Supporting Rhythm Activities of Deaf Children using Music-Sensory-Substitution Systems

Benjamin Petry
bpetry@acm.org

Thavishi Illandara
thavishi@ahlab.org

Don Samitha Elvitigala
samitha@ahlab.org

Suranga Nanayakkara
suranga@ahlab.org

Augmented Human Lab
Singapore University of Technology and Design, Singapore

Figure 1. In this paper, we (a) present MuSS-Bits++, a music-sensory-substitution system for music-making, (b) conducted a controlled study to investigate how it affects rhythm perception of deaf children, and (c) observed its use in music lessons at a deaf school.

ABSTRACT

Rhythm is the first musical concept deaf people learn in music classes. However, hearing loss limits the amount of information that allows a deaf person to evaluate his or her performance and stay in sync with other musicians. In this paper, we investigated how a visual and vibrotactile music-sensory-substitution device, MuSS-Bits++, affects rhythm discrimination, reproduction, and expressivity of deaf people. We conducted a controlled study with 11 deaf children and found that most participants felt more confident wearing the device in vibration mode even when it did not objectively improve their accuracy. Furthermore, we studied how MuSS-Bits++ can be used in music classes at deaf schools and what challenges and opportunities arise in such a setting. Based on these studies, we discuss insights and future directions that support the design and development of music-sensory-substitution systems for music making.

ACM Classification Keywords

K.4.2 Social Issues: Assistive technologies for persons with disabilities

Author Keywords

Music, Design, Sensory Substitution, Assistive Technology, Deaf

CHI 2018, April 21–26, 2018, Montreal, QC, Canada

DOI: https://doi.org/10.1145/3173574.3174060

INTRODUCTION

Music-making depends on a strong coupling of perception and motor processes [30, 29] that form a closed feedback loop consisting of (1) play, (2) perception, (3) interpretation, (4) evaluation, and (5) adjustment/synchronization. Hearing loss can create a gap in the feedback loop since it limits the auditory information that deaf musicians can employ to evaluate their performance. In the last decade, there have been several sensory-substitution approaches aiming to provide an alternative way for deaf people to perceive music. Most systems, such as the Emoti-Chair [15] and the Haptic-Chair [21], have been developed to enhance the *music-listening* experience for deaf people using visual and vibrotactile cues. Few systems, such as MOGAT [42] and "Tactile Sounds" [23], use sensory-substitution to support *music-making*.

Establishing a steady-beat and introducing rhythm to deaf children are the first exercises when teaching music to deaf children [8]. Thus, 'conveying rhythm information' should be an essential requirement for music-sensory-substitution systems that aim to support the music-learning process. However, there is little knowledge on how prior music-sensory-substitution systems affect rhythm perception. In this work, we present an implementation of a high-fidelity prototype of a music-sensory-substitution system that provides rhythm information to a deaf user in real-time. We deployed this prototype in a controlled and a field study to understand how it affects rhythm perception and how it can be used in music classes. This work contributes to the pool of music-sensory-substitution systems in four main ways:

- First, we present **a preliminary survey of challenges in teaching music classes** to deaf children based on a semi-structured interview with a music teacher.

- Second, we introduce the **implementation and technical evaluation of MuSS-Bits++**, a wearable plug-and-play music-sensory-substitution system which supports rhythm perception (see Figure 1a). MuSS-Bits++ is an improved version of MuSS-Bits [26, 25] as it (1) preprocesses the audio signal with a 512-bin FFT, enabling frequency filtering, (2) uses NRF communication to reduce power consumption, and (3) has a smaller and easy-to-wear form-factor, which was requested by participants of prior studies [26].
- Third, we **investigate how MuSS-Bits++ affects rhythm perception**. We collected performance and physiological data in a controlled study with 11 deaf children (see Figure 1b). Each child participated in seven sessions and went through rhythmic discrimination, reproduction, and expressivity tasks. Furthermore, we investigated the usage of MuSS-Bits++ in music classes to complement the observations from the controlled study (see Figure 1c). We found deaf children more confident wearing MuSS-Bits++ and more engaged in music classes. Moreover, they reported to understand the music teacher's instructions better with vibrations from MuSS-Bits++.
- Last, we discuss **insights** that will help other researchers to develop and investigate future music-sensory-substitution systems for music-making.

RELATED WORK & BACKGROUND

Music and the Deaf

Deaf people enjoy and participate in musical activities [22, 5, 11]. For example, deaf musicians such as Evelyn Glenny [13], Janine Roebuck [28], and Sean Forbes [10] made music their profession. Furthermore, organizations, such as 'Music and the Deaf' [1], aim to encourage deaf people in music-making activities. While the literature suggests that the interest in music highly depends on whether a deaf person associates him- or herself with the deaf or hearing culture [6], the interest in music is very much a personal choice.

Music-Sensory-Substitution Systems

While hearing aids and cochlear implants get increasingly common for speech recognition improvement, they are also known to distort music and hence, reduce the enjoyment of music [4, 7, 12, 20]. Furthermore, speaker listening, where a person places his or her hand close to a speaker or holding a balloon to feel the amplified vibrations, are common strategies among deaf people to listen to music [5] and are similar to the Tadoma method for speech perception. However, this is less practical for making music, where hands usually interact with an instrument. Music-sensory-substitution systems are another approach to convey music to deaf people. These systems expect music as input and map it to one or more alternative sensory channels while preserving one or more key characteristics of music (e.g., rhythm). The main challenge in building music-sensory-substitution systems is the design of the mapping [16]. Especially, the creation of a music-to-visual mapping turns out to be difficult, as the brain's processing of audio and visual information has little overlapping [2]. On the contrary, audio and tactile information are very similar in their nature [41], though their sensory bandwidths are very different (ear: 20Hz-20kHz, skin: up to 1000Hz [35, 40]). Nonetheless, Shibata [33] found that deaf people process tactile information in the same part of the brain where hearing people process auditory information.

Petry et al. [26] compiled a list of existing music-sensory-substitution systems and classified those systems among dimensions such as 'suitability to explore sound,' 'cater for user customization,' and 'feedback modality.' The authors found that most music-to-visual approaches used screens to project visual content, but the mappings were quite different among those systems. In contrast, vibrotactile approaches had same or similar mappings for pitch, loudness and time and mainly used voice coil actuators. Vibrotactile systems mostly differentiate in the body location where the user perceives the vibrotactile feedback. Recently, new music-sensory-substitution systems have been introduced to the market: (1) Das Sound Shirt [1] - a vibrotactile shirt that conveys classical music of an orchestra to a deaf person, (2) SUBPAC M2 [2] - a subwoofer that is carried on the back of a person, (3) The Basslet [3] - a watch-sized subwoofer worn around the wrist, and (4) Audiolux One [4] - a visual system representing sound in different shapes and places on a stage. While the pool of music-sensory-substitution systems is growing, little is known about how these systems affect deaf people in following and making rhythms.

