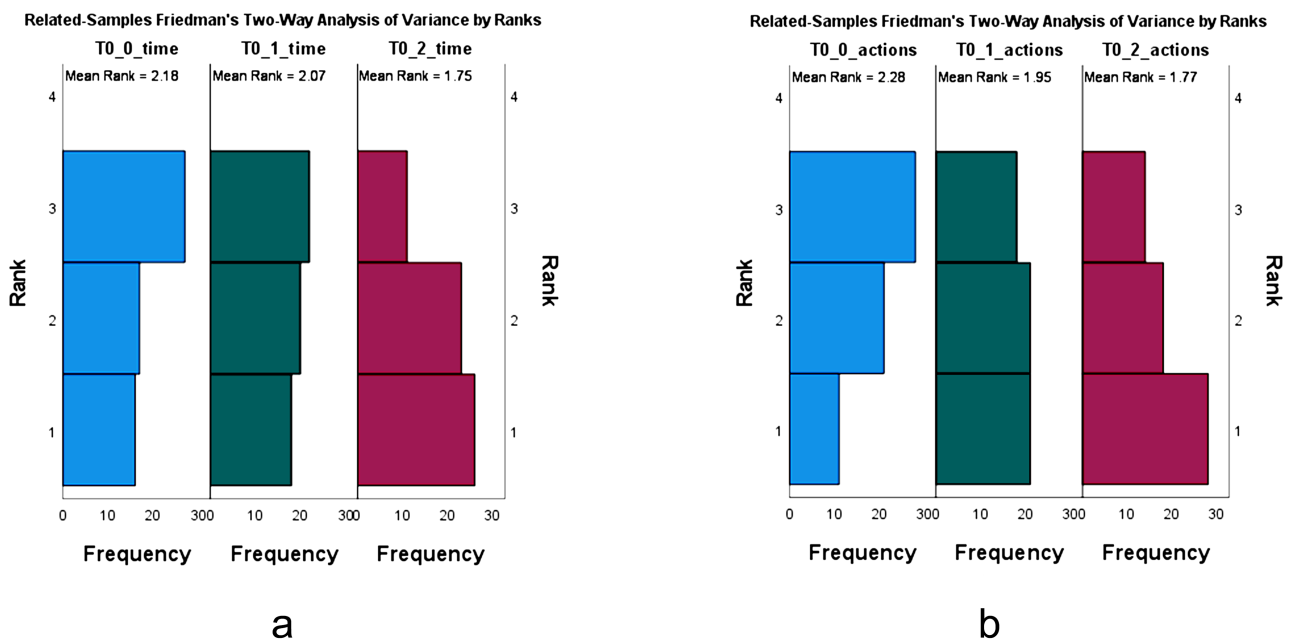
Figure 5. Intra-test Friedman rankings for auditory-only test case. (**a**) test iteration time, (**b**) test iteration action count.

All test cases exhibit statistically significant results for micro tuning time and micro tuning action count and similar rankings for both parameters, whose ranking charts are shown in Figures 7–10. There seems to be a pairwise similarity for both micro tuning time and action count between T0 and T2 and T1 and T3. In the case of T0 and T3, the second repetition is shown to be ranked third for most participants, meaning that they spent a

