論文 / 著書情報
Article / Book Information

Title	Selective listening attention enhancement, using a simultaneous visual and haptic stimuli
Authors	Alfonso Balandra, Hironori Mitake, Takako Yoshida, Shoichi Hasegawa
Citation	Journal of New Music Research, Volume 48, Issue 3, pp. 294-308
Pub. date	2019, 3
DOI	http://dx.doi.org/10.1080/09298215.2019.1575424
Note	This article has been accepted for publication in Journal of New Music Research, published by Taylor & Francis.

Selective listening attention enhancement, using a simultaneous visual and haptic stimuli

Alfonso Balandra[a], Hironori Mitake[a], Takako Yoshida[b] and Shoichi Hasegawa[a]

[a]Tokyo Institute of Technology, Computer Intelligence and Systems Science, Suzukakedai Campus, Kanagawa, Japan; [b]Tokyo Institute of Technology, Department of Mechanical Engineering, Ookayama Campus, Tokyo, Japan

ARTICLE HISTORY
Compiled January 16, 2019

ABSTRACT
In this paper, we explore the possibility of using a simultaneous visual and haptic stimulus to increase the selective listening attention of humans. Using a real-time self-reported evaluation method, the listening attention of several participants was measured. Specifically, the participants listened to a classical fugue while receiving a resemblant non-auditory stimulus synchronized with the notes of a single instrument in the song. Thereafter, the participants ($n = 30$) were asked to press down on a button if they were able to focus their listening attention on the instrument highlighted by the non-audio stimulus, while their initial detection time and total focus time were measured. Three combinations of two different stimuli modalities were compared: visual, haptic, and the combination of visual and haptic, using three classical polyphonic fugues. The empirical experiment results indicate that, regardless of a participants music skills or the voice pitch, the participants performance was improved by the visual-haptic stimuli, with longer selective listening periods and faster detection times compared with the single modality stimuli. At the end of the experiment, a subjective questionnaire was applied to measure the subjective participants ease between the stimulus conditions. The questionnaire results indicated that the participants preferred to use the visual-haptic stimulus, compared to the visual or haptic stimuli alone.

KEYWORDS
Multimodal perception; audio-tactile-visual stimulus; haptic music; multimodal stimulus

1. Introduction

An individuals selective listening attention clearly plays an important role in music appreciation and education. It is a key factor related to music education because it is essential to be able to distinguish musical attributes (Madsen & Geringer, 2000). Moreover, it factors into an exciting music listening experience (Goldstein, 1980; Madsen, 1997). Since, in most situations music listening is a multi-senstive experience, that may involve simultaneous vibratory and visual cues, like a live concert, then it seem simple and natural to use a non-auditive cue in order to help the listener to focus his/her attention into an specific instrument or voice in the music.

CONTACT A. Balandra. Email: balandra.a.aa@m.titech.ac.jp

Therefore, a synchronized external stimulus could be used to obtain listeners attention and focus it on an specific instrument. To avoid auditory distraction, this stimulus should be haptic, visual, or a combination of both. It seems logical that a simultaneous audio and visual stimulus may produce a better listening performance compared to a single modality stimulus, but if the Colavita visual dominance effect Spence (2009) is considered, then it seems necessary to evaluate whether a synchronized visual-haptic stimulus could in fact provide some perceptual advantage compared to a visual or haptic stimulus alone.

It is also important to determine whether the multidimensionality of music has an impact on participants selective listening. Therefore, factors such as the music structure, timbre, and the pitch differences or similarities between instruments should also be considered.

To the best of our knowledge, the impact of a multimodal stimulus on the music listening performance has not yet been directly evaluated. Therefore, we propose an empirical experiment to investigate whether the combination of visual and haptic stimuli is in fact more effective for highlighting a specific instrument compared to an individual haptic or visual stimulus. The results of this experiment may be directly applied to the effective design of new multi-modal interfaces based on our multimodal perception capabilities.

Our experimental approach is based on previous psychophysical real-time self-reported proposals (Madsen, Brittin, & Capperella-Sheldon, 1993) and prior results in experimental psychology, which regarded participants detection time as a consequence of a mental process (Sternberg, 1969).

The proposed evaluation method uses a synchronized non-audio stimulus to highlight a single instrument in a song, which will be refer as: *highlighted instrument* or

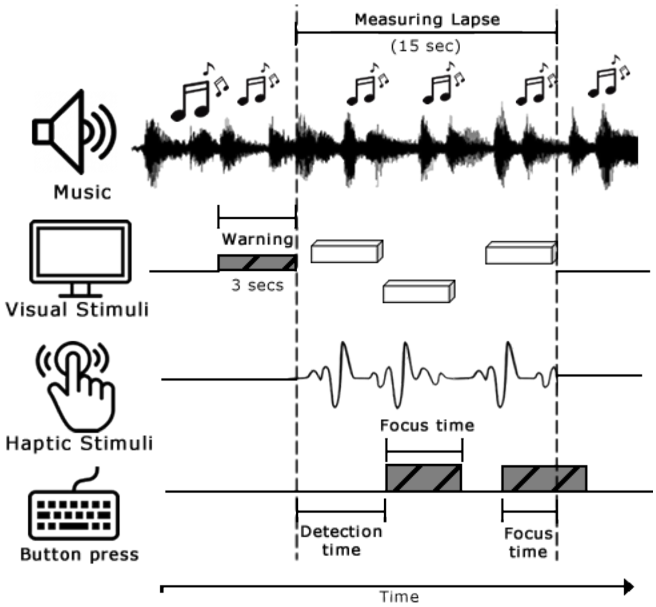


Figure 1. This figure shows a time diagram of a *measuring lapse*, which is the short period of time where the stimuli was displayed and participants' focus and detection time were measured. While listening to the music, each participant received a visual warning of 3 seconds before the non-audio stimulus was displayed. Then, the participant had 15 seconds to identify the instrument and press/or/release a key to report if he/she was able to find and follow the indicated instrument with the given stimulus.

highlighted voice. Three conditions of the non-audio stimulus were evaluated: visual-only, haptic-only, and a combination of visual and haptic stimuli. The haptic and visual stimuli were designed to represent the rhythm, pitch, and duration of the notes played by a single instrument in a polyphonic song. The visual-haptic stimulus was defined just by the synchronized combination of the visual and haptic stimuli. Thereafter, each stimulus condition was used to highlight a specific instrument for only a short period of time, which will be refer as: *measuring lapse*. The measuring lapses were predefined in diverse parts of the song. And during each measuring lapse, the participants reported their listening attention to the highlighted instrument by pressing down or releasing a button. In the proposed experiment, the participants' initial detection time and the total time the report button was pressed down were quantified. So, quicker initial detections and longer presses indicated better selective listening (see Figures 1 & 2).

2. Related Research & Motivation

The use of synchronized haptic and/or visual signals for music listening is relatively new. Many proposals have used a synchronized non-auditory stimulus to enhance individuals music listening experience, such as Yamazaki's study (Yamazaki, Mitake, & Hasegawa, 2016)in which a specific hardware solution was created to synchronize a haptic stimulus with the music to enhance listening experience. Another example