Understanding Challenges from a Teacher's Point-of-View

An observational study with deaf children was conducted by Petry et al. [24]. The authors provided different instruments to the children and asked them to perform musical tasks. They found two main strategies to receive feedback and two main strategies to play instruments that the deaf children employ: (1) focusing on the area of action and (2) looking at the audience, as well as (3) mimicking other people and (4) counting. Furthermore, half of the participants were not sure whether they played correctly, indicating a lack of confidence which in return may have affected the user experience. In this paper, we were motivated by the perspective of a teacher who runs music classes for deaf children.

Specifically, we conducted a semi-structured interview with a music teacher who has been conducting music lessons at a deaf school for three years. The interview aimed to understand challenges and strategies he developed to teach rhythm to deaf children. For the music teacher, teaching music to deaf children is different compared to teaching hearing children: *"[Hearing children] follow what I am saying, what I am doing. I can say what I am doing. But [teaching to deaf children] is difficult. How to think, how to get the rhythm, how to give the rhythm? [...] It is a challenge."*

Furthermore, keeping one rhythm seems to be a problem during drumming and dancing session: *"I make a group with a lot of drums. Then I start to play with them, and slowly everyone plays; but not rhythmic sounds [...] they cannot keep one*

---

[1] https://sound-shirt.jimdo.com/

[2] https://shop.subpac.com/

[3] https://eu.lofelt.com/products/basslet

[4] https://www.kickstarter.com/projects/audioluxdevices/audiolux-one-smart-led-system-for-sound-reactive-v

rhythm." Dancing *"is a challenge actually. They don't know rhythm [...] So I am following the rhythm and show [with my body movements] the rhythms to them, and they follow me."* His teaching strategy was to use counting: *"How to make beats [...] I draw numbers 1 2 3 4 1 2 3 4 [...] for counting."* For dancing, the children mainly deploy a mimicking strategy: *"I am dancing, I show [the] rhythms and they capture it, and they dance. [...] Two of them can hear a little bit, and they can follow rhythm; then the others follow them. [...] Their eyes are so [more] powerful than ours,[...] they capture the details."*

Our interview is an illustration of the difficulties that the teachers in deaf schools may face when they are not specially trained. In countries, such as the UK, where a bigger deaf community with an interest in music exists, music teachers can receive special training. However, developing countries (similar to where this teacher comes from) mostly do not have this support, and hence the teachers have to come up with their own strategies and resources. Thus, a music-sensory-substitution system that successfully provides rhythm information to a deaf user has the potential to improve and ease music-teaching.

MUSS-BITS++

MuSS-Bits++ was developed through an iterative user-centered design process and builds upon MuSS-Bits [26, 25]. User feedback suggested the device to be smaller and similar to modern smartwatches, as well as being able to differentiate instruments, particularly voice and drums. In this section, we describe the changes we made in hardware and firmware to satisfy these suggestions. We conducted a technical analysis, to ensure that MuSS-Bits++ can provide real-time feedback. The verification of real-time feedback is of particular importance, as rhythm information is sensitive to delays [18, 31].

Hardware

Like its predecessor, MuSS-Bits++ consists of wireless sensor-display pairs (Sensor- and Display-Bit, see Figure 1a). Major changes we made are (1) the use of a PIC33EP512GP502 processor to execute a real-time 512-bin FFT (Fast-Fourier-Transformation) for detailed filtering of frequencies and (2) reducing the form factor to a smartwatch size. The reduction of the form factor resulted in additional challenges such as the optimization for power consumption with an alternative wireless solution that reduces power consumptions but provides a large bandwidth for real-time feedback. To solve this problem we used the NRF24L01P module which has a maximum power consumption of 13.5mA for receiving and 11.3mA for transmitting. Compared to the previously used ESP8266-12F, this NRF module reduced the power consumption by a factor of 11.

Other changes include the replacement of a single RGB-LED by four NeoPixel LEDs for more visual possibilities. Furthermore, we used capacitive touch to control visual and vibrotactile intensity instead of potentiometers to minimize the form factor further. In addition, we changed the attachment mechanism to detachable straps and casings with micro-suction tape. Thus, MuSS-Bits++ can be combined with existing straps such as AppleWatch and running straps (e.g., the EdgeGear Shift™[5]) as well as can be attached to objects, such as a wall or instruments. The final size of both devices is 4.4cm × 3.5cm × 1.5cm, and their weight is about 25g, which is lighter than current smartwatches[6].

Firmware

MuSS-Bits++'s firmware is written in C. The Sensor-Bit firmware follows a modular architecture: (1) analog-to-digital conversion, (2) FFT, (3) musical element extraction, (4) music-to-visual and -vibrotactile mappings, and (5) NRF communication. The architecture was designed to easily exchange the musical element extraction algorithm as well as the mappings to make MuSS-Bits++ easily extendable. The Display-Bit firmware architecture is modular as well: (1) NRF communication, (2) intensity adjustments based on the capacitive touch input, and (3) generation of the motor and NeoPixel driving signal. As the motor has a non-linear response, we applied adjustments to the motor driving signal to create linear vibration feedback.

The musical element extraction algorithm has been derived from MuSS-Bits++ predecessor but uses the FFT to eliminate high-frequency noise. The algorithm calculates the average energy (e_{avg}) for all frequencies below 250Hz that contain energy values above 0. The mapping function uses e_{avg} as input and calculates the motor's and NeoPixel's driving signal, which results in vibration and brightness intensity.

EVALUATION

Our main question was whether and how MuSS-Bits++ can provide effective access to rhythm information so that deaf children could discriminate and reproduce rhythms, as well as expand their musical expressivity space. First, we conducted a technical evaluation to assess whether the MuSS-Bits++ prototype is sufficiently accurate in providing rhythm information in real-time. Second, we conducted a controlled user study with 11 deaf children to quantitatively describe the effect of MuSS-Bits++ on rhythm perception. We followed up with another study in a real setting with three of the deaf children and a music teacher who were using MuSS-Bits++ in music lessons. We first briefly report about the technical analysis followed by the controlled study. We then describe our qualitative findings from the music lessons and major insights.

