THE IMPACT OF TRANSPOSITION SKILLS ON INHIBITORY CONTROL PERFORMANCE

Álvaro M. Chang-Arana
Master's Thesis
Music, Mind & Technology
Department of Music, Art and Culture Studies
4 June 2018
University of Jyväskylä

JYVÄSKYLÄN YLIOPISTO

Tiedekunta – Faculty Humanities	Laitos – Department Department of Music, Art and Culture Studies
Tekijä – Author Álvaro Mario Chang Arana	
Työn nimi – Title The impact of transposition skills on inhibitory control performance	
Oppiaine – Subject Music, Mind & Technology	Työn laji – Level Master's Thesis
Aika – Month and year April, 2018	Sivumäärä – Number of pages 96

Tiivistelmä – Abstract

This study tests whether existing transposition skills have an impact upon inhibitory control. Differences in the degree of transposition practice could translate in different cognitive functionality of musicians and reveal further unexplored evidence of inhibitory control's plasticity. A total of 64 participants were divided into a group of musicians (n = 34) and a group of non-musicians (n = 30). A transposition task, a music Stroop task and a classic Stroop task were designed. Musicians played their main instruments to play-as-written or transposed conditions from which transposition levels were calculated. All participants responded to the music and the classic Stroop task. The former required participants to choose the note written name while ignoring its location on the staff. Notation system, note-naming system and familiarity with a specific clef was ensured for every music participant. The latter consisted on a motor adaptation of the classic Stroop task where the written word had to be ignored and instead of that pick the perceived colour. Accuracy, reaction time and a composite score was calculated for both tasks. Stroop tasks succeeded in eliciting an inhibitory control response. However no inhibitory control performance differences according to the transposition skill level were detected.

Asiasanat – Keywords

Inhibitory control, musical Stroop-like task, Stroop task, transposition

Säilytyspaikka – Depository	Department of Music, Art and Culture Studies
Muita tietoja – Additional information	

A Pablo Sabat.

Gracias por todo lo aprendido.

Acknowledgements

Perhaps the biggest lesson I learned from the Music, Mind and Technology programme is that big endeavours rarely are accomplished in solitude. Thus, seeing this thesis finished would not have been possible without the support of different people who contributed to its completion in different ways.

I would like to thank Geoff Luck for his valuable support as supervisor. His guidance, patience and warmth created the ideal learning conditions for the successful completion of this project. Martin Hartmann assistance was crucial for the audio data processing stage. Similarly, I would like to thank Kendra Oudyk who helped me sorting out methodological challenges through her illustrative and supportive support. My gratitude extends as well to the programme's teaching staff and classmates for their active listening, questioning and help.

This outcome is also due to two important figures: Jorge Arana, my grandfather, and Pablo Sabat, my piano teacher. If it were not for their guidance, I would not have come to understand and love music as I do now. The former introduced me to the classical music world, while the latter expanded my musical, cultural and symbolic personal universe.

Finally, the greatest of my gratitude is reserved for my parents. Their unconditional love made it possible for me to cross long distances in pursue of my personal development. Although insufficient, I hope this thesis can be a fair gesture of appreciation and love to them.

CONTENTS

1 INTRODUCTION

This study investigates whether transposition skills have an impact on inhibitory control. Inhibitory control "involves being able to control one's attention, behaviour, thoughts, and/or emotions to override a strong internal predisposition or external lure, and instead do what's more appropriate or needed" (Diamond, 2013, p. 137). Together with working memory and cognitive flexibility, they constitute executive functions (Diamond, 2013; Hofmann, Schmeichel, & Baddeley, 2012; Slevc, Davey, Buschkuehl, & Jaeggi, 2016). Executive functions refers to the set of the three previously mentioned mental processes associated with prefrontal lobe activity (Ardila, Pineda, & Rosselli, 2000) and needed when storing and manipulating information in our mind; inhibiting impulses and automatic responses as well as acting appropriately to the expected task; and adapting to unpredictable changes in the environment (Davidson, Amso, Anderson, & Diamond, 2006).

Why are executive functions relevant in contemporary psychological research? They have been shown to be an important predictor for adapting to virtually any life aspect, including mental and physical health, academic and professional success, interpersonal relationships and public safety (Diamond, 2013). For instance, in a nearly three decades longitudinal study, Moffit et al. (2011) found that self-control could predict young adults' health, wealth and public safety, even when controlling for the effect of other relevant variables such as intelligence, socioeconomic origin and mistakes done during adolescence such as dropping school, smoking or early parenthood.

But executive functions have also been object of interest in cognitive and neuropsychological research, especially when studying musicians (Slevc et al., 2016). The relevance of selecting musicians as research subjects lies in the fact that they provide a privileged window for exploring neuroplasticity effects through neuroimaging or behavioural methods (Herholz & Zatorre, 2012). Practice effects on executive functions are studied by comparing musicians to non-musicians, other musicians, bilinguals or multilinguals. However, it seems that music transposition has not been studied through the lens of executive function literature yet.

According to the new Grove dictionary of music and musicians (1980), transposition is defined as "the notation or performance of music at a pitch different from that in which it was originally conceived, by raising or lowering the notes in it by the same interval" (p. 121). Transposition can also be related to transposing instruments. A transposing instrument "produces pitches that sound different from what is notated in the score" (Adler, 2002, p. 167). In other words, there is an incongruence between the visual stimuli and the auditory output of it.