smaller percentage of the total test time in the micro tuning phase, also executing a lesser percentage of total tuning actions in this phase. Conversely, in the case of T1 and T3, participants spent more time micro tuning while performing more tuning actions in the second repetition. The common link between these pairs seems to be tied to visual feedback, with T1 and T3 being the test cases where this feedback method is present.
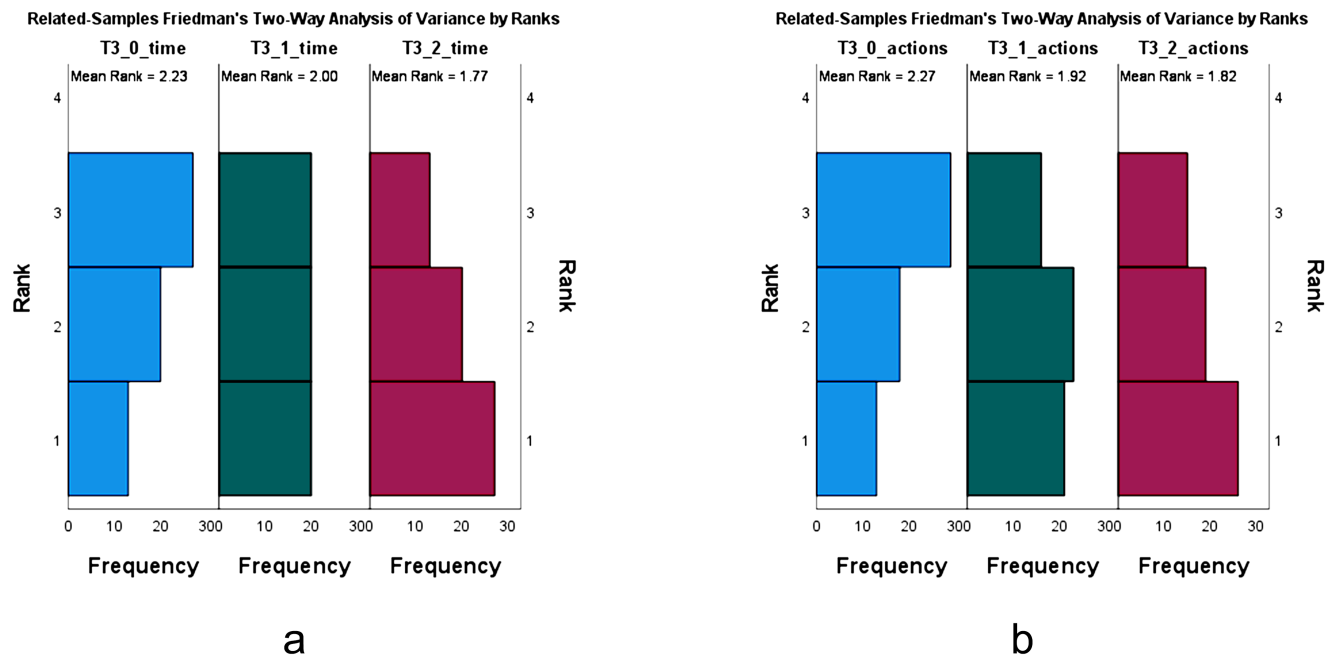

Figure 6. Intra-test Friedman rankings for auditory, visual, and haptic test case. (**a**) test iteration time, (**b**) test iteration action count.
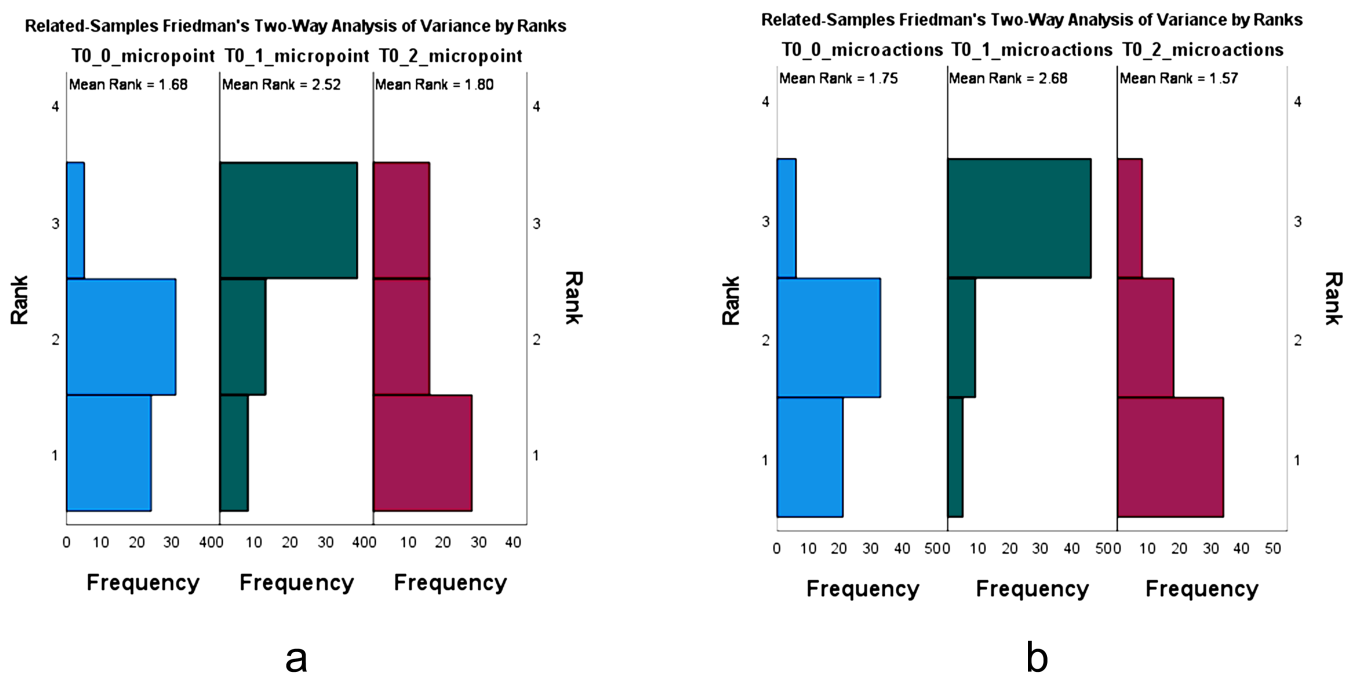