Analysis of Technical Accuracy

"Rhythm refers to the durations of a series of notes, and to the way that they group together into units" [19]. To convey rhythm accurately and enable a user to keep tempo, feedback about the occurrence of each note has to be provided almost instantaneously. Musicians can tolerate a delay of 30ms in collaborative performances [18, 31]. Hence, the time between sensing the audio by the Sensor-Bit until the Display-Bit conveys the rhythm information to a user has to be below this threshold. The theoretical delay is composed of (1) analog-to-digital conversion (500k samples per second; $\approx$ 0.002 ms), (2) FFT processing ($\approx$ 1ms), (3) NRF communication ($>$ 700 packets per second; $\approx$ 1.42ms, (4) motor lag (ERM Motor,

---

[5] https://www.getedgegear.com/

[6] https://support.apple.com/kb/SP766?locale=en_GB

model 307-103 from precisionmicrodrives; lag time = 8ms, rise time = 28ms) = 30.422ms delay.

Volume Level	Light Delay (ms)	Vibration Delay (ms)
50%	M=3.26, SD=1.54	M=26.66, SD=4.93
100%	M=2.87, SD=1.67	M=33.98, SD=7.08

Table 1. Delay measurements for vibration and light of MuSS-Bits++.

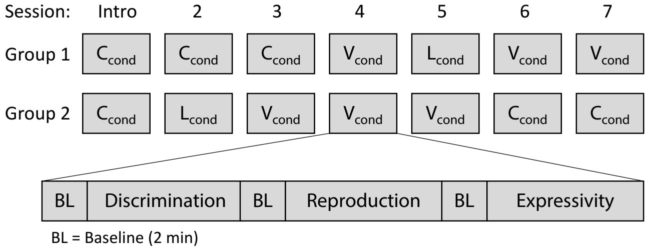
We measured the actual delay using a bass sound as stimulus, that was played from a computer (MacBook Air). We attached an accelerometer and a light dependent resistor (LDR) to the Display-Bit to measure vibration and light intensity. The computer was connected to the Sensor-Bit via the audio jack to eliminate environmental noise. The accelerometer and LDR, as well as the audio cable, were connected to an Arduino Mini, which measured the peak of the audio signal, as well as the peaks of the LDR and accelerometer values and calculated the time differences in microseconds. We conducted this experiment in two conditions: 50% and 100% of the stimulus' volume. In each condition, the stimulus was played 1000 times with a break of 1s (= 60bpm) between two stimuli. The results are shown in Table 1. As the light turns on almost immediately, the delay for the light condition roughly describes the conversion + FFT + NRF communication delay. As expected, the delay for vibration feedback increases with higher volume as the motor needs longer to speed up. Though vibration feedback sometimes exceeds 30ms for maximum volume, the delay is still around 30ms at an average volume.

Study 1: Rhythm Perception Study

Participants & Study Design

This study aimed to investigate how MuSS-Bits++ affects rhythm discrimination, reproduction, and expressiveness of deaf children. We recruited deaf participants from a deaf school who are interested in music. We faced the typical challenge of recruiting people in accessibility research [32] and were able to recruit only a small number of users that fit our criteria. Thus, we focused on extracting qualitative insights based on individuals (in the sense of [34]), rather than aiming for generalizability. The potential participants had to fill out a questionnaire comprising of five questions aiming on their musical interest. Based on our predefined inclusion criteria, we recruited 11 deaf children[7] (age: M=13.09, SD=2.02; six male, five female). Six participants had severe hearing loss (> 76db) and five participants had profound hearing loss (> 91db). The participants who answered the question 'Do you play an instrument?' from the questionnaire with 'yes' (P4, P6, P9) are considered as *musician participants* throughout the study as they are already familiar with music.

In the study, we tested three conditions: Control (C_{cond}), MuSS-Bits++ with light only (L_{cond}), MuSS-Bits++ with vibration only (V_{cond}). We excluded the combination of light+vibration (LV_{cond}) because we did not observe any difference between LV_{cond} and V_{cond} in a pilot test, and we wanted to keep the total duration of the study reasonably short for maximum attention. The whole study consisted of seven sessions per participant, each lasting about 30-45 minutes. Figure 2 gives an overview of the complete study design. The first session was conducted to familiarize the participant with the study procedure. Sessions 2-7 were block-wise counter-balanced arranged such that each condition was presented in two sessions. Throughout the whole study, we used a percussion instrument, the Thammattama drum (see Figure 1c), which was a familiar instrument for the children. Before we started the sessions, we obtained consent from the participants and their guardians, respectively. Furthermore, an interpreter was present to facilitate communication.



Session:	Intro	2	3	4	5	6	7
Group 1	C_{cond}	C_{cond}	C_{cond}	V_{cond}	L_{cond}	V_{cond}	V_{cond}
Group 2	C_{cond}	L_{cond}	V_{cond}	V_{cond}	V_{cond}	C_{cond}	C_{cond}

Figure 2. Overview of the study design of seven sessions.

Stimuli & Procedure

Each session consisted of three tasks: (1) discrimination task, (2) reproduction task, (3) expressivity task. For task 1 and 2, we took a rhythm set from Povel and Essens' work [27] with 105 different rhythms. Each beat had exactly eight onsets and was four beats long (with 16 notes). We recruited a musician who played these rhythms with the Thammattama drum. A close-up video recording showed the instrument and the musician's hands (see Figure 3c) for each rhythm. Since rhythm is time-sensitive, we discarded any rhythm recording where one or more played onset deviate more than 50ms compared to the ideal onset. The average deviation from the ideal onset across rhythms was 16.27 ms. Furthermore, the videos were post-processed so that each rhythm started exactly after 300ms from the start of the video.

The setup for the rhythm recording and the study was the same (see Figure 1c). We used a Shure SM57 X2U microphone to record the audio stream, a MacBook Pro to play the recorded videos and Create Labs speakers for sound output. In the L_{cond} and V_{cond} sessions, MuSS-Bits++ was attached to the palm (see Figure 3a). We tested other positions such as wrist and lower-arm in a pilot study, but the feedback from the participant suggests that the palm was a better location. This is in accordance with Goldstein [14] that shows that hands, particularly fingertips, have the best tactile resolution. However, feedback at the fingertips would interfere with holding the sticks, but the palm was a good compromise. Furthermore, in L_{cond}, we added a second Display-Bit in front of the computer screen (see Figure 3b), to minimize the distance between the area of action (video) and the visual feedback through MuSS-Bits++.

In the discrimination task, the computer showed two recordings of the rhythms with a break of 1s in between. The videos

---

[7] We initially recruited 13 participants but P3 and P5 could not finish all sessions due to sickness and were excluded from the analysis.