Although clearly defined in musicological terms, it seems that there is a gap of knowledge on transposition from a psychological perspective. Moreover, it seems that it could be possible to build a rationale for the claim that transposition could be intimately related with inhibitory control. Since transposition is a musical process more usual for some musicians than others could we identify changes in inhibitory control among musicians based on their transposition skills? In this study inhibitory control literature and techniques will be used to understand the psychological functioning of transposition.

A final warning before concluding this introduction is required. Executive functions are important in several life aspects such as the academic. Consequently, it is common to find examples of studies that have explored the relationship with intelligence (Ardila, Pineda, & Rosselli, 2000; Thorell, Lindqvist, Nutley, Bohlin, & Klingberg, 2009), or its transference effects to other cognitive abilities or different skills (Schellenberg, 2005; Moreno & Farzan, 2015; Schellenberg, 2011). Unfortunately, these research topics have proven to be easily misinterpreted by scientifically uninformed users, often giving rise to myths such as the "Mozart effect" (Pietschnig, Voracek, & Formann, 2010; Rauscher, Shaw, & Ky, 1993); or stating that playing a musical instrument increases people's intelligence, despite inconclusive long-term evidences (Costa-Giomi, 2014). The inaccuracies derived from this literature reveal an implicit bias against music education, as if its value would rely on extra-musical effects (Rauscher, 2009; Schellenberg, 2006a).

Therefore, it must be clearly stated that the motivation of this study is focused on exploring the psychological functioning of transposition through the lens of inhibitory control. Thus, research outcomes will be limited to this objective and clearly delimited within the studied sample. Any overgeneralization of the outcomes of this research must be rejected.

2 TRANSPOSITION AND TRANSPOSING INSTRUMENTS

According to the new Grove dictionary of music and musicians (1980), transposition is defined as “the notation or performance of music at a pitch different from that in which it was originally conceived, by raising or lowering the notes in it by the same interval” (p. 121). For example, a short melody composed of the notes C-G-A-G consist of the following intervallic relationship: perfect fourth, major second, and major second. If the melody would be transposed to a different tuning such as “A”, then every “C” in the score will turn into an A tone. The distance between C and A is a minor third below C. For transposition to occur the same intervallic distance needs to be kept for the rest of the notes. In other words, C-G-A-G will have to be lowered a minor third into: A-E-F#-E. Thus preserving the intervallic distances between the melody notes (i.e. the distances of perfect fourth, major second, and major second are still kept).

Figure 1. Example of a transposed melodic passage. Left score depicts a non-transposed melody (or an instrument in C) and right score depicts the actual tones that would be heard if it would have been transposed into A.

By extension, “transposing instruments produces pitches that sound different from what is notated in the score” (Adler, 2002, p. 167). The list of transposing instruments, according to Adler (2002, p. 169) and the new Grove dictionary of music and musicians (1980, p. 118) includes examples such as:

- B♭ clarinet and B♭ soprano saxophone (a major 2nd below written pitch)
- A clarinet (a minor 3rd below the written pitch)
- E♭ sopranino saxophone and E♭ clarinet (a minor 3rd above the written pitch)
- D clarinet (a major 2nd above the written pitch)
- F English horn and F basset horn (a 5th below the written pitch)
- G alto flute (a 4th below the written pitch)
- E♭ alto clarinet and E♭ alto saxophone (a major 6th below the written pitch)
- B♭ tenor saxophone and B♭ bass clarinet (a major 9th below the written pitch)
- E♭ baritone saxophone (a major 13th below the written pitch)

- B♭ contrabass clarinet and B♭ bass saxophone (two octaves and a major 2nd below the written pitch)
- Piccolo (an octave above the written pitch)
- Contrabassoon and double bass (an octave below the written pitch)
- Violino piccolo (a 4th over the written pitch)
- F French horn (a 5th below the written pitch)
- B♭ trumpet (a major 2nd below the written pitch)

As it will be explained in the Methodology section, a transposition task was developed for this study in which certain notes were asked to be transposed into higher intervals. Therefore, it was important to take into account the transposing characteristics and the playing range of each instrument to ensure that participants could perform this study's transposition task (see Section 6.4). For a full description of the main transposing instruments considered in this research, see Appendix A. Having explained what transposition is, the next section will elaborate on executive functions: the psychological context upon which this study is framed.

3 EXECUTIVE FUNCTIONS

Executive functions (EF) refer to a set of three mental processes needed when storing and manipulating information mentally; inhibiting impulses and automatic responses as well as acting appropriately to the expected task; and adapting to unpredictable changes in the environment (Davidson, Amso, Anderson, & Diamond, 2006). These three processes associated with prefrontal lobe activity (Ardila, Pineda, & Rosselli, 2000) are known as working memory, inhibitory control and cognitive flexibility (Diamond, 2013; Hofmann, Schmeichel, & Baddeley, 2012; Slevc et al., 2016).

There is a general agreement over the existence of three main EF: working memory, inhibitory control and cognitive flexibility (Davidson et al., 2006; Diamond, 2013; Slevc et al., 2016). In order to set a general context of EF literature, a brief definition of working memory and cognitive flexibility will be presented. Because this study focuses exclusively in inhibitory control, a more elaborated description of this EF will be reserved for Section 4.