Figure 7. Intra-test Friedman rankings for the auditory-only test case. (**a**) test iteration micro tuning time, (**b**) test iteration micro tuning actions.
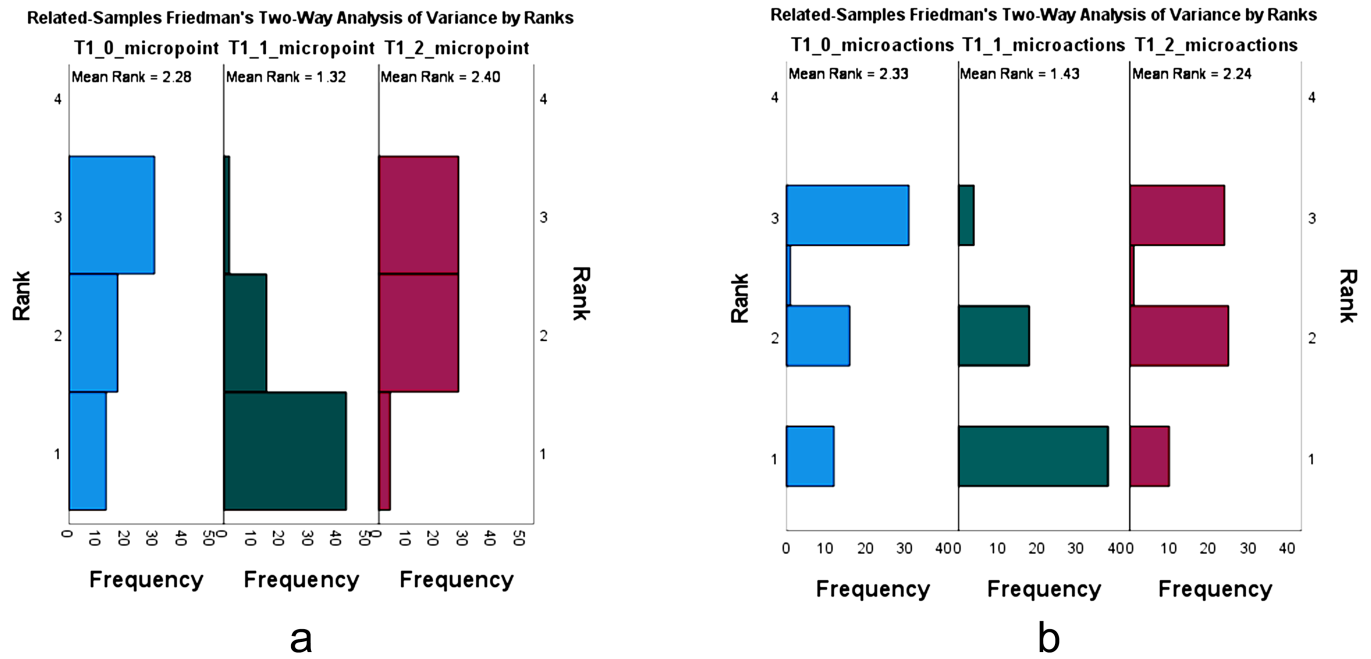
Figure 8. Intra-test Friedman rankings for the auditory and visual test case. (**a**) test iteration micro tuning time, (**b**) test iteration micro tuning actions.
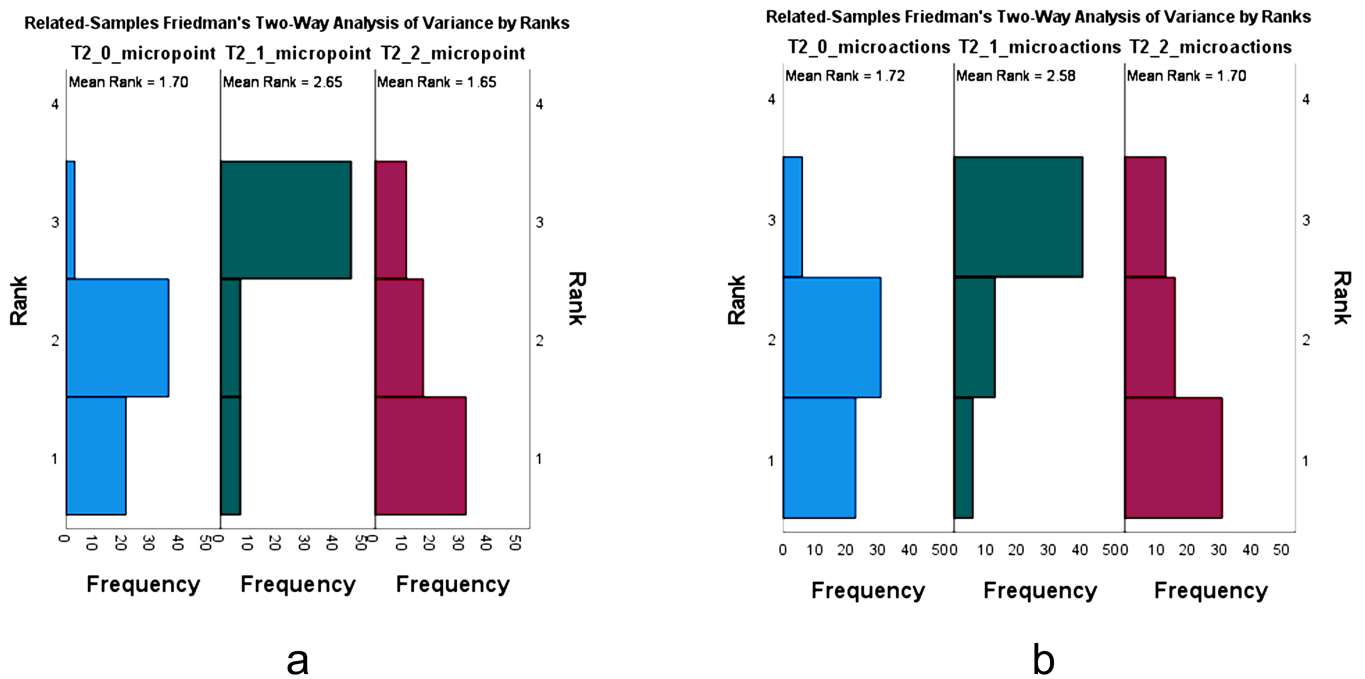