Figure 3. The setup and stimuli of the controlled user study: (a) MuSS-Bits was attached to the palm, (b) in light condition an additional Display-Bit was close to the screen, and (c) the participants could see the scores for the reproduction task.

were accompanied by sound (volume of speakers and computer set to 100%) to allow people with residual hearing to make use of it as they would do in real life. After that, the participants had to indicate whether the rhythms were the same or different. For same rhythm-pairs, we recorded two videos to avoid the possibility of children comparing the videos themselves rather than the rhythms presented in them. Furthermore, participants had the option to replay the rhythm as many times as they want. In each session, participants were shown 20 rhythm-pairs.

The reproduction task asked the participants to reproduce a given rhythm. For that purpose, the computer played one of the 105 rhythm recordings with sound as well as displayed the rhythm's score (see Figure 3c). Then the participant was asked to reproduce the presented rhythm with the Thammattama drum. We decided to show the rhythm's score on the screen because the participant in the pilot study struggled without having a visual anchor and this made the task too difficult. Participants had the option to replay the rhythm. In each session, participants were asked to reproduce 20 rhythms.

The expressivity task asked the participants to express bipolar themes. We chose themes from prior work [24]: (1) *happiness* and *sadness*, (2) a *rabbit* running over grass and a *turtle* creeping over the floor, (3) *bird* flying up and *rain* falling down. The themes were displayed in block-wise counter-balanced order.

Data Gathering and Analysis

For the discrimination test, we collected the responses (same/different) and the number of replays. We conducted Chi-Square tests to find differences between conditions based on correct responses and replays. Additionally, we conducted a Spearman correlation between accuracy and similarity of a rhythm-pair (calculated using the 'chronotonic distance' [38]) for each condition.

In the reproduction task, we collected replay instances and the onsets (with timestamp) the participant played. To ensure that the onsets were recorded and classified properly by our algorithm, an experimenter checked each rhythm and, where necessary, conducted adjustments. We compared the participants' onsets with the recorded onsets among three dimensions: (1) number of onsets (though all rhythms had eight onsets, some participants played less or more onsets), (2) average deviation of onsets (only for the rhythms where the participants played eight onsets), and (3) deviation in tempo. As these data turned out to be non-parametric, we conducted Kruskal-Wallis tests for each participant between conditions on each metric. As post-hoc tests, we conducted Mann-Whitney-Wilcoxon tests on condition pairs. Furthermore, we ran a Spearman correlation between complexity (calculated using the 'metricalonormpk-MUS-44' algorithm [37]) and each of the metrics for each condition.

As statistical methods are limited on small sample sizes, these methods have been complemented with qualitative data. Thus, all tasks were video recorded, and an experimenter took notes during the session. Furthermore, an independent researcher qualitatively analyzed the video-recordings of task 3 for differences between conditions and themes. Additionally, we asked the participants to fill out the SUS questionnaire [3] after the study to receive feedback about the usability of MuSS-Bits++.

Orthogonal to the tasks we also collected heart rate (HR) data from the participants using a TomTom Touch[8]. HR has been found to rise in situations of stress [36]. We were interested whether there is a difference in HR between conditions, which might suggest different stress levels. We included a two minutes baseline before each task in which the child was supposed to sit or stand, depending on the task to match with the physical and postural requirement of the task. We normalized the HR ($\mathrm{HR}_{\mathrm{norm}}$) data of a task by subtracting the mean HR of the baseline to compensate for day-dependent differences in HR. We conducted Kruskal-Wallis tests for each participant to find differences between conditions (the HR data turned out to be non-parametric as all Levene's tests were significant). For each significant difference, we conducted Mann-Whitney-Wilcoxon tests on each condition pair (as suggested in [9]) to find out which pair was significantly different and calculated the effect size. We were only interested in significant medium and large effects.

Results

For the discrimination task (see Figure 4), the Chi-Square test did not reveal any significant difference on correct replies, though P9 improved from 57.5% in C_{cond} to 85% in L_{cond} and 90% in V_{cond}. Furthermore, there was no significant correlation between accuracy of participants and similarity of rhythm-pairs. A significant difference between conditions could be found on replay frequencies for P2 (χ^2= 10.69, with 2 df, $p < 0.01$) and P4 ($\chi^2 = 15.2$, with 2 df, $p < 0.001$). P4 had a higher replay frequency in C_{cond} (20) compared to L_{cond} (4) and V_{cond} (6). P2 needed more replays for V_{cond} (14) and L_{cond} (11) than for C_{cond} (1). As expected, there was a significant difference

---

[8] https://www.tomtom.com/en_us/sports/fitness-trackers/fitness-tracker-touch/black-large/

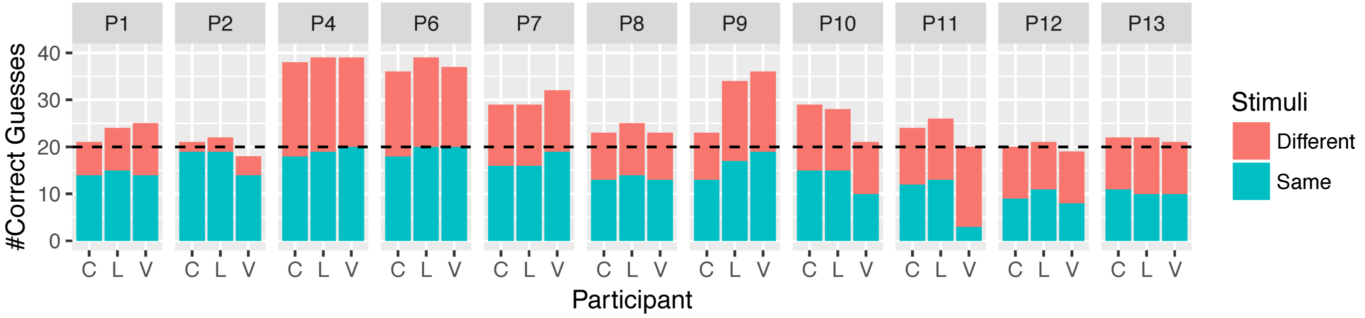


Figure 4. Task 1: Overview of the participant's performance of the discrimination task. Conditions are marked with C (Control), L (Light), and V (Vibration).