3.1 Working memory

According to Diamond (2013), working memory is an EF "which involves holding information in mind and mentally working with it" (p. 142). This definition is based on Baddeley and Hitch's working memory model (1994). According to them, working memory is made of three distinct systems: a phonological loop which processes verbal information; a visuospatial sketchpad which processes nonverbal information; and a central executive which coordinates the functioning of the two other components. The meaning of working memory within the EF context adds an emphasis on its updating function, meaning "the ability to continuously monitor information and to rapidly add and remove information from working memory" (Slevc et al., 2016, p. 199).

3.2 Cognitive flexibility

This EF appears later in development and feeds on the functioning of working memory and inhibitory control (Diamond, 2013). This function allows to engage in task-switching, meaning "the ability to shift back and forth between multiple tasks or mental sets" (Hofmann et al., 2012,

p. 174). Cognitive flexibility allows a change in perspective both on a spatial (imagining seeing something from another location) and interpersonal sense (being able to understand another person's point of view), and come up with different and creative solutions for everyday life problems (Diamond, 2013).

4 INHIBITORY CONTROL

Having described briefly working memory and cognitive flexibility, it is time to complete the picture of EF by delving into the main focus of this study: inhibitory control. According to Diamond (2013) inhibitory control "involves being able to control one's attention, behaviour, thoughts, and/or emotions to override a strong internal predisposition or external lure, and instead do what's more appropriate or needed" (p. 137). Banich and Depue (2015) also coincides with Diamond (2013) concerning the override function, but adds that to inhibit is to "interrupt, or abort ongoing processes, especially when those processes are well engrained" (p. 17). Taken together the core functions of inhibitory control are to refrain an automatized and salient behaviour and change it towards a required one.

Inhibitory control has been also studied at a neural level. Associated inhibitory functions have been shown to be disrupted in psychiatric disorders such as ADHD and substance abuse disorders (Diamond, 2013; Hofmann et al., 2012; Mullane, Corkum, Klein, & McLaughlin, 2009; Moffitt et al., 2011). These associated areas involve lateral regions of the right prefrontal cortex although the exact computational function of this region is still unknown. The prefrontal cortex is a heterogeneous structure with distinguishable functional organization that control this top-down cognitive control system. Although much research has been done on the left ventrolateral prefrontal cortex (VLPFC), much less is known about the functions of its right counterpart (Levy & Wagner, 2011). According to Levy and Wagner (2011) the specific function of the right VLPFC is unclear and there are two main hypotheses: either it is crucial for motor inhibition, or this region is a subcomponent of a broader functional system (including temporoparietal areas) designed to reorient attention after abrupt changes in the environment. Functionally, it has been showed that both hypotheses share a great similarity in terms of the regions recruited (Levy & Wagner, 2011). Nevertheless, specific activation patterns of sub regions in the right VLPFC have been identified and are summarized from Levy and Wagner (2011):

- Inferior frontal junction: detects salient stimuli in the environment, generating signals every time a matching signal appears in the environment. It might be related in the detection of infrequent stimuli as measured by Go/No-Go tasks. This area has also been reported to be activated during Stroop tasks, task switching, and verbal n-back tasks (Cramon, Brass, & Derrfuss, 2004).

- Right posterior-VLPFC: involved in stopping motor tasks, playing an important role in motor inhibition and refreshing the execution of action plans.
- Right mid-VLPFC (pars opercularis): this region is activated when there are equal chances that go and no-go tasks are likely to occur. In other words, subjects cannot plan any differential response.
- Right anterior insula (AI)/frontal operculum (FO): tasks involved in refocusing attention seems to involve the AI; while motor inhibition elicited by stopping tasks involved the AI and the FO. In other words, AI is involved both in reorienting tasks and motor inhibition tasks.

So far, the definitions of inhibitory control presented at the beginning of this section have been supported by neuroscientific evidence. This evidence has showed the complex and modular nature of inhibitory control. To complete its conceptualization, a description of the most common behavioural measurements of inhibitory control will be presented.

4.1 Inhibitory control tasks

This section presents the most common inhibitory control tasks used in EF research according to Diamond (2013). The description of the different tasks will be valuable for complementing the understanding of what inhibitory control is and for justifying the methodology that was chosen in this study. As a guiding principle, all of these tasks require participants to “override, interrupt, or suppress an ongoing cognitive, emotional or behavioural response” (Banich & Depue, 2015, p. 17). Next, six of the most well-known inhibitory control task will be described.

4.1.1 Antisaccade task

In this motoric inhibition task (Luna, 2009), participants are presented with a centred target signal (e.g. a dot in a screen) which is immediately followed by a peripheral stimuli (e.g. dots of different colours). Motoric inhibition is elicited when participants are required to resist the pro-saccade eye movement or the reflex of staring into the appearing stimuli and instead perform an antisaccade task or look in the opposite direction of the appearing target (i.e. if a red dot appears at the right of the centred target signal, participants must look at the left side of the latter and vice versa; Hutton & Ettinger, 2006; Luna, 2009; Munoz & Everling, 2004).

4.1.2 Delay of gratification task

The delay of gratification task has had different set ups in its implementation (Casey et al., 2011). Nonetheless, it is characterized by measuring the amount of time that a participant can resist an immediate reward in order to receive a delayed and larger outcome (Casey et al., 2011; Mischel, Shoda, & Rodriguez, 1989). A paradigmatic setting of this task involves placing a child in front of a cake tin containing a marshmallow and a pretzel and the experimenter will ask the child which one of those two he would like to eat. After the child chose, the experimenter will say that he needs to leave the room and if the former waits until the latter returns, then the child will receive the desired sweet. Nevertheless, if the child does not wants to wait she or he can ring a bell and the experimenter will return immediately, but the child will receive the non-desired sweet (Mischel, Ebbesen, & Raskoff, 1972).