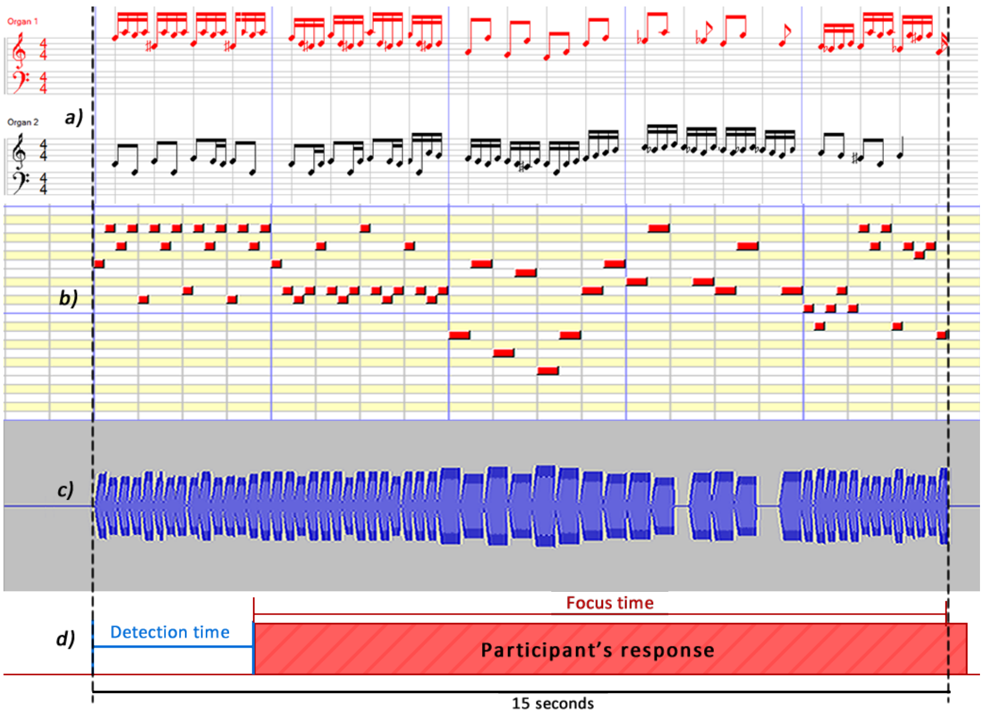


Figure 2. The figure shows the signals used in a measuring lapse with two voices. *a)* Shows the music score from the first measuring lapse of Bach's fugue (BWV 578), with the highlighted voice in red. *b)* Shows the visual stimuli of each note played by the highlighted voice. *c)* Shows the haptic stimuli used to represent the highlighted instrument's pitch and timing. *d)* Shows the response of a specific subject to the given stimuli, appeared in a red rectangle, and his/her detection time appeared in blue.

is Hwang's proposal (Hwang, Seo, Kim, & Choi, 2013), in which a psychophysical experiment was performed to measure the magnitude function of a vibratory stimulus at diverse frequencies. Subsequently, Hwang proposed (Hwang & Choi, 2014) a real-time auditory saliency method to extract the amplitude, loudness, and energy of the audio signals. By using these properties, a haptic signal was created by applying the previously mentioned magnitude functions. Hwang's method was based on the idea of engaging the users listening attention by emphasizing the salient parts of music with a vibratory stimulus, thus leading to an enhanced music listening experience.

In addition to haptics, other studies defined visual-tactile music listening environments for individuals with a hearing impairment (Karam, Russo, & Fels, 2009; Nanayakkara, Taylor, Wyse, & Ong, 2009). More recently, Tanabe used a piano with special vibrotactile actuators attached to each key to provide a synthetic tactile sensation, while a visual stimulus was displayed over the piano keys (Tanabe et al., 2016). In general, the main purpose of the visual and haptic stimuli in the previous studies was to provide more information to improve the users' listening or playing performance.

In addition to these examples, there are other studies that synchronized music with a haptic stimulus (Armitage & Ng, 2016; Merchel, Altinsoy, & Stamm, 2010), a visual stimulus (Meuer et al., 2017), or a combination of both (Rodet et al., 2005; Tanabe et al., 2016). Merchel et al. and Armitage and Ng designed specific hardware solutions to improve participants playing performance. However, they only Merchel et al. considered previous research on audio and tactile human perception in their design. The approaches of Meuer et al. and Rodet et al. were more open because both defined their own integration methods for the audio, visual, and tactile stimuli, though neither gave any specific consideration to audio or tactile perception.

Unfortunately, most of the previously mentioned studies did not consider any previous multimodal perceptual analyses in their design. To the best of our knowledge, the only exception is Hwang, who performed a psychophysical experiment (Hwang et al., 2013) in order to improve his own multimodal music listening environment (Hwang & Choi, 2014).

Contrary to our hypothesis, the Colavita effect (Spence, 2009) indicates that humans are more likely to respond to a visual stimulus when they are exposed to a combination of auditory and visual signals. However, Hecht and Reiner found that the visual dominance disappears when simultaneous visual-audio-tactile stimuli are presented. According to Hecht and Reiners explanation (Hecht & Reiner, 2009), it seems that when the tri-sensory stimulus is given, the probability of missing two signals is much smaller than missing only one signal; therefore, the visual dominance disappears.

Hecht and Reiners conclusions may indirectly support our assumption that a visual-haptic stimulus may have more of an impact on subjects' listening perception. Yet, due to the multidimensional properties of music, it is not possible to generalize these conclusions to music listening along with a synchronized non-audio stimulus. Therefore, it is necessary to consider that specific properties of music may also have an impact on the selective listening of a single melody in a polyphonic composition. For example, in polyphonic music, two voices are more recognizable if their respective melodies have a closer relation in pitch and when their notes' onsets have the same tempo. In such a case, participants begin to rely on timbre to distinguish them (Gregory, 1990). In addition, the instruments timbre seems to play an important role as well. For instance, a very homogeneous timbre can be used to obfuscate the listeners' abilities to identify the concurrent number of voices present in a polyphonic song (Huron, 1989).

The psychoacoustic theory of "auditory stream segregation" proposes that humans can effortlessly segregate tones sequences into individual streams or melodies based on the tone (Bregman, 1978). Additionally, it has been demonstrated that our perception of melodic structures can also be manipulated if the piece is cleverly composed, such as in many of Bach's contrapuntal compositions (Wright & Bregman, 1987). Further, the temporal relation between the voices seems to be an important clue for stream segregation (Ragert, Fairhurst, & Keller, 2014). As such, for selective listening, the specific structure of the melody and its registry are influential factors that must be considered.

For the effective design of new multimodal listening or playing interfaces, one must go beyond a consideration of the fundamental capabilities of the somatosensory and auditory senses. It is also important to consider humans intrinsic capabilities to perceive music. Thus, it seems necessary to perform a direct analysis of the use of multimodal stimuli on selective music listening to lay a foundation for future multimodal music studies and the design of new music interfaces.

3. Evaluation

To measure participants selective listening, a real-time self-reported evaluation was used (Madsen et al., 1993). In the empirical experiment, all of the participants ($n = 30$) listened to a song while a synchronized non-audio stimulus was displayed to highlight a single instrument in the song. The participants were then asked to press down a button if they were able to audibly identify and follow the highlighted instrument. Or release the button if they were not able to match the given stimulus with an instrument in the song.

3.1. Participants

Thirty participants, inclusive of 23 males and 7 females and aged between 24 to 60 years old, participated in the experiment. All the participants were students or members of the university who agreed to take part in the study. Due to the nature of the evaluation, it was necessary to identify the participants musical capabilities. Therefore, we asked the participants whether they were able to play a musical instrument and if they were able to currently play a complete song with their instruments. The participants who answered affirmatively to these specific questions were considered music players. On the other hand, those who answered '*no*' to any the questions were considered non-players. In summary, from the 30 participants, 11 were music players, and the other 19 were non-players.

Only 5 participants from the non-players group received extracurricular instrument lessons before they were 9 years old (Gordon, 1980). However, none of these participants reported being able to nowadays fluently play their studied instrument.

Song	Instrument	High	Low	Average
Mozart	Violin-1	89 - G5	55 - G2	75 - Eb4
	Violin-2	84 - C5	55 - G2	67 - G3
	Viola	68 - G#3	48 - C2	60 - C3
	Cello	61- C#3	36 - C1	48 - C2
Haydn	Violin-1	88 - E5	57 - A2	74 - D4
	Violin-2	83 - B4	56 - G#2	69 - A3
	Viola	76 - E4	49 - C#2	63 - Eb3
	Cello	66 - F#3	37 - C#1	52 - E2
Bach	Organ-1	84 - C5	57 - A2	73 - C#4
	Organ-2	80 - G#4	50 - D2	65 - F3
	Organ-3	74 - D4	38 - D1	59 - B2
	Organ-4	50 - D2	24 - C0	38 - D1

Table 1. The instruments' pitch register and average played note.