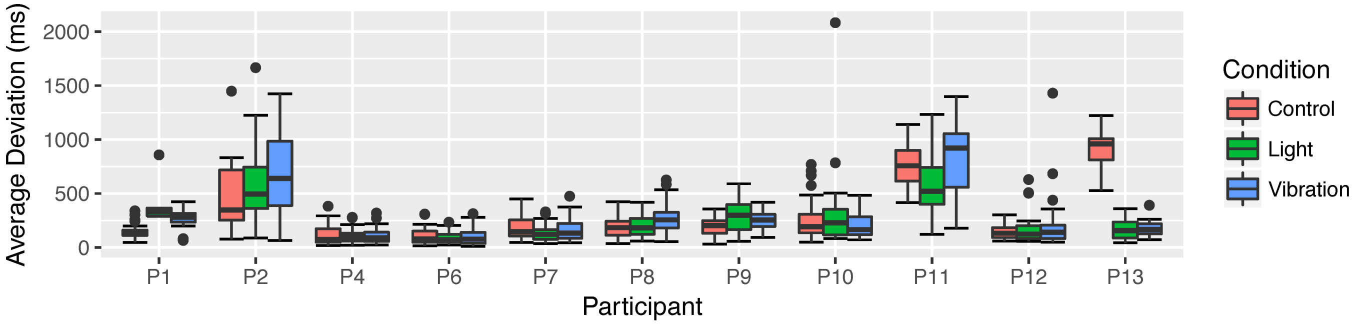


Figure 5. Task 2: Overview of the average deviation of onsets (only rhythms where the participants played eight onsets).

in accuracy between musician and non-musician participants ($t(3.01) = 4.61$, $p < 0.05$, $r = 0.94$). The Kruskal-Wallis tests revealed no significant differences among conditions for tempo variations. A similar effect was observed for onsets except for P13, who deviated the most in C_{cond} (Md = 11) from eight beats. Even in L_{cond} and V_{cond} his median was 9 but was significant lower compared to C_{cond} ($H(2) = 34.72$, $p < 0.001$; C_{cond}–L_{cond}: $U = 1332.5$, $p < 0.001$, $r = -0.58$; C_{cond}–V_{cond}: $U = 1274$, $p < 0.001$, $r = -0.52$). As shown in Figure 5, P13's average deviation from the given rhythm significantly reduced from C_{cond} (Md = 960ms) to L_{cond} (Md = 156ms) as well as V_{cond} (Md = 166ms) ($H(2) = 13.7$, $p < 0.01$; C_{cond}–L_{cond}: $U = 84$, $p < 0.001$, $r = -0.91$; C_{cond}-V_{cond}: $U = 48$, $p < 0.001$, $r = -0.91$). For P1, this was reversed as he performed worse in L_{cond} (Md = 335ms) compared to C_{cond} (Md = 128ms) ($H(2) = 15.22$, $p < 0.001$; $U = 5$, $p < 0.001$, $r = -0.67$). Apart from that, there were various medium effects between conditions, though they do not have a clear tendency towards one or the other condition to reduce average deviation. Furthermore, we did not find any correlation between complexity and one of the metrics. For P6 there was a significant difference in replaying the rhythm video. He replayed it more often in L_{cond} (21 times) compared to V_{cond} (4 times). Splitting the participants into musicians and non-musicians revealed significant difference among all conditions for the metric average deviation resulting in medium effects ($U = 156740$, $p < 0.001$, $r = -0.40$). The other two metrics were either nonsignificant or just revealed small effects.

In the expressivity task, we saw that the theme pair *happiness* and *sadness*, as well as *rabbit* and *turtle*, have been mainly expressed through fast and slow tempo. This is the same observation that Petry et al. [24] made in their observational study. For *bird* and *rain*, there was no specific difference in tempo or volume. A further observation was that most participants converged to a more regular beat and controlled their volume at one level more often when using MuSS-Bits++ (in L_{cond} but even more often in V_{cond}). Apart from that, some participants intentionally varied volume and tempo between conditions. P1 for example, lowered his volume for both *turtle* and *bird* themes when using MuSS-Bits++ compared to C_{cond}. Furthermore, P7 slowed down the tempo in L_{cond} and V_{cond} compared to C_{cond} for the themes *rabbit*, *sadness*, and *turtle*.

Across all participants, the average SUS score was 80.91 (SD=11.69). This score is above the threshold of 68 [3] and indicates MuSS-Bits++ to be a usable system. Looking more detailed at the questions of the SUS we found that two questions were very positively rated across all participants: 'I think that I would like to use this system frequently' and 'I felt very confident using the system.' In contrast, participants agreed with the statement 'I think that I would need the support of a technical person to be able to use this system.' This could be because we always helped them to set up MuSS-Bits++, which could have let them believe they needed technical support.

In task 1, for six participants there was a clear order in which condition the HR_{norm} was lowest. Three participants had a significant lower HR_{norm} in V_{cond} compared to L_{cond} and C_{cond}, two for L_{cond} and one for C_{cond}. In task 2, only one person per condition had the lowest HR_{norm}. Particularly P13 reduced his HR_{norm} from C_{cond} to V_{cond} at 12.1bpm and from L_{cond} to V_{cond} at 9.4bpm. He also performed better in V_{cond} compared to C_{cond}, as such, it seems that vibration affects his performance as well the associated stress with the task. In task 3, also his

Figure 6. Music lesson with MuSS-Bits++: (a) Music teacher shows rhythm to participant, (b) participant reproduces the rhythm, (c) collaborative synchronization session.

HR_{norm} was much lower for V_{cond} compared to the other two conditions.

Study 2: MuSS-Bits++ in a Music Lesson

This study aimed to understand how MuSS-Bits++ performed in a real-world setting. A music teacher selected three participants from the previous study (P4, P6, P9) and conducted three music lessons with them in a group (see Figure 6). Each lesson was about 45 minutes, and the music teacher could decide what to teach with two sets of MuSS-Bits++. We suggested that he conducts half of the last lesson without MuSS-Bits++ to observe how things would change without MuSS-Bits++. We conducted a semi-structured pre- and post interview with the music teacher as well as collected feedback from the participants after the last lesson. Furthermore, a researcher video recorded the lessons and took notes during the study.

Findings

Before the study, the teacher suggested three positions where he could imagine that MuSS-Bits++ would have a good effect: (1) the forehead for vibration, (2) the hand for the light feedback, since they already have a good connection with their hands due to sign language, and (3) on the table for the light feedback. During the study, we observed that one of the Display-Bits was attached to one participant and the other was put on the table to emit light feedback.