4.1.3 Flanker task

Flanker task demands activating focused attention (Luna, 2009) in order to ignore visual distractions that “prime different motor responses” (Cragg, 2016, p. 242). It was first designed by Eriksen and Eriksen (1974). Participants had to pay attention to a target letter which will always appear in the same location and ignore all other stimuli that could be showed simultaneously. If the presented letter was H/K, then subject had to press a lever either to the right or left and if letter was S/C then he or she had to press the lever into the opposite direction. In the original experimental design (Eriksen & Eriksen, 1974, p. 144), five noisy conditions were used:

1. Noise same as target: target was flanked by three identical letters to each side.
2. Noise response compatible: target letter was flanked by a three times repeated letter but which was compatible with the response set (i.e. if target was H, then flanking letters were three copies of K to each side).
3. Noise response incompatible: target letter was flanked by three copies of a letter belonging to the other set (i.e. if target stimuli is H, then flanking letters could be three copies of S/C to each side).
4. Noise heterogeneous-Similar: target stimuli is flanked by three different letters from the stimuli set (i.e. letters that resemble but are others than H/K) that have similar features.

5. Noise heterogeneous-Dissimilar: target stimuli flanked by dissimilar letters and excluding letters from the stimuli set (i.e. letters that does not resemble and are others than H/K).

These five conditions were grouped into three kind of displays: compatible (1 and 2), incompatible (3) and neutral (4 and 5) (Eriksen, 1995).

4.1.4 Go/No-Go task

The go/no-go task (Donders, 1969) is another cognitive task used in response inhibition research (Luna, 2009; Verbruggen & Logan, 2008). In a paradigmatic setting of this task, participants are required to execute a motor command every time they see a particular stimuli (i.e. press a red button every time they see an apple) and to refrain from responding every time they see another particular stimuli (i.e. do not press the red button if they see a watermelon). Normally, the frequency of the “go” task will be higher than the “no-go” ones, thus creating a trend to execute a motor command on every trial. This elicits the cognitive task of inhibiting the prepotent response and refrain action (Cragg & Nation, 2008; Rubia et al., 2001).

4.1.5 Simon task

In the Simon task (Simon & Rudell, 1967) participants listened to a series of 132 pre-recorded commands in which the word “left” or “right” were randomly announced through a headphone, either through its left or right speaker. Participants were placed in front of two telegraph keys, one at their right and another one at their left, and were instructed to press either of them according to the speakers’ announcement and, regardless of the direction from which the order came (i.e. if the right speaker said “left” participants still needed to press the left telegraph button). What Simon and Rudell (1967) found was that participants took more time to reply to a verbal command in which the auditory origin was inconsistent with the appropriate behavioural response (i.e. left speaker reproduce “right” and hence participants had to press the right button) than when the auditory origin of the command was consistent with the appropriate behavioural response (i.e. left speaker reproduce “left” and hence participants had to press the left button; Hommel, 2011).

4.1.6 Stroop task

In the seminal 1935 Stroop's paper, two research questions were raised. First, Stroop asked what is the time difference in interference time when comparing the "interfering effect of color stimuli upon reading names of colors… with the interfering effect of word stimuli upon naming colors themselves" (1935, p. 646). Second, Stroop asked "what effect would practice in reacting to the color stimuli in the presence of conflicting word stimuli have upon the reaction times in the two situations described in the first problem?" (p. 647). Out of these questions, three experiments were proposed.

The first experiment was called "the effect of interfering color stimuli upon reading names of colors serially" (Stroop, 1935, p. 647). For this experiment, a 10 x 10 stimuli matrix was designed. In it, 5 colours (i.e. red, blue, green, brown and purple) were printed following a series of specifications. Stroop made sure that the ink colour did not appeared twice in each column and row. Moreover, they should not succeed immediately in columns or rows. Meanwhile, word names should not repeat more than twice in each line. A list of colour names printed in black ink was created by duplicating the arrangement of the colourful list. Finally, each test had a second form by printing a reverse order of the stimuli. The colour list test (and its two formats) was called "reading color names where the color of the print and the word are different (RCNd)"; while the black list test (and its two formats) was called "reading color names printed in black (RCNb)".

Seventy undergraduates (14 male; 56 female) were recruited. They were divided by sex groups and each one was randomly assigned to two possible arrangements of test orders:

- RCNb (1st format)
- RCNd (2nd format)
- RCNd (1st format)
- RCNb (2nd format)

Or:

- RCNb (2nd format)
- RCNd (1st format)
- RCNd (2nd format)
- RCNb (1st format)

This arrange was chosen to counterbalance for practice and fatigue effects. Before the first reading of each test, participants were shown a 10 words sample list. They were asked to respond as quickly and accurate as possible, avoiding leaving any errors uncorrected. The starting signal was “Ready! Go!” (p. 648). Stroop reported a difference of 2.3 seconds between RCNb and RCNd tests, which represent an increase of 5.6% of the time spent to read the colour names in black ink. He and concluded that “this increase is not reliable” (p. 659).