3.2. Stimuli

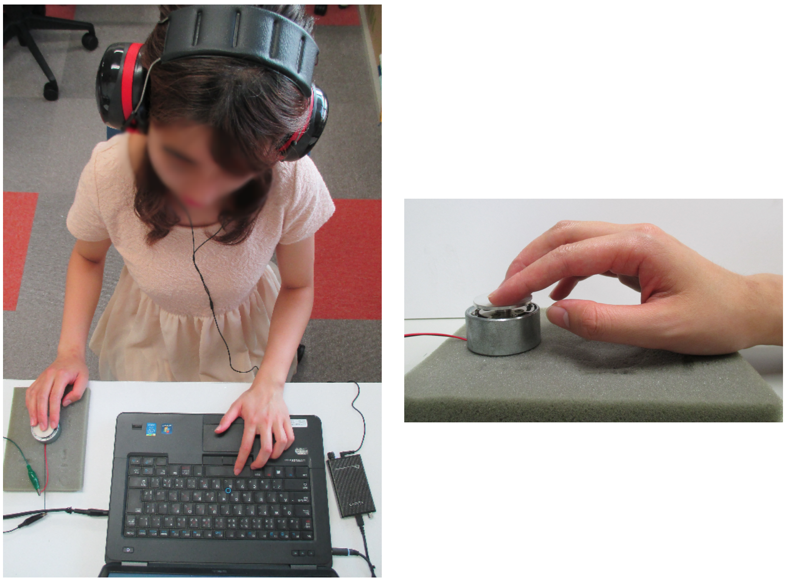
The visual and haptic modalities of the non-audio stimulus were designed to have the same rhythm, duration, and pitch as a single instrument in the song so that the results across stimulus conditions could be directly compared. The visual-haptic stimulus was generated by simultaneously combining the visual and haptic stimuli, and the single modality stimuli were considered as controls.

3.2.1. Musical Stimuli

The songs used in the evaluation were MIDI renditions of the following classical songs: Haydn - String Quartet Op. 20 No. 6 Final movement (2 violins, viola, and cello), Mozart - Adagio & Fugue K.546 Final 4 minutes (2 violins, viola, and cello), and Bach - Little Fugue BWV-578 (4 voices with organ only). For the experiment, Windows 7 default synthesizer (Microsoft GS Wavetable Synth) was used to play these songs.

When Mozart and Haydn songs were played with the selected synthesizer, the violin (MIDI-40), viola (MIDI-41) and cello (MIDI-42) timbres did not contrast too much between each other, so any of these instruments highlighted itself due to its particular timbre. For Bach, the 4 voices were played by a MIDI church organ (MIDI-19); therefore, there was no contrast in the timbre. Also, all the notes of all of the instruments were played at the same volume to avoid any variations that could have inadvertently captured the participants listening attention.

All of these selected songs are polyphonic fugues, which allowed us to define challenging but not trivial selective listening tasks. In general, a fugue is a compositional technique in which two or more short melodies, called voices, are introduced and then successively taken up by other voices in a continuous interweaving of the melodies. The melodic complexity between the different voices in fugues makes them ideal for creating challenging selective listening tasks. After all, if the selected song is too simple, then a focused listening task may become trivial. The specific MIDI renditions of these polyphonic songs were selected because their individual voices do not play overlapped notes, thus simplifying the design of the visual and haptic stimuli. In other words, the songs themselves are polyphonic, but the notes of each voice have a unique start and duration that do not overlap with any other note played by the same voice. Therefore, the properties of each individual note can directly define the haptic and visual stimuli properties.



(a) (b)

Figure 3. a) While listening to the music, the participant faced a laptop where the visual stimulus was displayed. He/she used his/her left hand to report his/her selective listening by pressing the spacebar key. b) To feel the haptic signal, the participant's right-hand index finger touched the surface transducer.

The instruments' pitch registers, between songs, are quite similar. Since Haydn's and Mozart's songs are played with the same instruments, their pitch registers are similar. In the case of Bach, the organ plays the pitch register of each voice, but their individual voice registers are also comparable to the register played by the violin, viola, and cello in Mozart's and Haydn's songs (see Table 1). The similarity between the instruments pitch registers allowed the inclusion of the pitch register as an effect in the statistical analysis. For detailed information about the songs' scores and the specific sections of the songs selected as measuring lapses, please see the Appendix.

3.2.2. Visual stimulus

The visual modality was designed using a simplistic and self-explanatory visual stimulus. For this stimulus, one by one, the notes played by the highlighted instrument were represented with a single yellow rectangle, which appeared and disappeared from the screen according to a note's properties. The length of the rectangle represented the note's duration, while the note's pitch was represented by the relative position of the rectangle across the Y-axis on the screen. Therefore, a note with a higher pitch was displayed higher on the screen, while a note with lower pitch was displayed lower on the screen. In order to hide the temporal melodic structure of the highlighted voice, only one single rectangle was displayed on the screen, which represented the properties of every single note the highlighted instrument was playing at the time, so once the note ended, its respective rectangle disappeared (see Figure 2).

3.2.3. Haptic stimulus

To design the haptic stimulus, it was necessary to consider the capabilities and limitations of the somatosensory sense in order to propose an effective method to represent the notes' pitch of the highlighted instrument as a haptic signal. Compared to the extraordinary capability of the human ear to detect frequency variations (Wolfe et al., 2006) the touch sense performance to distinguish frequency changes is not so effective, with a noticeable difference around 18% (Pongrac, 2006).

Therefore, we developed a simple method to map the notes key in a predefined haptic frequency range. Specifically, the notes with a higher key had a higher haptic frequency, while the notes with a lower key had a lower haptic frequency. Additionally, this method maps the notes' pitch in a predefined haptic frequency range ($\triangle f$) while trying to maximize the haptic frequency variations between the notes. As such, they can be easily perceived. Therefore, $\triangle f$ can be defined as being in the human tactile frequency range (5Hz to 700Hz) (Lofvenberg & Johansson, 1984).

To perform the frequency mapping, a range between the lowest key note (k_{min}) and the highest key note (k_{max}) was defined as: $\triangle k$ (see Table 1). Additionally, a maximum (f_{max}) and minimum (f_{min}) haptic frequency range was defined as: $\triangle f$ (See Equations 1 & 2). In this case, $\triangle f$ was set between $f_{max} = 300$Hz and $f_{min} = 30$Hz. This specific frequency range was selected because humans have a better frequency discrimination at low frequencies (Verrillo, 1992).

$$\triangle k = k_{max} - k_{min} \tag{1}$$

$$\triangle f = f_{max} - f_{min} \tag{2}$$

$$\triangle a = a_{max} - a_{min} \tag{3}$$

The proposed algorithm uses two lineal relations to map the notes' key in the predefined frequency range ($\triangle f$). The first lineal relation calculates the note's relative position in the song (see Equation 4) by using the note's key (k), the range between the lowest and highest key in the song ($\triangle k$), the number of notes with a different pitch (k_ϵ) and the key of the lowest note ($\triangle k$).

$$n = \frac{k_\epsilon \cdot (k - k_{MIN})}{\triangle k} \tag{4}$$

The other lineal relation is used to calculate the final haptic frequency (f_h) between the predefined range ($\triangle f$) by using the note's relative position (n) (see Equation 5).

$$f_h = f_{MIN} + \left(\frac{\triangle f}{k_\epsilon} \cdot n \right) \tag{5}$$

In addition, the initial amplitude of the haptic signal (a_h) is defined using the same strategy applied for the haptic frequency. The note's key (k_ϵ) is also used to define the initial attack amplitude of the signal (a_h). Therefore, notes with a higher pitch will have a higher initial amplitude, and the notes with a lower pitch will have a lower initial amplitude. First, a haptic amplitude range is predefined as $\triangle a$; the note's relative position is then calculated using the equation 4. Finally the haptic attack amplitude (a_h) is calculated using the equation 6.

$$a_h = a_{min} + \left(\frac{\triangle a}{k_\epsilon} \cdot n \right) \tag{6}$$

Instrument	$\boldsymbol{t_a}$ ms	$\boldsymbol{t_d}$ ms	$\boldsymbol{t_s}$ ms	$\boldsymbol{s} = a_h\%$	$\boldsymbol{r} = a_h\%$
Organ	5	0	t_4-25	100	1
Violin	66	376	t_4-225	57.14	1
Cello	40	290	t_4-200	47.45	1
Viola	66	180	t_4-125	80.0	1

Table 2. ADSR timing and amplitude parameters for the haptic envelopes defined for each instrument.

The variations in pitch, represented in the visual stimulus by the rectangle position, are evidently perceivable. Therefore, to compensate for the limitation of the haptic sense to perceive frequency variations, we did not rely on the haptic frequency shift alone to represent the notes of the highlighted instrument. We took advantage of the noticeable capabilities of the somatosensory sense to distinguish small amplitude variations (1.7~19 mN for frequencies at 320Hz) (Hatzfeld, Cao, Kupnik, & Werthschtzky, 2016), so the waveform characteristics of the highlighted instrument were also included in the haptic signal envelope.