Already after the first lesson, the music teacher said that it was easier for him to teach rhythm to the participants compared to the past. According to him, the children quickly understood what tempo he wanted them to play, and they were able to keep the tempo better than before: *"Previously it was difficult for them to keep tempo. However, with [MuSS-Bits++] they were able to keep tempo."* The participants agreed and said that they better understood the instructions during the lessons. In the second lesson, the music teacher played the Thammattama drum, and we observed that P6 was able to adjust to the music teacher's tempo while wearing MuSS-Bits++. Furthermore, in a synchronization part of the last lesson, we observed that the participants were able to quickly notice a change in tempo or rhythm when MuSS-Bits++ was present. Moreover, P9 and P6 had problems to keep the tempo and loudness without MuSS-Bits++ and maintained the tempo better with MuSS-Bits++. The music teacher commented this: *"Synchronization task was successful. That was something unexpected. They composed nice music."*

P9 seems to benefit from light feedback. She said that she was looking at MuSS-Bits++ to keep her tempo. However, she preferred the Display-Bit to be on the table rather the hand for the light feedback. Furthermore, all participants emphasized the use of vibration. P4's usual feedback strategy is to put his hand on the table, while the teacher plays. However, MuSS-Bits++'s vibration provided him with a more powerful cue to perceive feedback, and he could *"feel more music."* The overall consensus across the participant was that participants enjoyed the lessons more than before. The teacher observed this as well: *"Response was really good. They were enthusiastic to use it. I would like to use it with other students as well."*

The music teacher also tried to do a jamming session in the second lesson. He planned to play a rhythm using a guitar and expected the others to follow him with the Thammattama drum (one-by-one). However, this did not work well: *"Guitar session was not successful. I think MuSS-Bits++ didn't work for the frequency range of guitars. They didn't feel the sound of strings."* Furthermore, from the video recordings, we observed that MuSS-Bits++ did not work for instruments such as tambourine and guitar. This could be expected as the implemented rhythm extraction algorithm focuses on lower frequencies and is optimized for the Thammattama drum. The music teacher suggested that MuSS-Bits++ could be used for dancing as well, as it provides a good feedback of the tempo. In addition, he proposed that MuSS-Bits++ could be equipped with a recording functionality: *"If [Muss-Bits++] can record the vibrations [when I am playing] it should be useful. Then it's easy for them to learn and match beats again while they are playing."*

DISCUSSION

Insights

In this section, we combine and discuss the results of the controlled study and the music lessons. We hope that these insights will inform other researchers about the design of music-sensory-substitution systems for music-making.

User Experience vs. Performance: Six participants explicitly mentioned that they like the vibrations; out of that three went on to indicate that it made the task easier (P4, P6, P9). However, their performance did not improve significantly in any of those tasks, and sometimes their heart rate increased significantly for L_{cond} or V_{cond} which may indicate positive or also negative stress. In contrast, two participants (P10 and P12) mentioned that they had problems with feeling the vibrations, but there was no significant difference in their performance in any of the tasks. This raises the question, why do most of them appreciate vibrations and even perceived the tasks easier, while there was no change in their performance? If vibration does not improve their ability to play rhythms, it would be dubious whether this should be explored further.

We can think of the following factors, why the participants did not improve: (1) lack of skill to play the Thammattama drum, (2) it takes time to learn to use MuSS-Bits++'s feedback, (3) they have a different concept of music compared to hearing people and thus needed another mapping to improve, (4) the tasks were too difficult or too easy for them, so that no effect was quantitatively visible. Most of the participants did not have experience with the Thammattama drum which could affect their performance in tasks. However, three of them have played the instrument before, and all of them said vibration makes the task easier, though their performance did not improve. Thus, we assume that the performance improvement does not depend on the participant's instrument skill. However, participants may need time to be able to use the feedback they receive from MuSS-Bits++. Kristjánsson et al. [17] stated that sensory-substitution systems always have a learning curve, where steepness depends on the mapping. Learning to use a sensory-substitution device's feedback can happen almost instantly as in [39], but can also take several weeks or months. Also, it could be that the concept of music and rhythm, in particular, is very different for deaf people than for hearing. Thus, our mapping might not provide the information that they need to improve their performance. It would be interesting to investigate different mappings and enable deaf users to create their personal ones as we discuss in the future work section. We selected rhythms with various complexities to avoid having only too simple or too difficult rhythms. In the pilot study, we found a significant difference between C_{cond} and V_{cond} in performance and thus did not look into selecting a subset of rhythms with a certain complexity. Furthermore, we did not find a correlation between performance and complexity. Hence, we assume participants did not improve with MuSS-Bits++ due to factor (2) or (3).

Rhythm Communication: The controlled experiment did not reveal a significant effect on MuSS-Bits++ on performance. However, based on the qualitative feedback, especially from the music lesson, we believe, MuSS-Bits++ still can be a good tool to communicate rhythm. In the music lessons, the teacher observed that children understood and adopted rhythm information (tempo and intensity) faster when he conveyed them via MuSS-Bits++: *"This device helped them to control intensity of the sound. Without [Muss-Bits++] I have seen them struggle to keep their playing intensity and tempo."* Also, the participants felt that they understood the instructions better than before. Hence, MuSS-Bits++ can be valuable for music educators in the deaf community. Especially developing countries with not specialized music teachers, could make use of MuSS-Bits++ as a musical communication tool.

Vibrotactile vs. Visual Feedback: The participants with musical background preferred the vibration feedback from MuSS-Bits++. P4 pointed out that the vibration feedback complements the natural vibrotactile feedback from instruments as MuSS-Bits++ provides a stronger cue. P13, the youngest participant, benefited from vibration feedback as his performance improved significantly with vibration feedback compared to C_{cond}. P6 seems to be particularly sensitive to vibrations as well as his heart rate was always comparatively lower for V_{cond} than for C_{cond} or L_{cond}. However, similar to what has been suggested in other assistive technology work [34], given the diversity of skills and hearing abilities, a vibration feedback that works across everyone may not even exist.

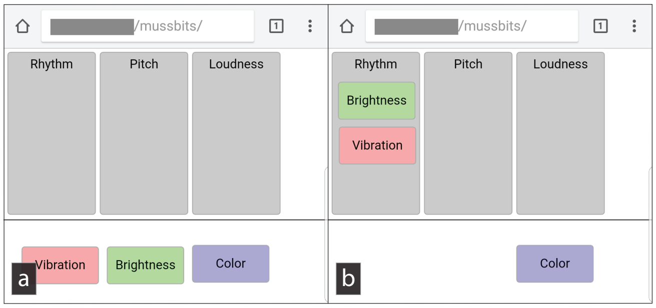


Figure 7. Our user interface to customize mappings of the Sensor-Bit: (a) no mapping selected, (b) rhythm is mapped to vibration intensity and visual brightness.