The second experiment designed by Stroop was called “the effect of interfering word stimuli upon naming colors serially” (Stroop, 1935, p. 649). For this experiment, the RCNd tests were modified. A new test called “naming color test” (NC) was designed by printing the stimuli in RCNd in the same order, but in the form of solid squares. Also, the RCNd was used differently. In this experiment, participants would have to ignore the written colour names and name the colours’ ink serially. This test was called “naming color of word test where the color of the print and the word are different” (NCWd). Just as in Experiment 1, participants read two forms of each test in a single sitting.

A hundred students (29 male; 71 female, included undergraduate and graduate students) were recruited. They were divided by sex groups and each one was randomly assigned to two possible arrangements of test orders:

- NC
- NCWd
- NCWd
- NC

Or:

- NCWd
- NC
- NC
- NCWd

This arrange was chosen to counterbalance for practice and fatigue effects. Before the first reading of each test, participants were shown a 10 words sample list. They were asked to name the colours’ ink as quickly and accurate as possible and to correct all errors. The starting signal was the same as Experiment 1. Stroop reported a major increase of 47.0 seconds between NC and NCWd, which represents an increase in delay time of 74.3% of the time required to name colours' ink in printed squares. He attributed this remarkable difference in time to a difference in association strength between stimuli and responses: "the associations that have been formed between the word stimuli and the reading response are evidently more effective than those that have been formed between the color stimuli and the naming response" (pp. 659-660). In other words, he uncovered a strong interference effect for verbal habitual responses.

Finally, Experiment 3 was named "the effects of practice upon interference" (Stroop, 1935, p. 652). For this experiment, RCNb, RCNd, NC and NCWd were used. A modification was introduced in NC, where the solid squares were replaced by swastikas in order to approach a closer resemblance to printed words. Additionally, the order of presentation of stimuli was changed: every line still contained to repetitions of one colour name, but they were separated just by one other colour. The purpose of this change was to equate for difficulty level on every line of the task. Once again, there were two forms for these tests, in which the second form was an inverter order of the first.

Thirty two undergraduates (17 male; 15 female) were recruited. Stroop designed a 14 days training program in which at each training day participants had to read 4 half-sheets of a particular test. Stroop registered the average time and chose this value as the day's score. The training schedule was as follows:

Table 1

Experiment 3 planning

Day	Test
1	RCNb
2	RCNd
3	NC
4	NCWd
5	NCWd
6	NCWd
7	NCWd
8	NCWd
9	NCWd
10	NCWd
11	NCWd
12	NC
13	RCNd
14	RCNd

Note. Adapted from "Studies of Interference in Serial Verbal Reactions," by J. R. Stroop, 1935, *Journal of Experimental Psychology*, *18*(6), p. 653.

Stroop reported five effects after practicing the NCWd task for eight days. The first one was a mild decrease in the interference effect of reading the printed colour names over naming the ink. Second, a practice curve similar to other experimental reports was obtained. Thirdly, the group's variability was increased. Fourth, the reaction time in NC (where solid squares were used) was decreased. Fifth, the conflict between reading words when presented in conflicting ink colours increased.

Ever since its publication, the classic Stroop task has become a paradigmatic cognitive psychology experiment. It has been used to study automatization, emotional processing, neuroplasticity, inhibitory control, among other processes (MacLeod, 1991). This extended description of Stroop's classical study served a relevant purpose in this research. In the Methodology section, Stroop's set up will be used as a framework for the experimental design. Having presented the main inhibitory control tasks, the next session deals with previous EF studies with samples of musicians.

4.2 Music, executive functions and inhibitory control research

Musicians provide a privileged window for studying neuroplasticity either by neuroimaging or behavioural data (Herholz & Zatorre, 2012). This is because their extensive musical training over the course of the lifespan involves a multiplicity of cognitive, emotional and behavioural responses which are distinguishable from other non-musically trained populations. Thus, it is assumed that behavioural and neurological differences between populations of musicians and non-musicians will be attributed to musical training. The EF literature has not been oblivious to this and has compared musicians with samples of non-musicians, bilinguals or multilinguals (Kunert, Willems, Casasanto, Patel, & Hagoort, 2015; Patel, 2003; Patel, 2008). In this section, examples of EF research conducted with musicians across the lifespan will be presented. It aims at justifying the population selected for this study and to show that no previous EF research has studied music transposition.

Moreno, Wodniecka, Tays, Alain and Bialystok (2014) found subtle differences in EEG processing of inhibitory control task (Go/No-go task) over behaviour in a group of musicians and bilinguals. Musicians showed an earlier enhanced P2 and a reduced N2 signals, which were associated with early processing of stimuli-response representation (for instance, needed when performing fast and complicated musical tasks). Meanwhile, bilinguals had larger N2 and P3 signals, associated with later inhibition of conflicting information (for instance, needed when detecting a sounds from other language that they might not need at the moment). Hence, there are subtle differences between music and language despite the many similar characteristics.

In an earlier study, Moreno et al. (2011) conducted a 20 days program of music-listening training or visual arts training in children between 4-6 years old. After the intervention only the music training group resulted in an enhanced performance on a measure of verbal intelligence. Meanwhile the art group had no practical effects to report. This could be a matter of developmental processes: visuo-motor skills may develop further in life than auditory skills which are crucial for other uses such as language acquisition. Other behavioural changes in the music group were on a better performance of Go/No-Go task and an increase P2 which becomes consistent with Moreno et al. (2014) findings. So, how early can executive functions develop? Janus, Lee, Moreno and Bialystok (2016) develop a 20 days training program with children (4 and 6 years old) who were assigned either to French or music conditions. Interestingly, both showed an improvement in tasks of executive control; although not even after training did they differ in tasks such as receptive vocabulary, Raven test, or Corsi Blocks.