To be able to control the haptic signal frequency and amplitude with precision, an ADSR (Attack-Decay-Sustain-Release) envelope generator (Jensen, 1999) was used in order to define the envelope shape of the haptic signal. The ADSR parameters were configured so the the haptic signal envelope would have the same characteristics as the general audio envelope shape of the highlighted instrument. In our specific implementation, a logarithmic curve was used for the release phase, while linear equations were used for the rest of the ADSR phases.

For the haptic waveform, the timing of each ADSR phase was defined as follows: t_a defined the attack-decay inflexion time, t_d defined the decay-sustain inflexion, time and t_s defined the sustain-release inflexion. The specific amplitude levels of each ADSR phase were determined by the peak amplitude at attack a_h, the amplitude at sustain s, and the amplitude at the end r. Table 2 details the specific parameters used to define the haptic signal for each instrument present in the songs used in the evaluation.

3.2.4. Visual and Haptic stimuli

The visual-haptic stimulus was created using a synchronized combination of the previously described visual and haptic stimuli, which were created and displayed using the same aforementioned strategies. Since the goal of this evaluation was to estimate the effectiveness of a multimodal haptic-visual stimulus for selective listening perception and since the visual stimulus and haptic stimulus were designed to provide the same amount of information, we considered both single modality conditions as control conditions.

3.3. Procedure

For each song, 9 measuring lapses, which were each 15 seconds long, were predefined, with a break of at least 5 seconds in between each. During each measuring lapse, the participants' selective listening and detection time were measured. The selective listening time was quantified by saving the total time the response button was pressed down, while the detection time was quantified by saving the time from the beginning of the lapse until the first time the button was pressed down. (see Figures 1 & 2).

The measuring lapses start time and the highlighted instrument were predefined according to the particular structure of the song during each lapse. A different highlighted instrument was selected for each lapse, so all of the instruments present in the song were highlighted at least once. Therefore, the measuring lapse's start and highlighted voice were specifically predefined to avoid creating confusing tasks as well as tasks with obvious answers. The measuring lapses were defined in places where the notes of the highlighted instrument were not masked by other instruments playing the exact same melody. Additionally, to avoid confusion, we abstained from placing the measuring lapses prior to the instruments playing of a single note for the same amount of time. Thus, by these means, we attempted to control the difficulty of the tasks and to avoid impossible or arbitrary inquiries (see Appendix). In addition, the lapses' start and highlighted voice were predefined across conditions in order to be able to directly compare the participants' responses to each measuring lapse.

A MIDI sequencer was used to ensure that the lapses' start and the voices fulfilled the conditions mentioned above.

To prevent the participants from remembering the pre-defined highlighted voices of each measuring lapse, the complete experiment occurred on 3 different days, with at least 4 days between each application. On a single day, the songs and the feedback condition order were randomly selected without repetition so that the participants would not repeat the same song or condition during the same day.

To verify their understanding of the instructions, the participants performed a practice run with a polyphonic song that was not used in the main experiment (Bach BMW 1079 - Largo). Apart from the practice, no other instructions or information about the music was given to the participants.

In a quiet room with a comfortable temperature, the participants sat in front of a notebook PC placed on a desk, using their left hand to press the space bar while their right hand was placed over the haptic device. The vibration was displayed with an external DAC (xDuoo XD-01) and a surface transducer (Adafruit 4Ω 5Watt), while the music was played using the PC's audio card (Realtek ALC3226) with stereo in-ear earphones (Sennheiser MX 475). The haptic and visual signals were created and syncronized with the music with an control program, written in C++. Also the same program captured and saved the participants' responses.

Most digital audio cards, due to their inherent design, have a random delay output. Therefore, it was necessary to verify that the delay between the audio and haptic signals was below the reported human audiotactile asynchrony perception threshold (23ms) (Adelstein, Begault, Anderson, & Wenzel, 2003). The random delay between the audio cards was measured and controlled to be ±2.5ms. In order to have such control on the signals timing, it was necessary to use real-time audio drivers (Tippach, 2003) to reduce the audio card's buffer size. It was also necessary to use real-time audio libraries that provided support for such drivers (Cook & Scavone, 1995).

In order to isolate the transducer vibration sound from the music, the participants used noise-cancelling earmuffs over the earphones. Additionally, to avoid vibration transmission to the desk, the transducer was placed on 2cm layer of casing foam. Moreover, all of the participants were instructed to touch the surface transducer using only their fingertips. They were instructed to place them on any part of the transducer surface where they could encounter an stronger sensation (see Figure 3).

If applied incorrectly, self-reported evaluation methods could generate fragile, random, or biased data (Stone, Shiffman, Atienza, & Nebeling, 2007). To avoid these issues, several measures were taken in order to avoid participant distraction, which may have affected the participants responses.

The amount of time the participants performed the selective listening task was limited. Therefore, the predefined measuring lapses were exactly 15 seconds long. In addition, a visual warning, which was 3 seconds in length, was displayed before each measuring lapse began. This visual warning captured the participants attention so they could be completely focused on finding and following the indicated instrument during the measuring lapse (see Figure 1).

Additionally, and prior to the main experiment, a single participant, who performed the described experiment three times with four songs and using all conditions, tested the repeatability of the proposed method. The generated results showed a clear data correlation between the experiments of this particular participant. Also, the test's data consistency was checked by calculating the Cronbach's alpha coefficient (Cronbach, 1951; Tavakol & Dennick, 2011), which showed good consistency between songs with a Cronbach's alpha coefficient of: $\alpha \geq 0.7$. Additionally, we calculated the Cronbach's alpha coefficient to confirm the data consistency of the main evaluation results. A value of $\alpha > 0.72$ was obtained. Therefore, the final experiment did not generate random results, and the measurements were consistent among the participants.

3.4. Questionnaire

Immediately after the participants completed all of the trials of the experiment, a subjective questionnaire was used in order to measure the participants' different perspectives of the three feedback conditions. A graphic rating scale was used to evaluate their preferences, understanding, and the ease with which they performed the task associated with each stimulus condition. Three graphical rating scale questions were asked: "*Did you like the experience?*", "*Was the stimulus easy to understand?*" and "*Was the stimulus easy to follow?*". The rating scale was from -100 to 100 for a completely negative response to a completely positive response, respectively. The descriptive adjectives that appeared in the questions were: "*Like the experience*", "*Dislike the experience*", "*Easy to understand*", "*Difficult to understand*", "*Easy to follow*" and "*Hard to follow*", respectively.

4. Results & Discussion

4.1. General Analysis

To determine whether the different factors had an influence on the participants responses, a multivariate analysis of variance (MANOVA) was performed. Since all of the songs included voices with a similar pitch register (see Table 1), it was possible to average the participants responses with a similar pitch register across different lapses and songs. Hence, the selective listening and detection responses of each measuring lapse were grouped and averaged, across songs, based on their respective highlighted voice pitch register. So for this analysis the lapses were grouped as follows: High (Mozart:Violin-1, Haydn:Violin-1 & Bach:Organ-1), Medium-1 (Mozart:Violin-2, Haydn:Violin-2, Bach:Organ-2), Medium-2 (Mozart:Viola, Haydn:Viola, Bach:Organ-3) and Low (Mozart:Cello, Haydn:Cello, Bach:Organ-4).

Therefore, a three-factor multivariate ANOVA was performed using the following main effects and levels: the stimulus condition (with 3 levels: visual, haptic, and visual-haptic), the participant's skill (with 2 levels: players and non-players), and the voices' pitch register (with 4 levels: High, Medium-1, Medium-2, and Low); while the

Effect	**F**	**df**	**Error df**	**P value**
Stimulus	11.400	4.000	672.000	**0.000**
Skill	30.816	2.000	335.000	**0.000**
Pitch	5.100	6.000	672.000	**0.000**
Stimulus*Skill	0.192	4.000	672.000	0.943
Stimulus*Pitch	0.873	12.000	672.000	0.574
Skill*Pitch	0.519	6.000	672.000	0.794
Stimulus*Skill*Pitch	0.323	12.000	672.000	0.985

Table 3. Multivariate ANOVA results table, inclusive of Pillai's trace. All of the main effects (stimulus, skill, & pitch) presented were significant ($p < 0.01$). However there was not a significant interaction between any of the simple main effects, which indicates the lack of any particular significant interaction between the participant's skill level or the voice's pitch with the stimulus condition.

participant's detection and listening time were the dependent variables.