Figure 9. Intra-test Friedman rankings for the auditory and haptic test case. (**a**) test iteration micro tuning time, (**b**) test iteration micro tuning actions.
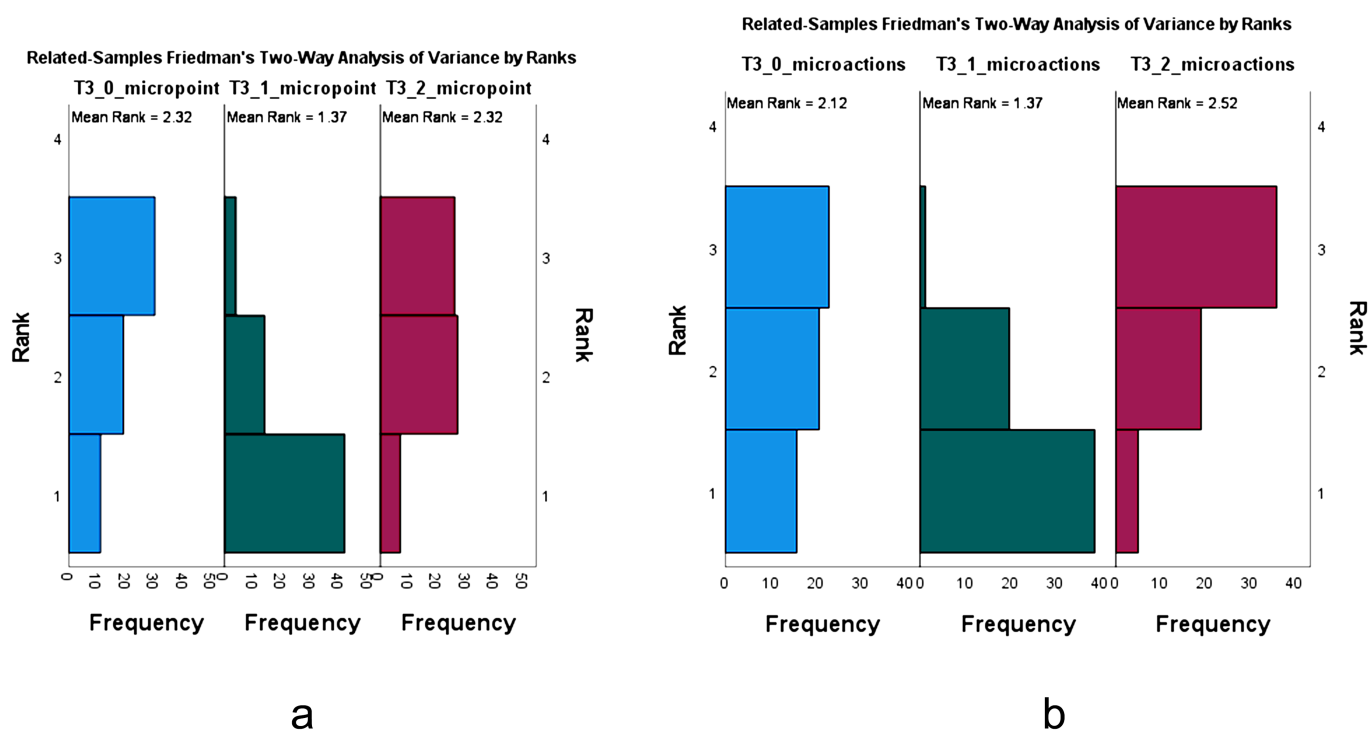
Figure 10. Intra-test Friedman rankings for the auditory, visual and haptic test case. (**a**) test iteration time, (**b**) test iteration action count.