It seems clear that EF can be developed from early in life. What about with older children? Joret, Germeys and Gidron (2016) focused on cognitive inhibitory control among children between 9-12 years old. The music group presented a greater resistance to interference when compared with a non-musical control group. This differences could be explained by the attention demanded by musical education as well as the need to ignore other students if they were surrounded by other students during the lessons. This differences in performance are related with cerebral activity.

Zuk, Benjamin, Kenyon and Gaab (2014) found differences in EF measures between adult musicians and children with musical training. Children with musical training also showed greater cerebral activity when compared with non-musician peers. Particularly the SMA and the right VLPFC. Interestingly they did not find differences in inhibitory control measures between the group of children, perhaps as a consequence of sample size or other subject selection criteria.

Differences in EF can be spotted also in adulthood. Bialystok and DePape (2009) found that intense musical training enhances performance in executive functioning. Three groups were compared: musicians (but not bilinguals), bilinguals (but not musicians) and control group (none of the previous). Both musicians and bilinguals outperformed the control group on conflict auditory (auditory Stoop) and spatial tasks (Simon arrows), but, interestingly, musicians were better at the auditory task than the bilingual group. The authors conclude that although generalization of increased performance to other domains is possible, the greatest effects will be obtained in the tasks which are closer to the core experience. Hence, there are domain-general effects but also domain-specific effects.

Finally, there is some evidence that EF can still be developed in late adulthood and contribute to reduce cognitive decay. Bugos, Perlsetin, McCrae, Brophy and Bedenbaugh (2007) found that individualized piano instruction in a group of older adults resulted in an increase in measures related to EF. Particularly, there was an increase cognitive abilities related to working memory, such as concentration and attention. Even if results cannot be overgeneralized, it shows some suggestive evidence towards plasticity of executive functions in late life.

In sum, evidence presented in this section points to the fact that musicians are an ideal study population for EF research across different age groups and types of populations. Although, interestingly none of the studies cited above had paid closer attention to more subtle differences within musical skills. That is, no previous EF research has targeted music transposition as a study object. However, before diving deeper, a thorough description of different music Stroop tasks needs to be developed. As it will be clear in the Methodology section, a music Stroop task will be designed following the strengths and limitations from previous similar experiences.

4.3 The music Stroop task

In this section four research papers in which a music Stroop task was used will be discussed (Akiva-Kabiri & Henik, 2012; Grégoire, Perruchet, & Poulin-Charronnat, 2013; Stewart, Walsh, & Frith, 2004; Zakay & Glicksohn, 1985). A detailed summary of every methodology will be presented chronologically. This section finalizes with a comparison between the methodologies followed in each study. The rationale for this extended description is to identify the main strengths and limitations of every study. These will inform the design of the music Stroop task described in the Methodology section.

4.3.1 Zakay and Glicksohn (1985)

This study explored the interaction between stimulus-response compatibility (SRC) and the amount of congruence between task-relevant and task-irrelevant dimensions of the stimulus (CRN) and how this interaction influences the Stroop effect was investigated. Twenty 20 pianists (14 females, age range: 20-26, *M* age of piano initiation: 6 years old) were recruited for their study. Four music Stroop-like tasks were presented in a paper: word (W), note (N), note-word (NW) and word-note (WN); which were based on the four classic Stroop tasks (W, colour [C], CW, WC). These tasks were to be completed by using three experimental stimuli which are reproduced in Figure 2. Each stimuli consisted of 10 notes randomly selected from 2 octaves of the keyboard.

In the upper staff (condition A), 10 written names of notes were printed in corresponding spatial position; in the middle staff (condition B), 10 note symbols were printed in their corresponding spatial location. But on the lower staff (condition C), 10 written note names were incongruent with their staff location. Hence, conditions A and B (comprising W and N tasks) were labelled as “congruent” and condition C (comprising NW and WN tasks) as “incongruent”.

Two respond conditions were also included: one verbal and another motor. On the one hand, in the verbal W condition participants had to read aloud the 10 written note names; while in the N condition, the 10 note symbols had also to be read aloud. On the other hand, in the NW subjects had to read aloud the 10 notes symbols according to their position on the staff while ignoring the printed name (e.g. if “Re” was printed in the “G” line, then the participant had to say “Sol (G)” and not “Re (D)”); while in the WN condition, the same 10 written note names had to be read while ignoring their position in the staff. For the motoric condition, participants received the same instructions with the only difference that the response had to be done by pressing the adequate keys on a piano keyboard. Given that there were two respond conditions, two series of three experimental stimuli were created for the four tasks. In total, this means that it was a 2x2x4 stimuli design. The presentation of the eight tasks was randomized and participants were measured on their speed and accuracy. Researchers calculated two sets of results: response times and number of error between the eight conditions; and repeated measures ANOVA.

Zakay and Glicksohn (1985) reported a similar effect as the one obtained in the classic Stroop task. In their study, the most difficult condition was NW, while W was the easiest one. They concluded that when SRC and CRN were high, response impairment was the lowest; while when SRC and CRN were low, response impairment was the highest.