The data obtained from the evaluation satisfied almost all of the ANOVA assumptions. The variance and errors between the cells were similar. The measurements were independent from each group because all the participants performed the experiment with a random stimuli and song order. However, as expected for the reaction measurements (Roberts & Russo, 2014), the detection and the listening time measurements were skewed. This skewness was normalized using a logarithmic transformation; thereafter, the same multivariate ANOVA test was applied to the original and transformed data in order to verify any difference in the statistical analysis between the transformed and untransformed data. Since both tests had the same results and interactions, the original untransformed data were used in this report.

In some cases, the participants were unable to provide an answer or their initial detection was unreflective and immediate (i.e., faster than the average reaction to haptic and visual stimuli < 200ms (Frith & Done, 1986) (Shelton & Kumar, 2010)). In these cases, the generated data were replaced by the lapse's average in that stimulus condition. Only less than the 3% of the data were eccentric or missing, which allowed the use of this specific measure to fix this issue.

The results of the multivariate ANOVA, as shown in Table 3, were calculated with SPSS using Pillai's trace. These results show a significant main effect of the stimulus condition ($F(4.0, 672.0) = 11.4, p < 0.01$) and a significant main effect of the participant skill level ($F(2.0, 335.0) = 30.816, p < 0.01$). In addition, the voices' pitch effect was significant ($F(6.0, 672.0) = 5.1, p < 0.01$).

On the other hand, in regard to the simple main effects, a significant effect did not exist between the stimulus type and the participant's skill level ($F(4.0, 672.0) = 0.192, p > 0.01$), suggesting that the performance difference of each stimulus condition was the same, regardless of the participants' abilities. Moreover, a significant effect was not found between the stimulus type and the voices' pitch ($F(12.0, 672.0) = 0.873, p > 0.01$), thus showing that the performance of each stimuli was the same, regardless of the voice's pitch. Further, a significant effect was not present between the participants skill level and the voices pitch ($F(6.0, 672.0) = 0.519, p > 0.01$), meaning that players and non-players performances were the same regardless of the voice's pitch. Finally, a significant effect involving the combination of the three main effects was not discovered (Stimulus*Skill*Pitch) ($F(12.0, 672.0) = 0.323, p > 0.01$). In summary, none of the simple main effects combinations presented a significant interaction, as shown in the four last rows of

Table 3.

Since the stimulus main effect was significant, a Tukey HSD pairwise comparison was performed in order to assess which of the feedback stimuli means were significantly different. The results of this analysis, depicted in Figures 4a & 4b, indicate that the participants performance with the visual-haptic was significantly better, with faster detection times and longer listening periods, compared to the visual-only and haptic-only conditions. The visual-only and haptic-only means do not show any significant difference in their detection time or listening time.

A pairwise comparison between the players and non-players groups was also performed. The results indicate a significant difference between the player and non-player means, for listening time as well as for detection time (see Figures 4c & 4d). As expected, the participants who had experience playing an instrument obtained significantly faster detections and longer listening periods compared to the participants who did not have experience playing an instrument.

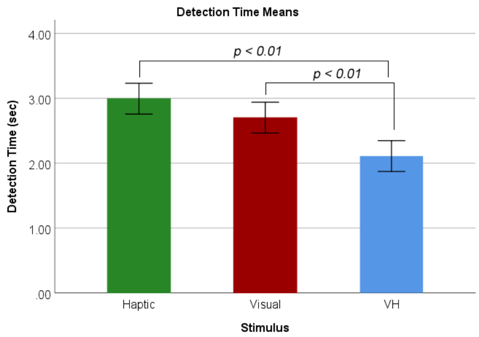
The multivariate ANOVA results indicate that the performance of skilled participants was significantly better compared to that of unskilled participants, although the participants performance was significantly better in both groups when using the visual-haptic stimuli. As shown in the interaction plots between participants skill level and the stimulus condition(see Figures 5a & 5b). No interactions were detected between the participants skill level and the given stimulus in the detection and listening time measurements. This indicates that the visual-haptic stimuli performance improvement was the same regardless of the participants' music skills.

Furthermore, as shown in the interaction plots featuring the voices' pitch and the stimuli (see Figures 5c & 5d), the participants performed better when using the visual-haptic stimulus across voices with a different pitch. This indicates that, regardless of the voices' pitch, the participants had longer listening periods and faster detections with the visual-haptic stimulus, compared with the single modality stimuli. The same plots showed an interaction between visual and haptic conditions only for voices with a low pitch register, where the haptic stimuli obtained better results. Although, the lack of interaction between the voices' pitch and the stimuli conditions (see Table 3) do not justify a post-hoc comparison.

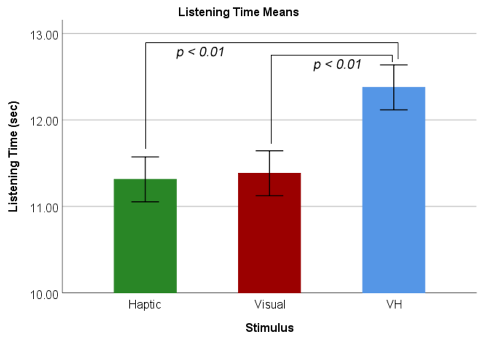
In addition to the psychometric evaluation, a subjective perception questionnaire was used to compare the participants' perception of the performed task in different stimulus conditions. A graphic rating scale with three questions was used: "*Did you like the experience?*", "*Was the stimulus easy to understand?*" and "*Was the stimulus easy to follow?*". To compare the results, a single factor ANOVA, between the conditions, was performed. The results, as shown in Figure 6, indicate that the participants found the visual-haptic condition more understandable and easier to follow compared to the visual-only and haptic-only conditions. Also, the participants showed a high likeness for the haptic-stimuli; it was almost as high as the visual-haptic likeness results.

4.2. Detailed analysis

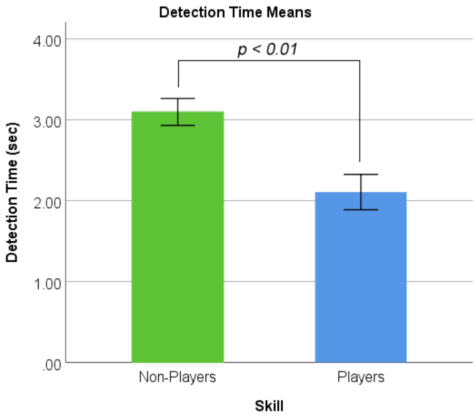
In the previous section, we demonstrated that the participants performance improved with visual-haptic stimuli, regardless of the participants skill level and the voices' pitch registry. However, the variance differences between the lapses measurements strongly



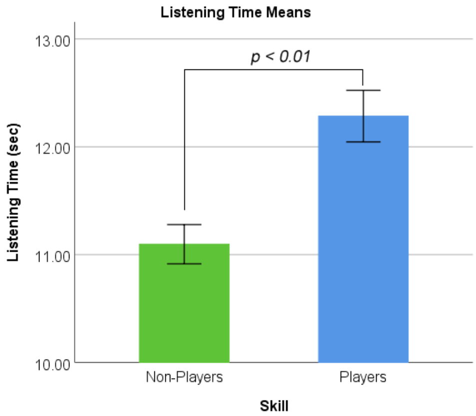
(a) Detection mean time between stimulus conditions



(b) Listening mean time between stimulus conditions



(c) Detection mean time between players and non-players



(d) Listening mean time between players and non-players

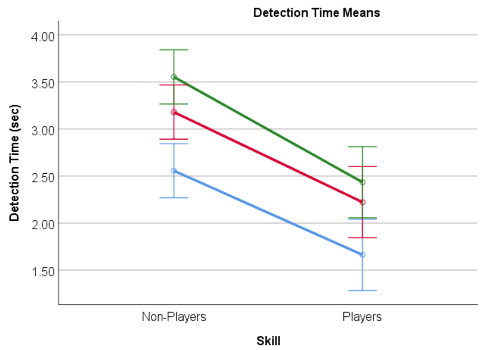
Figure 4. Bar plots that compare the detection and listening time means between stimuli conditions and the participants skill level, where error bars indicate a 95% confidence interval. Plots (a) and (b) show a significant better performance of the players versus the non-players, with faster detection times and longer selective listening periods, while (c) and (d) show the significant performance increase of the visual-haptic stimulus (VH) compared with the single modality stimuli, with faster detection times and longer listening periods.

indicate that the unique musical structure of each measuring lapse was a factor in the experiment (see Figure 7). Yet, since an objective or numerical strategy to measure the complexity of a polyphonic melody does not exist, the melody complexity of each measuring lapse was not viewed as a factor in the previous multivariate ANOVA.