Vibrotactile feedback is a discrete private feedback that only the wearer perceives. In some situations, this might be desirable, whereas light feedback attracts attention and acts as a public display. This, in fact, could be useful in collaborative music sessions as it communicates the beat of the user to other musicians. In the music lessons, this was particularly important when the music teacher demonstrated rhythms on the Thammattama drum. Besides focusing on the drum and the teacher's movements, the participants very often looked at the Display-Bit's visual feedback to understand the music teacher's instructions. Apart from these observations, the participants rarely commented about light feedback. Vibrotactile feedback seems to be the more enjoyable and, according to the musician participants, more communicative modality. This is in accordance with Nanayakkara et al. [21] who found that vibration has a higher impact on the user's experience than visual feedback.

Limitations & Future Work

Wireless Communication: The current MuSS-Bits++ prototype has some limitations. As we use NRF instead of WiFi, the range of MuSS-Bits++ less (10m with in-line-of-sight). With objects in between, the communication distance drastically drops. To avoid users wondering what happens with MuSS-Bits++, we implemented a visual feedback mechanism (pulsing lights in different colors) that turns on when no packet has been received after 300ms.

Potential Confounding Effects: Our findings might have been confounded by the cultural bias of the local community where the study took place. Furthermore, there could have been novelty effect as the study was conducted over a few weeks. In fact, we have started a long-term study at the deaf school where 12 deaf children are expected to use MuSS-Bits++ in music lessons over a period of 6 - 9 months.

User Defined Mappings: MuSS-Bits++ is conceptualized as a building block that is not only restricted to rhythm. The FFT information can also be used to extract other musical elements, such as pitch or timbre. Furthermore, new mappings can be implemented and tested. In prior work, researchers fixed the mapping. However, each deaf person may benefit from a different mapping. Hence, MuSS-Bits++ can be used as a tool that enables deaf users to create their personal mappings. The challenge in creating the mapping is to design an interface that lets users customize the connection between musical elements as input and the visual/vibrotactile output. We started looking into this and built a working user interface that communicates with the Sensor-Bit via the audio jack (see Figure 7). In the design of the user interface, we first defined musical elements (e.g., rhythm, pitch, loudness) and output classes (e.g., vibration intensity, visual-color, visual-brightness). Furthermore, we implemented mappings that take any musical element as input, but only belong to exactly one output class. Through drag-and-drop, the mapping can be dragged to any musical element to establish a connection, whereas only one mapping of each output class can be active at a time. In future, we plan to test this, starting with deaf musicians.

Instrument Specific Music Element Extraction: In the music lessons we observed that guitar and some other instruments did not work well with MuSS-Bits++. This is due to different characteristic sound profiles of instruments. Implementing instrument specific music element extraction algorithms could improve the perception of musical elements on instruments. This would require a way to select an instrument which could be established either through a user interface or automatically by analyzing the sound's harmonics.

CONCLUSION

In this paper, we investigated how MuSS-Bits++, a music-sensory-substitution systems for music-making, affects rhythm perception. We first interviewed a music teacher from a deaf school and identified that communicating rhythm information is a challenging task. In a controlled study, we found that participants with musical background perceived MuSS-Bits++'s vibrations as making the task easier, though no significant difference in their performance was observed. We argued that this could be due to two reasons: (1) they need time to learn how to use the feedback from MuSS-Bits++, or (2) the mapping we provided is not matching the mapping they need to improve their accuracy. Furthermore, in music lessons that were conducted by a music teacher from a deaf school, the teacher observed, and the participants agreed, that it was easier for the participants to understand the music teacher's instructions. Furthermore, collaborative synchronization activities among three participants with the music teacher were observed to work better with MuSS-Bits++ compared to the same activity without MuSS-Bits++. We believe that music-sensory-substitution systems hold a lot of potential for music-making and hope that this work will help other researchers to develop our vision further.

ACKNOWLEDGMENTS

This work is partially funded by the SUTD President's Graduate Fellowship. The authors gratefully acknowledge the support of Dr. Reijntjes School for the Deaf in Sri Lanka. In addition, the authors appreciate the feedback and time of all participants and the music teacher who made this work possible. Moreover, we would like to thank Attila Victor Achenbach for the thoughtful industrial design and Juan Pablo Forero for the ingenious technical realization of the MuSS-Bits++'s prototype.

REFERENCES

1. 2017. Music and the Deaf. (March 2017). Retrieved March 6, 2017 from `http://matd.org.uk/`.
2. Caterina Bertini, Fabrizio Leo, Alessio Avenanti, and Elisabetta Ladavas. 2010. Independent mechanisms for ventriloquism and multisensory integration as revealed by theta-burst stimulation. *European Journal of Neuroscience* 31, 10 (2010), 1791–1799.
3. John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
4. Marshal Chasin. 2003. Music and hearing aids. *The Hearing Journal* 56, 7 (July 2003), 36. `http://dx.doi.org/10.1097/01.HJ.0000292553.60032.c2`
5. Summer Crider. 2009. *Re-Defining Music Through Deaf Lens*. Master's thesis. Gallaudet University, Washington, DC.
6. Alice-Ann Darrow. 1993. The Role of Music in Deaf Culture: Implications for Music Educators. *Journal of Research in Music Education* 41, 2 (July 1993), 93–110. `http://dx.doi.org/10.2307/3345402`
7. Ward R. Drennan and Jay T. Rubinstein. 2008. Music perception in cochlear implant users and its relationship with psychophysical capabilities. *Journal of rehabilitation research and development* 45, 5 (2008), 779–789.
8. Wiliam G Fawkes. 2006. *The Teaching of Music to Hearing Impaired Children and Teenagers.* `http://www.maryharehistory.org.uk/articles/fawkes/fawkes_2006.pdf`
9. Andy Field and Graham Hole. 2002. *How to Design and Report Experiments*. SAGE. Google-Books-ID: LN6QAwAAQBAJ.
10. Sean Forbes. 2017. Sean Forbes. (2017). Retrieved March 6, 2017 from `http://www.deafandloud.com/`.
11. David Fourney. 2015. *Making The Invisible Visible: Visualization Of Music And Lyrics For Deaf And Hard Of Hearing Audiences*. Ph.D. Dissertation. Ryerson University, Toronto, Ontario, Canada.
12. John J. Galvin, Qian-Jie Fu, and Robert V. Shannon. 2009. Melodic Contour Identification and Music Perception by Cochlear Implant Users. *Annals of the New York Academy of Sciences* 1169, 1 (July 2009), 518–533. `http://dx.doi.org/10.1111/j.1749-6632.2009.04551.x`
13. Evelyn Glennie. 2003. How to truly listen. (2003). `https://www.ted.com/talks/evelyn_glennie_shows_how_to_listen`
14. E Goldstein. 2009. *Sensation and perception*. Cengage Learning.
15. Maria Karam, Carmen Branje, Gabe Nespoli, Norma Thompson, Frank A. Russo, and Deborah I. Fels. 2010. The Emoti-chair: An Interactive Tactile Music Exhibit. In *Proc. CHI EA 2010*. ACM, New York, NY, USA, 3069–3074. `http://dx.doi.org/10.1145/1753846.1753919`