Figure 2. Example of the music Stroop-like task designed by Zakay and Glicksohn (1985). Upper staff depicts W task, middle staff N task and lower staff shows stimuli used either in NW or WN task. Reproduced and adapted from Zakay, D., & Glicksohn, J. (1985). Stimulus congruity and S-R compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology/Revue Canadienne De Psychologie, 39*(3), 414-423. doi: 10.1037/h0080069

4.3.2 Stewart, Walsh and Frith (2004)

Two research questions were raised: first, is the execution speed of pianist on a sequence of number to finger mapping affected significantly by reading irrelevant musical notation? Second, what is the nature of the representation of musical notation and can it be generalized outside a musical context? For answering them, two experiments were conducted.

In Experiment 1, two group of participants were formed. Group 1 was made of 12 professional piano students (12 female, *M* age = 26 years old, *M* age of piano experience = 20 years). Meanwhile, Group 2 included 14 non-musicians without any experience in music reading or playing (10 female, *M* age = 22 years old). All participants were right-handed.

In their music Stroop task five tasks were designed. In four stimuli, numbers ranging from 1 to 5 (each one referring to fingers in the right hand; e.g. 1 = thumb, 2 = index, etc.) were superimposed on musical notes symbols. Only the baseline task differed. On it, a black strip background line on which five numbers printed in white were located. Participants were required to execute a series of keypresses by mapping from the numbers to the respective fingers. Each musical stimuli showed a motor sequence of five notes which ranged from G4 to D4. An example of this set of tasks can be seen in figure 3.

The experiment consisted of five conditions: baseline, congruent, incongruent (random), incongruent (systemic), and catch. As explained above, the baseline condition consisted of a row of five white numbers against a black strip background. In the congruent condition, there was a correspondence between the depicted notes and their spatial finger allocation: "notes extending from the bottom to the top of the staff map respectively onto digits extending from the left to the right hand" (p. 184). In the incongruent random condition there was an inconsistency between notes and numbers among all stimuli (e.g. G-3, A-1, etc.). This inconsistency was determined following a systematic method which ensured combining each of the five notes only once. In the incongruent (systemic) condition, the correspondent relationship between numbers and notes was inverted (e.g. G-5, A-4, B-3, etc.), thus obtaining a "musically incongruent but spatially systemic" (p. 184) condition. Lastly, in the catch condition the first three notes were congruent and the las two incongruent random. This was used in order to prevent participants from using note-reading strategies on congruent trials.

Participants responded to the stimuli using a computer keyboard in order to control for a facilitated response in the musicians group. They were instructed of ignoring the musical notation and use only the number to perform the trial. Previous to the actual experiment, participants went through five practice sessions. Before each experimental trial, a middle fixation point was shown for a second, after which a stimuli was presented until a response was executed or up to 3 seconds for the pianists' group or 4 seconds for the non-musicians' group. After a response or the time limit, another middle fixation point was showed, thus repeating the procedure in this way. In total, participants had to respond to two trials of 12 motor sequences in five possible conditions, adding up a total of 120 trials per participant. The 120 trials were divided in 12 blocks of 10 stimuli. Thus, the stimuli dimension for this experiment was 5x12x2. Motor sequences and trial types were pseudorandomly presented. Errors and response time were calculated. Response times of key presses comprised two different calculations: cumulative analysis and itemized analysis. The statistical test selected for testing their hypotheses was a mixed-design repeated measures ANOVA.

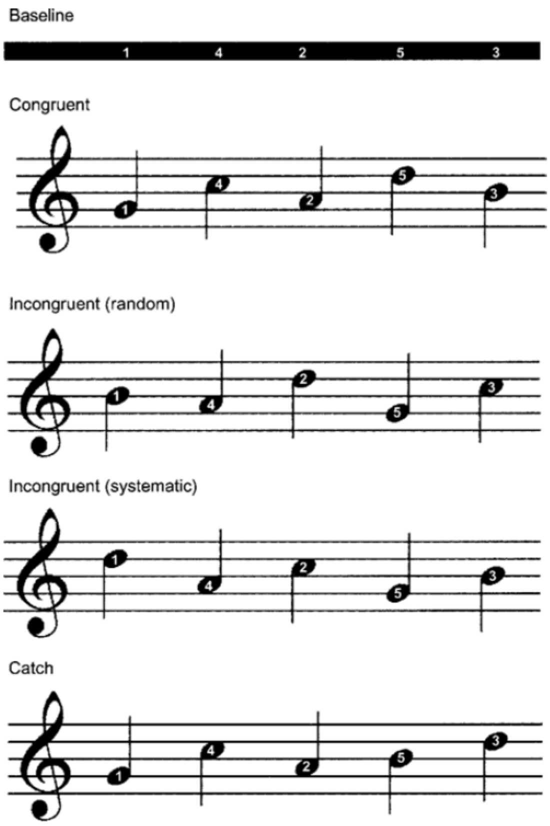

Figure 3. Example of the music Stroop-like task designed by Stewart et al. (2004). All motor sequences had five numbers which (for the musical stimuli) could be musically-spatially correspondent or not. Adapted from “Reading Music Modifies Spatial Mapping in Pianists” by L. Stewart, V. Walsh, and U. Frith, 2004, *Perception & Psychophysics, 66*(2), p. 185. Copyright 2004 by the Psychonomic Society, Inc.