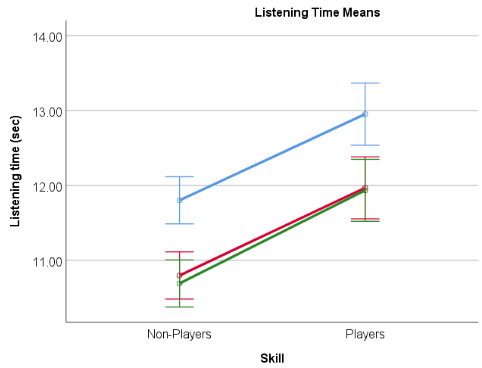
Instead, we performed an individual multivariate ANOVA analysis of each song. Then, considering the measuring lapses and the stimuli conditions as the main effects, a two-factor multivariate ANOVA was performed using the following main levels: the stimuli condition (with 3 levels: visual, haptic, and visual-haptic) and the measuring lapses (with 9 levels, one for each measuring lapse). The participants detection and listening times were considered as dependent variables. By these means, a significant interaction between the measuring lapses and the stimulus conditions was identifiable.

The results of the multiple multivariate ANOVA are shown in Table 4. The table of results shows the significant main effect of the stimulus condition across the three songs (Mozart $F(16, 1566) = .352, p < .001$, Haydn $F(16, 1566) = .097, p < .001$ and Mozart $F(16, 1566) = .216, p < .001$). Also, there is significant main effect of the measuring lapses, although this interaction was expected due to the melodic difference of each measuring lapse.

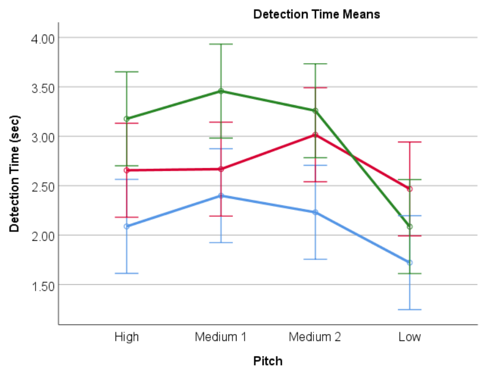
On the other hand, as in the general analysis, there was no significant effect between stimuli and the measuring lapses in any of the three songs (Mozart $F(32, 1566) = .051, p > .05$, Haydn $F(32, 1566) = .038, p > .05$ and Mozart $F(32, 1566) = .052, p > .05$). The lack of any significant interaction between the measuring lapses and the stimulus condition indicates that the performance difference



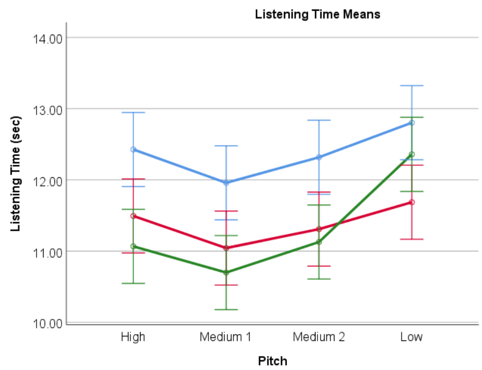
(a) Interaction plot of the detection time between stimulus conditions and the participants' skill.



(b) Interaction plot of the listening time between stimulus conditions and the participants' skill.



(c) Interaction plot of the detection time between the voices' register and the stumulus conditions.



(d) Interaction plot of the listening time between the voices' register and the stumulus conditions.

Figure 5. Interaction plots between skill*stimulus and voice*stimulus, where error bars indicate a 95% confidence interval. Plots (a) and (b) suggest that players and non-players had a better performance with the visual-haptic stimulus compared to single modality stimuli. The performance difference between stimuli was the same, regardless of the participants' music skills. Plots (c) and (d) also show a performance improvement with the visual-haptic stimuli across voices with different pitches. Therefore, the improvement caused by the visual-haptic stimuli was not related to the participants skill level or the voice's pitch.

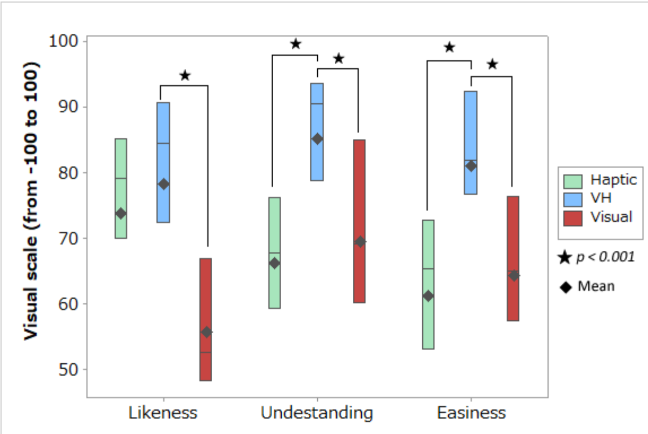


Figure 6. Box plot of the visual scale questionnaire results, which shows the participants' likeness, understanding and the easiness of the task across modalities. The boxes demonstrate the median confidence area, the diamonds show the measurements' mean and the stars indicate the mean statistical difference between stimulus conditions.

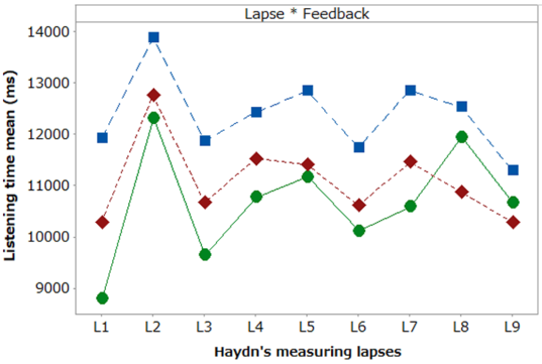
Song	Effect	F	df	Error df	P value
Mozart	Lapse	.352	16.000	1566.000	**.000**
	Stimulus	.048	4.000	1566.000	**.000**
	Lapse*Stimulus	.051	32.000	1566.000	.144
Haydn	Lapse	.097	16.000	1566.000	**.000**
	Stimulus	.068	4.000	1566.000	**.000**
	Lapse*Stimulus	.038	32.000	1566.000	.563
Bach	Lapse	.216	16.000	1566.000	**.000**
	Stimulus	.034	4.000	1566.000	**.000**
	Lapse*Stimulus	.052	32.000	1566.000	.114

Table 4. The individual multivariate ANOVA results of each song were performed with Pillai's trace. The lapse and the stimulus conditions had a significant effect ($p < 0.05$) in all songs. However, there is no significant interaction between the lapses and the stimuli in any of the songs. This finding indicates that the participants performance was not affected by any particular combination of the stimulus condition and the melody.

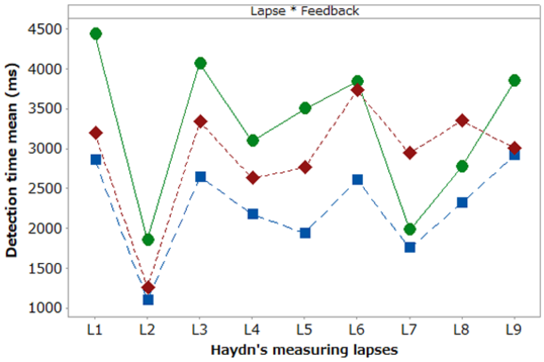
between the stimuli was the same across measuring lapses and songs. The interaction plots also reflected the same results, with parallel lines across most of the measuring lapses (see Figure 7) showing no significant individual interaction across the stimuli and the measuring lapses.

Since the stimulus main effect was significant, three Tukey HSD pairwise comparisons were performed, for each song, in order to find which feedback stimuli means were significantly different. The results, as shown in Figures 8a & 8b, indicate a significantly better participant performance when using the visual-haptic stimulus compared to the single modality stimuli. The results also reveal the lack of a significant difference between visual-only and haptic-only conditions, as they had the exact same results across the three songs.