The expectancy, in the presence of a clear tendency, would be to observe an improvement or degradation with each subsequent repetition (indicative of some sort of cumulative influence of repetition) or to simply observe consistent results (indicative of no influence of repetition). This apparent alternation of results, with the first and third repetitions having somewhat similar ranking distributions, and the second showing a completely reversed distribution, indicates no perceivable tendency and allows no grounds for establishing a clear conclusion—other than the need for additional testing to ascertain if there is any pattern behind this or if it is a product of randomness.

5. Conclusions

Building upon our previous research concerning design guidelines for mapping user input in mobile-based DMIs [5], we wanted to evaluate how these tools could ideally provide feedback to users during operation. Specifically, we wanted to ascertain if visual and haptic feedback would be viable ways of complementing the obvious auditory feedback of these instruments. We designed an experiment and protocol allowing us to analyze the performance of 60 participants over a set of parameters relating to tuning time and accuracy, and precision pitch tuning.

We found that while visual feedback does provide positive reinforcement by seemingly helping experiment participants reach the vicinity of targeted notes faster and with less hesitation, for the other considered metrics, there seems to be little benefit to introducing either additional feedback method. The experiment results provide evidence concerning the four questions behind our study, which concern the impact of auditory, visual, and haptic feedback methods on the effectiveness of mobile-based DMIs' musical note reproduction. Next, we enunciate the main conclusions in response to each question. Furthermore, the results point to hypotheses and open questions to be tackled in future work.

1. Does visual feedback impact pitch tuning time and accuracy?
    We found that visual feedback did not show any pronounced impact on pitch tuning time and accuracy. No significant results were found to indicate that participant

results improved or worsened for either of the three considered parameters (iteration time, tuning action count, and note pitch distance) due to visual feedback;

2. Does haptic feedback impact pitch tuning time and accuracy?
    We found that haptic feedback did not show any pronounced impact on pitch tuning time and accuracy. No significant results were found to indicate that participant results improved or worsened, for either of the three considered parameters, due to the presence of haptic feedback;