16. M. Karam, F.A. Russo, and D.I. Fels. 2009. Designing the Model Human Cochlea: An Ambient Crossmodal Audio-Tactile Display. *IEEE Transactions on Haptics* 2, 3 (July 2009), 160–169.
17. Árni Kristjánsson, Alin Moldoveanu, Ómar I. Jóhannesson, Oana Balan, Simone Spagnol, Vigdís Vala Valgeirsdóttir, and Rúnar Unnthorsson. 2016. Designing sensory-substitution devices: Principles, pitfalls and potential. *Restorative Neurology and Neuroscience* 34, 5 (2016), 769–787. `http://dx.doi.org/10.3233/RNN-160647`
18. Zefir Kurtisi, Xiaoyuan Gu, and Lars Wolf. 2006. Enabling Network-centric Music Performance in Wide-area Networks. *Commun. ACM* 49, 11 (Nov. 2006), 52–54. `http://dx.doi.org/10.1145/1167838.1167862`
19. Daniel J Levitin. 2011. *This is your brain on music: Understanding a human obsession*. Atlantic Books Ltd.
20. Charles Limb. 2011. Building the musical muscle. (2011). Retrieved November 11, 2014 from `http://www.ted.com/talks/charles_limb_building_the_musical_muscle`.
21. Suranga Nanayakkara, Elizabeth Taylor, Lonce Wyse, and S H. Ong. 2009. An Enhanced Musical Experience for the Deaf: Design and Evaluation of a Music Display and a Haptic Chair. In *Proc. CHI 2009*. ACM, New York, NY, USA, 337–346. `http://dx.doi.org/10.1145/1518701.1518756`
22. Carol A. Padden and Tom L. Humphries. 1990. *Deaf in America: Voices from a Culture*. Harvard University Press, Cambridge, Mass.
23. Russ Palmer. 1994. Tac-Tile Sound System. (1994). `http://www.russpalmer.com/tactile.html`
24. Benjamin Petry, Jochen Huber, and Suranga Nanayakkara. 2018. Scaffolding the Music Listening and Music Making Experience for the Deaf. In *Assistive Augmentation*. Springer, Singapore, 23–48. `https://link.springer.com/chapter/10.1007/978-981-10-6404-3_3` DOI: 10.1007/978-981-10-6404-3_3.
25. Benjamin Petry, Thavishi Illandara, Juan Pablo Forero, and Suranga Nanayakkara. 2016b. Ad-Hoc Access to Musical Sound for Deaf Individuals. In *Proc. ASSETS 2016*. ACM, New York, NY, USA, 285–286. `http://dx.doi.org/10.1145/2982142.2982213`
26. Benjamin Petry, Thavishi Illandara, and Suranga Nanayakkara. 2016a. MuSS-Bits: Sensor-Display Blocks for Deaf People to Explore Musical Sounds. In *Proc. OzCHI 2016*. ACM, New York, NY, USA, 72–80. `http://dx.doi.org/10.1145/3010915.3010939`
27. Dirk-Jan Povel and Peter Essens. 1985. Perception of Temporal Patterns. *Music Perception: An Interdisciplinary Journal* 2, 4 (July 1985), 411–440. `http://mp.ucpress.edu/content/2/4/411`
28. Janine Roebuck. 2007. I am a deaf opera singer. (Sept. 2007). Retrieved August 26, 2015 from `https://www.theguardian.com/theguardian/2007/sep/29/weekend7.weekend2`.
29. Gottfried Schlaug. 2015. Musicians and music making as a model for the study of brain plasticity. In *Progress in Brain Research*. Music, Neurology, and Neuroscience: Evolution, the Musical Brain, Medical Conditions, and Therapies, Vol. 217. Elsevier, 37–55.
30. Gottfried Schlaug, Eckart Altenmüller, and Michael Thaut. 2010. Music listening and music making in the treatment of neurological disorders and impairments. *Music Perception: An Interdisciplinary Journal* 27, 4 (2010), 249–250.
31. Nathan Schuett. 2002. The effects of latency on ensemble performance. *Bachelor Thesis, CCRMA Department of Music, Stanford University* (2002).
32. Andrew Sears and Vicki L. Hanson. 2012. Representing Users in Accessibility Research. *ACM Trans. Access. Comput.* 4, 2 (March 2012), 7:1–7:6. `http://dx.doi.org/10.1145/2141943.2141945`
33. D Shibata. 2001. Brains of deaf people hear music. *International Arts-Medicine Association Newsletter* 16 (2001), 4.
34. Kristen Shinohara and Josh Tenenberg. 2009. A Blind Person's Interactions with Technology. *Commun. ACM* 52, 8 (Aug. 2009), 58–66. `DOI:` `http://dx.doi.org/10.1145/1536616.1536636`
35. Carl Stumpf. 1883. Tonpsychologie. *Leipzig: Hirzel* 1 (1883).
36. Joachim Taelman, S. Vandeput, A. Spaepen, and S. Van Huffel. 2009. Influence of Mental Stress on Heart Rate and Heart Rate Variability. In *4th European Conference of the International Federation for Medical and Biological Engineering*. Springer, Berlin, Heidelberg, 1366–1369. 10.1007/978-3-540-89208-3_324.
37. Eric Thul. 2008. *Measuring the complexity of musical rhythm*. Ph.D. Dissertation. McGill University.
38. Godfried T Toussaint. 2004. A Comparison of Rhythmic Similarity Measures. In *Proc. ISMIR 2001*.
39. Mitchell Tyler, Yuri Danilov, and Paul Bach-Y-Rita. 2003. Closing an open-loop control system: vestibular substitution through the tongue. *Journal of Integrative Neuroscience* 02, 02 (Dec. 2003), 159–164. `http://www.worldscientific.com/doi/abs/10.1142/S0219635203000263`
40. Ronald T Verillo. 1991. Vibration sensing in humans. *Music Perception* 9, 3 (1991), 281–302.
41. Georg Von Bekesy. 1959. Similarities between hearing and skin sensations. *Psychological Review* 66, 1 (1959), 1–22. `http://dx.doi.org/10.1037/h0046967`
42. Yinsheng Zhou, Khe Chai Sim, Patsy Tan, and Ye Wang. 2012. MOGAT: Mobile Games with Auditory Training for Children with Cochlear Implants. In *Proc. MM 2012*. ACM, New York, NY, USA, 429–438. `http://dx.doi.org/10.1145/2393347.2393409`