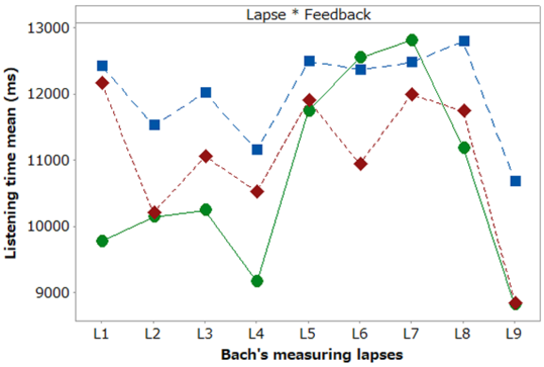
In conclusion, the multivariate ANOVA results and the pairwise comparison results indicate that regardless of the song or the measuring lapse, the participants' performance was significantly better with the visual-stimuli, while there was no significant difference between the single modality conditions. And the lack of conclusive evidence of the interaction between the stimulus condition and the measuring lapses do not allow to precisely indicate a relation lapses melodic structure with an specific stimulus



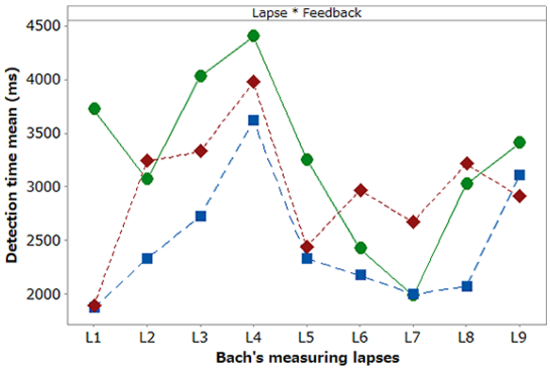
(a) Haydn's listening time interaction plot



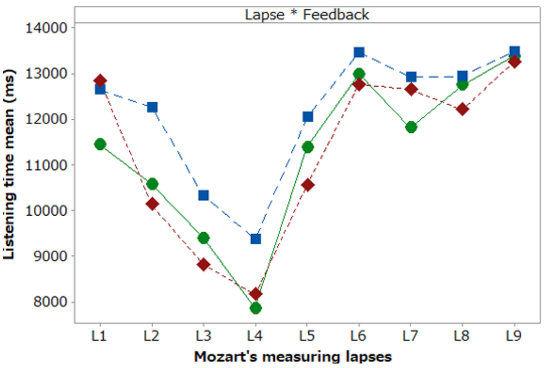
(b) Haydn's detection time intereaction plot



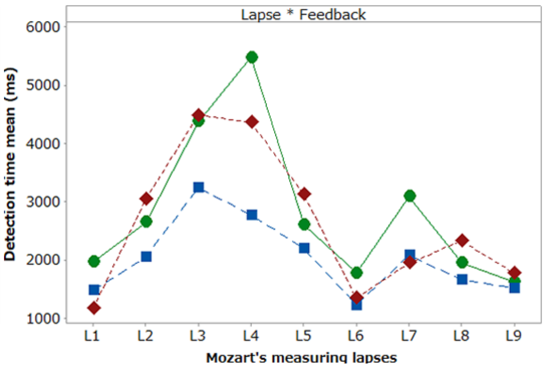
(c) Bach's listening time interaction plot



(d) Bach's detection time interaction plot



(e) Mozart's listening time interaction plot



(f) Mozart detection time interaction plot

Figure 7. The detection and selective listening time interaction plots of each song. In general, the visual-haptic condition presented faster detection times and longer selective listening periods. Additionally, in most cases, the interaction lines between the stimuli ran parallel with each other, which demonstrated the lack of a significant performance difference caused by an specific combination of the given stimuli and measuring lapse. Therefore, the stimuli condition performance difference did not significantly change across measuring lapses.

condition.

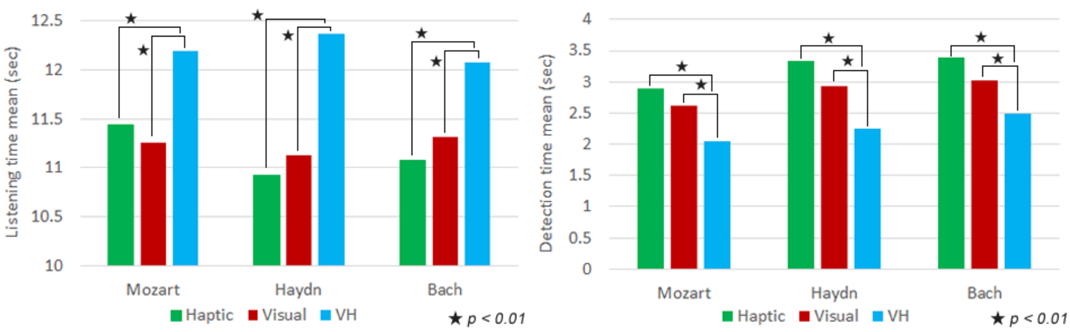
Overall, in the general and detailed multivariate analyses, none of the simple main effect combinations presented a significant interaction (see Table 3 & 4)As such, the lack of clear statistical evidence does not allow for a post-hoc analysis of any of the simple main effects interactions. Therefore, consistent evidence of the interaction between and specific stimulus with an specific kind of melody does not exist. However, we suspect that, due their own nature, the visual stimuli should more effective to display the melodic structure of the song, while the haptic stimuli should more effectively highlight the pitch of a specific instrument. Yet, in order to demonstrate this hypothesis, it may be necessary to perform a different study with specific music stimuli.

5. Conclusion

The results obtained through this analysis support our hypothesis. The multivariate analysis of the participants' performance between conditions demonstrated significantly longer periods of focused listening during most listening tasks, when the participants used a synchronized combination of haptic and visual stimuli. In addition, the multivariate analysis did not show any significant interactions between the highlighted voice pitch registry, the participants musical skills, and the task complexity with the provided stimuli. Hence, regardless of the listening task, voice, and the participants music skills, the subjects performed better when using the redundant visual-haptic stimulus.

These conclusions do not come across as surprising, but if the Colavita visual dominance effect (Spence, 2009) is considered, then it seems necessary to evaluate if in fact a synchronized visual-haptic stimulus could provide some perceptual advantage compared to a visual or haptic stimulus alone. Other factors must also be considered, such as the participants music skills, the voices' pitch (Bregman, 1978), and the complexity of the listening task (Wright & Bregman, 1987). Therefore, even if the hypothesis seems logical, all of these factors need to be considered as well.

The performed empirical evaluation shows that the listening performance improve-



(a) (b)

Figure 8. Results of the pairwise comparison between the stimulus' modalities. The results, obtained with Tukey HSD, indicate a significant difference between the visual-haptic stimulus and the single modality conditions in all of the songs. Also, the results associated with the songs do not show any statistical differences between the visual-only and haptic-only conditions.

ment, which was associated with the visual-haptic stimulus, was not affected by: the participants music skills, the voices' pitch register or the melody. Therefore, the proposed visual-haptic stimulus come across as an ideal mean to enhance users ability to focus on a specific voice in a polyphonic melody. Hence, this kind of multimodal stimulus may have a direct application for entertainment purposes or children's music education (Gordon, 1980).

Further, the obtained results are similar to Hechts findings (Hecht & Reiner, 2009). We could not find any evidence of visual dominance or haptic-dominance when the participants performed the selective listening task with the visual-haptic condition (tri-modal). In fact, the multivariate analysis did not show any difference between the visual-only and haptic-only conditions, while both demonstrated a significantly worse performance compared to the redundant visual-haptic condition, even when the visual-haptic stimuli were only a part of the synchronized combination along with haptic-only stimuli. Therefore, as also mentioned by Hech, we suspect that the participants performance was better with the visual-haptic stimulus because the probability of missing a note with two redundant non-audio signals was much smaller that missing a note with only one non-audio signal.

Also, we are aware that the main constrain of the proposed experiment was the diversity of the music, because only 27 different measuring lapses were used in the experiment. Although the clear tendency in the obtained statistical results strongly suggest that a redundant visual-haptic stimuli is in fact effective to improve the selective listening. So as a future work it is necessary to perform a similar experiment using a larger diversity of music stimuli, to be able to confirm if there is any relation between any of the stimuli modalities with an specific type of melody structure.