3. Do these additional visual and haptic feedback methods impact macro and micro note tuning?
    Experiment results suggested that visual feedback reinforcing auditory feedback impacts macro and micro note tuning. Visual feedback was shown to positively impact participant performance, allowing them to reach the vicinity of the targeted note more easily, thus reducing the need for pitch micro tuning. Haptic feedback was found to have no significant impact on macro and micro note tuning;

4. Is there a perceivable degree of familiarity for visual and haptic feedback methods in note reproduction on mobile-based DMIs?
    For the three main parameters (iteration time, number of tuning actions, and note distance), both the auditory feedback and auditory, visual and haptic feedback test cases (seemingly unrelated) showed similar tendencies. However, these were not the expected positive tendencies indicative of performance amelioration with familiarity, but rather the opposite—a worsening in performance. Were it not for the randomization of the order of test iterations, one could attempt to find justification for some type of gradual fatigue, resulting in this degradation of performance. Considering the implemented experiment procedure, this aspect cannot be seen as the main driver for this observed tendency. At the same time, contrary to inter-test case testing, micro tuning time and micro tuning action count percentages have shown somewhat inconsistent results, with no clear tendency but with an unexplained repeated pattern. Based on these results, there seems to be no perceivable influence of these feedback methods over training or familiarity.

These conclusions, however, are accompanied by a need to explore this field further. There remains a lot to be studied in terms of the impact of multimodal feedback for mobile-based DMIs in the hopes of finding, or at least building towards, a common practice and approach to designing these instruments. Results are conclusive regarding the impact of the specific additional feedback methods used to reinforce auditory feedback on tuning time and accuracy. Nonetheless, it would also be interesting to extend this experiment to a greater number of participants and run parametric analyses to ascertain if the overall results for each parameter do indeed show pragmatically better or worse results with each of the additional feedback methods (e.g., do participants average better times with or without visual feedback?). The conclusiveness of these results was accompanied by the questions brought up by the intra-test case testing, bringing up the need for further study in terms of familiarity and training potential. Specifically, the results concerning this aspect exhibited patterns requiring additional or different testing to better describe and validate. Perhaps using only three repetitions in a short and contained experiment is not enough to properly gauge the effects of training and familiarity, and a longer, more ongoing experiment would be best suited for this end.

In this particular experiment, we chose to focus on the general public as the target of the study. We took this decision based on the assumption that just as traditional instruments are accessible to anyone, this kind of instrument should also be. Nonetheless, another very important avenue for exploration lies in the potential distinction between musically trained and non-trained individuals, particularly considering instrument performance experience. This would allow for further understanding of the impact of familiarity and training on the operation of MHD-based DMIs. Pitch perception is something that improves with training [36], and musical training would be expected to impact the results of these tests. Similarly, it can be expected for experienced instrument performers to be more sensitive to

having vibrotactile cues as a natural feedback mode of musical instruments and potentially be better suited for assimilating additional haptic feedback.

An evolution, or extension of this experiment and protocol, extending not only the number of participants but the time of contact (or times of contact) with the application, would complement these results.

Institutional Review Board Statement: Ethical review and approval were waived for this study, as no ethical issues were involved (e.g., no vulnerable populations, no collection of sensitive issues, no distress situations, invasive activities, or collection of biological materials).

Informed Consent Statement: Patient consent was waived as no identifiable data was collected or registered.

Data Availability Statement: The data presented in this study are openly available in Zenodo at https://doi.org/10.5281/zenodo.6794197, accessed on 29 June 2022.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MHD	Mobile Handheld Device
DMI	Digital Music Instrument
DMCI	Digital Music Control Interface
LED	Light-Emitting Diode
PWM	Pulse-Width Modulation