Finally, we believe that the results associated with the performed experiments can be directly applied to the upgrade of the design of recent virtual environments focused on enhancing individuals music listening experience. Therefore, it is necessary to study the limits and capabilities of humans multimodal perception in order to provide a psychophysical foundation for future VR and AR research.

6. Appendix - Detailed Music Information

The song's MIDI files and all of the information related to the measuring lapses (start, ending and highlighted voice) are available in this url: https://goo.gl/3Z8oJR.

References

Adelstein, B. D., Begault, D. R., Anderson, M. R., & Wenzel, E. M. (2003). Sensitivity to haptic-audio asynchrony. In *Proceedings of the 5th international conference on multimodal interfaces* (pp. 73–76). New York, NY, USA: ACM.

Armitage, J., & Ng, K. (2016). Feeling sound: Exploring a haptic-audio relationship. In R. Kronland-Martinet, M. Aramaki, & S. Ystad (Eds.), *Music, mind, and embodiment: 11th international symposium, cmmr 2015, plymouth, uk, june 16-19, 2015, revised selected papers* (pp. 146–152). Springer International Publishing.

Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance*, *4*(3), 380.

Cook, P. R., & Scavone, G. P. (1995). *Ccrma - the synthesis toolkit in c++ (stk).* Retrieved from https://ccrma.stanford.edu/software/stk/

Cronbach, L. J. (1951, Sep 01). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. Retrieved from https://doi.org/10.1007/BF02310555

Frith, C. D., & Done, D. J. (1986, Nov 01). Routes to action in reaction time tasks. *Psychological Research*, *48*(3), 169–177. Retrieved from https://doi.org/10.1007/BF00309165

Goldstein, A. (1980, Mar 01). Thrills in response to music and other stimuli. *Physiological Psychology*, *8*(1), 126–129. Retrieved from https://doi.org/10.3758/BF03326460

Gordon, E. E. (1980). The assessment of music aptitudes of very young children. *Gifted Child Quarterly*, *24*(3), 107–111.

Gregory, A. H. (1990). Listening to polyphonic music. *Psychology of Music*, *18*(2), 163-170. Retrieved from https://doi.org/10.1177/0305735690182005

Hatzfeld, C., Cao, S., Kupnik, M., & Werthschtzky, R. (2016, Oct). Vibrotactile force perception absolute and differential thresholds and external influences. *IEEE Transactions on Haptics*, *9*(4), 586-597.

Hecht, D., & Reiner, M. (2009, Feb 01). Sensory dominance in combinations of audio, visual and haptic stimuli. *Experimental Brain Research*, *193*(2), 307–314. Retrieved from https://doi.org/10.1007/s00221-008-1626-z

Huron, D. (1989). Voice denumerability in polyphonic music of homogeneous timbres. *Music Perception: An Interdisciplinary Journal*, *6*(4), 361–382. Retrieved from http://mp.ucpress.edu/content/6/4/361

Hwang, I., & Choi, S. (2014). Improved haptic music player with auditory saliency estimation. In M. Auvray & C. Duriez (Eds.), *Haptics: Neuroscience, devices, modeling, and applications: 9th international conference, eurohaptics 2014, versailles, france, june 24-26, 2014, proceedings, part i* (pp. 232–240). Berlin, Heidelberg: Springer Berlin Heidelberg. Retrieved from https://doi.org/10.1007/978-3-662-44193-0_30

Hwang, I., Seo, J., Kim, M., & Choi, S. (2013, July). Vibrotactile perceived intensity for mobile devices as a function of direction, amplitude, and frequency. *IEEE Transactions on Haptics*, *6*(3), 352-362.

Jensen, K. (1999). Envelope model of isolated musical sounds. In *Proceedings of the 2nd cost g-6 workshop on digital audio effects (dafx99)* (pp. W99-1–W99-4).

Karam, M., Russo, F. A., & Fels, D. I. (2009, July). Designing the model human cochlea: An ambient crossmodal audio-tactile display. *IEEE Transactions on Haptics*, *2*(3), 160-169.

Lofvenberg, J., & Johansson, R. (1984). Regional differences and interindividual variability in sensitivity to vibration in the glabrous skin of the human hand. *Brain Research*, *301*(1), 65 - 72. Retrieved from http://www.sciencedirect.com/science/article/pii/0006899384904037

Madsen, C. K. (1997). Focus of attention and aesthetic response. *Journal of Research in Music Education*, *45*(1), 80–89.

Madsen, C. K., Brittin, R. V., & Capperella-Sheldon, D. A. (1993). An empirical method for measuring the aesthetic experience to music. *Journal of Research in Music Education*, *41*(1), 57–69.

Madsen, C. K., & Geringer, J. M. (2000). A focus of attention model for meaningful listening. *Bulletin of the Council for Research in Music Education*(147), 103-108. Retrieved from http://www.jstor.org/stable/40319396

Merchel, S., Altinsoy, E., & Stamm, M. (2010). Tactile music instrument recognition for audio mixers. In *Audio engineering society convention 128.*

Meuer, E. M., Kern, E. A., Andrews, M., Tenhoff, A., Andrews, K., Huschka, P., … Thomas, A. P. (2017). Board# 150: Maker: Painting pitches. In *2017 asee annual conference & exposition.*

Nanayakkara, S., Taylor, E., Wyse, L., & Ong, S. H. (2009). An enhanced musical experience for the deaf: design and evaluation of a music display and a haptic chair. In *Proceedings of*

the sigchi conference on human factors in computing systems (pp. 337–346).
Pongrac, H. (2006). Vibrotactile perception: Differential effects of frequency, amplitude, and acceleration. In *2006 ieee international workshop on haptic audio visual environments and their applications (have 2006)* (p. 54-59).
Ragert, M., Fairhurst, M. T., & Keller, P. E. (2014, 01). Segregation and integration of auditory streams when listening to multi-part music. *PLOS ONE*, *9*(1), 1-9. Retrieved from `https://doi.org/10.1371/journal.pone.0084085`
Roberts, M., & Russo, R. (2014). *A student's guide to analysis of variance.* Routledge.
Rodet, X., Lambert, J.-P., Cahen, R., Gaudy, T., Guedy, F., Gosselin, F., & Mobuchon, P. (2005). Study of haptic and visual interaction for sound and music control in the phase project. In *Proceedings of the 2005 conference on new interfaces for musical expression* (pp. 109–114). Singapore, Singapore: National University of Singapore. Retrieved from `http://dl.acm.org/citation.cfm?id=1085939.1085971`
Shelton, J., & Kumar, G. P. (2010). Comparison between auditory and visual simple reaction times. *Neuroscience & Medicine*, *1*(1), 30–32.
Spence, C. (2009). Explaining the colavita visual dominance effect. *Progress in brain research*, *176*, 245–258.
Sternberg, S. (1969). Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, *57*(4), 421-457. Retrieved from `http://www.jstor.org/stable/27828738`
Stone, A., Shiffman, S., Atienza, A., & Nebeling, L. (2007). *The science of real-time data capture: Self-reports in health research.* Oxford University Press.
Tanabe, K., Takahashi, A., Hoshino, K., Ogawa, D., Hachisu, T., & Kajimoto, H. (2016). Haptone: Haptic instrument for enriched musical play (ii)?system detail. In *International asiahaptics conference* (pp. 461–465).
Tavakol, M., & Dennick, R. (2011). Making sense of cronbach's alpha. *International journal of medical education*, *2*, 53.
Tippach, M. (2003). *Asio4all - universal asio driver for wdm audio.* Retrieved from `http://www.asio4all.org/`
Verrillo, R. T. (1992). Vibration sensation in humans. *Music Perception: An Interdisciplinary Journal*, *9*(3), 281–302. Retrieved from `http://mp.ucpress.edu/content/9/3/281`
Wolfe, J. M., Kluender, K. R., Levi, D. M., Bartoshuk, L. M., Herz, R. S., Klatzky, R. L., ... Merfeld, D. M. (2006). *Sensation & perception.* Sinauer Sunderland, MA.
Wright, J. K., & Bregman, A. S. (1987). Auditory stream segregation and the control of dissonance in polyphonic music. *Contemporary Music Review*, *2*(1), 63-92.
Yamazaki, Y., Mitake, H., & Hasegawa, S. (2016). Tension-based wearable vibroacoustic device for music appreciation. In *International conference on human haptic sensing and touch enabled computer applications* (pp. 273–283).