References

1. Ardiel, E.L.; Rankin, C.H. The importance of touch in development. *Paediatr. Child Health* **2010**, *15*, 153–156. [CrossRef] [PubMed]
2. Blackwell, P.L. The Influence of Touch on Child Development: Implications for Intervention. *Infants Young Child.* **2000**, *13*, 25–39. [CrossRef]
3. Playability-Cambridge Dictionary Definition. 2022. Available online: https://dictionary.cambridge.org/dictionary/english/playability (accessed on 28 June 2022).
4. Holst, A. Number of Smartphone Users Worldwide from 2016 to 2021 (in Billions). 2019. Available online: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide (accessed on 9 February 2022).
5. Clément, A.; Moreira, L.; Rosa, M.; Bernardes, G. Musical Control Gestures in Mobile Handheld Devices: Design Guidelines Informed by Daily User Experience. *Multimodal Technol. Interact.* **2021**, *5*, 32. [CrossRef]
6. Maes, P.J.; Leman, M.; Lesaffre, M.; Demey, M.; Moelants, D. From expressive gesture to sound: The development of an embodied mapping trajectory inside a musical interface. *J. Multimodal User Interfaces* **2010**, *3*, 67–78. [CrossRef]
7. Hattie, J.; Timperley, H. The Power of Feedback. *Rev. Educ. Res.* **2007**, *77*, 81–112. [CrossRef]
8. Dumas, B.; Solórzano, M.; Signer, B. Design guidelines for adaptive multimodal mobile input solutions. In Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services—MobileHCI'13 , Munich, Germany, 27–30 August 2013; p. 285. [CrossRef]
9. Sá, M.D.; Carriço, L.; Duarte, C.; Reis, T. *Designing Mobile Multimodal Applications*; IGI Global: Hershey, PA, USA, 2010. [CrossRef]
10. Jacko, J.A.; Scott, I.U.; Sainfort, F.; Moloney, K.P.; Kongnakorn, T.; Zorich, B.S.; Emery, V.K. *Effects of Multimodal Feedback on the Performance of Older Adults with Normal and Impaired Vision*; Springer: Berlin/Heidelberg, Germany, 2003; Volume 2615, pp. 3–22.
11. Sun, M.; Ren, X.; Cao, X. Effects of Multimodal Error Feedback on Human Performance in Steering Tasks. *J. Inf. Process.* **2010**, *18*, 284–292. [CrossRef]
12. Vitense, H.S.; Jacko, J.A.; Emery, V.K. Multimodal feedback: Establishing a performance baseline for improved access by individuals with visual impairments. In Proceedings of the Fifth International ACM Conference on Assistive Technologies—Assets '02, Edinburgh, Scotland, 8–10 July 2002; p. 49. [CrossRef]

13. Vitense, H.S.; Jacko, J.A.; Emery, V.K. Multimodal feedback: An assessment of performance and mental workload. *Ergonomics* **2003**, *46*, 68–87. [CrossRef] [PubMed]
14. Di Domenico, A. An Evaluation of the Impact of Multimodal Feedback on Interaction Performance and Experience in Mobile Gaming. Master's Thesis, Utrecht University, Utrecht, The Netherlands, 2014.
15. Seebode, J.; Schleicher, R.; Möller, S. Investigating Multimodal Feedback in a Mobile Interaction Paradigm. In Proceedings of the CD of FORUM ACUSTICUM 2014, Krakow, Poland, 7–12 September 2014.
16. Broll, G.; Palleis, H.; Richter, H.; Wiethoff, A. Exploring multimodal feedback for an NFC-based mobile shopping assistant. In Proceedings of the 2013 5th International Workshop on Near Field Communication (NFC), Zurich, Switzerland, 5 February 2013; pp. 1–4. [CrossRef]
17. Burke, J.L.; Prewett, M.S.; Gray, A.A.; Yang, L.; Stilson, F.R.B.; Coovert, M.D.; Elliot, L.R.; Redden, E. Comparing the effects of visual-auditory and visual-tactile feedback on user performance: A meta-analysis. In Proceedings of the 8th International Conference on Multimodal interfaces—ICMI '06, Banff, AB, Canada, 2–4 November 2006; p. 108. [CrossRef]
18. Hsu, H.Y.; Lin, C.W.; Lin, Y.C.; Wu, P.T.; Kato, H.; Su, F.C.; Kuo, L.C. *Effects of Passive Haptic Learning in Music-Supported Therapy on Sensorimotor Performance in Older Adult Hands: A Randomized Crossover Trial*; 16 July 2021, PREPRINT (Version 1) available at Research Square. Available online: https://doi.org/10.21203/rs.3.rs-594506/v1 (accessed on 4 July 2022).
19. Zand, M.; Kyrarini, M. Sensing Visual Art by Relatable Music and Haptic Feedback for Individuals with Visual Impairments. In Proceedings of the 15th International Conference on PErvasive Technologies Related to Assistive Environments, Corfu, Greece, 29 June–1 July 2022; pp. 310–311. [CrossRef]
20. Lee, M. Entangled: A Multi-Modal, Multi-User Interactive Instrument in Virtual 3D Space Using the Smartphone for Gesture Control. In Proceedings of the International Conference on New Interfaces for Musical Expression, Shanghai, China, 14–18 June 2021. [CrossRef]
21. Piao, Z.; Xia, G. Sensing the Breath: A Multimodal Singing Tutoring Interface with Breath Guidance. *arXiv* **2022**, arXiv:2202.01439. https://doi.org/10.21428/92fbeb44.143bb1d4.
22. Zhang, Y.; Li, Y.; Chin, D.; Xia, G.G. Adaptive Multimodal Music Learning via Interactive-haptic Instrument. In Proceedings of the International Conference on New Interfaces for Musical Expression, Porto Alegre, Brazil, 3–6 June 2019; p. 6.
23. Chang, A.; O'Sullivan, C. Audio-haptic feedback in mobile phones. In Proceedings of the 2005 Conference on Human Factors in Computing Systems, Portland, OR, USA, 2–7 April 2005; p. 1264. [CrossRef]
24. Roberts, C.; Forbes, A.; Höllerer, T. Enabling Multimodal Mobile Interfaces for Interactive Musical Performance. In Proceedings of the International Conference on New Interfaces for Musical Expression, Ann Arbor, MI, USA, 21–23 May 2012; p. 4.
25. Altinsoy, M.E.; Avanzini, F.; Beck, S.D.; Berdahl, E.; Blandino, M.; Bouwer, A.; Cadoz, C.; Castagné, N.; Fontana, F.; Fritz, C.; et al. *Musical Haptics*; Springer: Charm, Switzerland, 2018. [CrossRef]
26. Turchet, L.; Barthet, M. Co-Design of Musical Haptic Wearables for Electronic Music Performer's Communication. *IEEE Trans. Hum. Mach. Syst.* **2019**, *49*, 183–193. [CrossRef]
27. Turchet, L.; West, T.; Wanderley, M.M. Touching the audience: Musical haptic wearables for augmented and participatory live music performances. *Pers. Ubiquitous Comput.* **2021**, *25*, 749–769. [CrossRef]
28. Boyer, E.O.; Vandervoorde, L.; Bevilacqua, F.; Hanneton, S. Touching sounds: Audio virtual surfaces. In Proceedings of the 2015 IEEE 2nd VR Workshop on Sonic Interactions for Virtual Environments, SIVE 2015—Proceedings, Arles, France, 24 March 2015; pp. 1–5. [CrossRef]
29. Williams, J. TouchTable: A Musical Interface with Haptic Feedback for DJs | Haptic Intelligence. Available online: https://hi.is.mpg.de/research_projects/touchtable (accessed on 22 July 2022).
30. Fontana, F.; Paisa, R.; Ranon, R.; Serafin, S. Multisensory Plucked Instrument Modeling in Unity3D: From Keytar to Accurate String Prototyping. *Appl. Sci.* **2020**, *10*, 1452. [CrossRef]
31. Hwang, S.; Bianchi, A.; Wohn, K.Y. VibPress: Estimating pressure input using vibration absorption on mobile devices. In Proceedings of the MobileHCI 2013—15th International Conference on Human-Computer Interaction with Mobile Devices and Services, Munich, Germany, 27–30 August 2013; pp. 31–34. [CrossRef]
32. Tung, Y.C.; Shin, K.G. ForcePhone: Software Lets Smartphones Sense Touch Force. *IEEE Pervasive Comput.* **2016**, *15*, 20–25. [CrossRef]
33. Brinkmann, P.; Mccormick, C.; Kirn, P.; Roth, M.; Lawler, R. Embedding Pure Data with libpd. In Proceeding of the Fourth International Pure Data Convention, Weimar, Germany, 8–12 August 2011; pp. 291–301.
34. Hilton, E.; Gottwald, R. Differences in Visual and Auditory Short-term Memory. *Undergrad. Res. J.* **2001**, *4*, 47–50.
35. Lindner, K.; Blosser, G.; Cunigan, K. Visual versus auditory learning and memory recall performance on short-term versus long-term tests on short-term versus long-term tests. *Mod. Psychol. Stud.* **2009**, *15*, 6.
36. Fyk, J. Duration of Tones Required for Satisfactory Precision of Pitch Matching. In *Bulletin of the Council for Research in Music Education*; University of Illinois Press: Champaign, IL, USA, 1987; pp. 38–44.