Experiment 2 recruited the same group of participants who took part in Experiment 1. One group of 8 pianists and another one of 14 non-musicians responded to a non-musical analogue of the music Stroop task. Two tasks were designed: horizontal-to-horizontal stimulus-response task and vertical-to-horizontal stimulus-response task. For each of the tasks, three stimuli were created. In turn, these three stimuli were used for three conditions: baseline, congruent and incongruent (systematic). An example of both tasks are presented in figures 4 and 5.
According to Stewart et al. (2004), the rationale behind the horizontal-to-horizontal task is that we will have a fast reaction when presented with a stimulus-response task which is congruent or “overlap on some physical or representational dimension” (p. 190). For instance, in Simon tasks, participants will react faster when a right-side-presented stimuli needs to be responded with a right-handed response, but not if there is a mismatched between stimuli location and

required motor response. Here, there is an overlearned "horizontal meridian" response. However, in the vertical-to-horizontal stimulus-response task there is a crucial difference. For non-musicians participants, there is no learned correspondence between appearing stimuli in the vertical axis with responses in the horizontal one. For Stewart et al. (2004), pianists would "be characterized by a set of vertical-to-horizontal stimuli-response mappings" (p. 190).
In the horizontal-to-horizontal task, stimuli were shown in a parallel way as in Experiment 1. When presented with the congruent condition, in the leftmost position a "1" appeared, and in the rightmost position a "5" was shown. In the incongruent (systematic) condition, the number presentation was reversed (e.g. leftmost position was matched with "5").

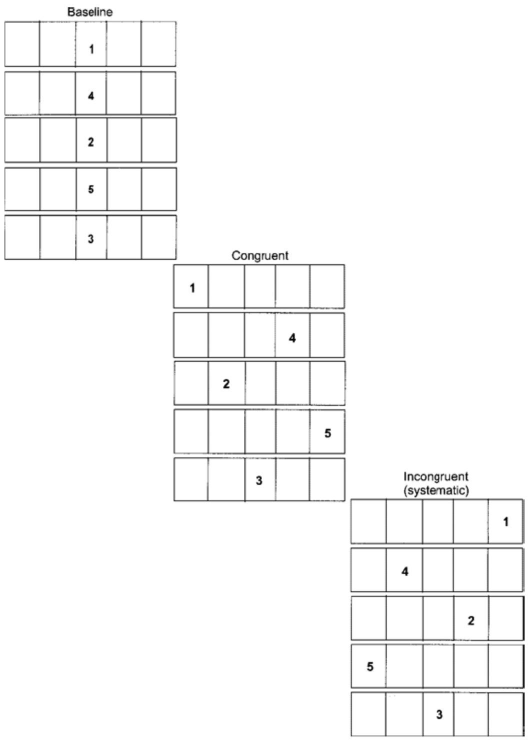

Figure 4. Example of horizontal-to-horizontal task designed by Stewart et al. (2004). The three conditions used are presented. Reproduced and adapted from Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception & Psychophysics, 66*(2), 183-195. doi: 10.3758/BF03194871.

Meanwhile, in the vertical-to-horizontal task, stimuli were inspired by the way in which pianists read note in the scores. When facing a congruent condition, the lowest position showed a "1", and the highest position a "5". Contrary, in the incongruent condition, the lowest position showed a "5" and the highest a "1". Finally, in the baseline condition for both horizontal-to-horizontal and vertical-to-horizontal task, all numbers appeared in the middle box. Again, participants had to press the correct key while ignoring their position in the horizontal or vertical boxes. Contrary to Experiment 1, every time a participant pressed a key, it triggered the appearance of the next one.

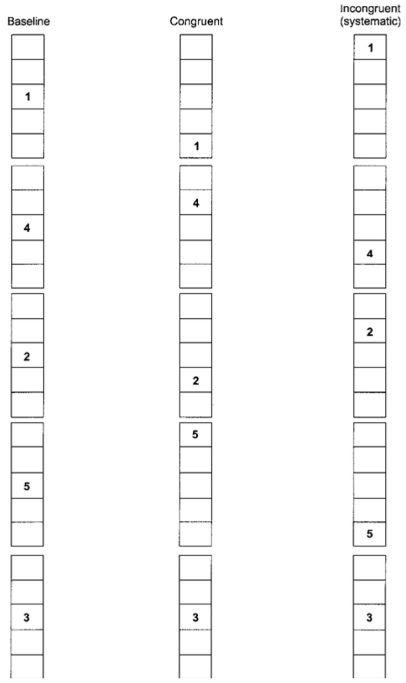

Figure 6. Example of vertical-to-horizontal task designed by Stewart et al. (2004). The three conditions used are presented. Reproduced and adapted from Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception & Psychophysics, 66*(2), 183-195. doi: 10.3758/BF03194871.

Participants responded to the stimuli using a computer keyboard in order to control for a facilitated response in the musicians group. They were instructed to ignore the horizontal or vertical position of the numbers and use them only to perform the trial. Each number appeared one at the time and the next one was shown as soon as the participant pressed a key (which made the previous one to disappear). After completing five stimuli, a 1 second pause was made before the appearance of the next series. In total, participants had to respond to two kind of tasks, three trial types (baseline, congruent and incongruent) and on two groups of participants. Trials were pseudorandomly presented. Errors and response time were calculated. Response times of key presses comprised two different calculations: cumulative analysis and itemized analysis. The statistical test selected for testing their hypotheses was a mixed-design repeated measures ANOVA.

This research presented two relevant results. The first one was that irrelevant notated music can produce an interference effect on a group pianists who are asked to perform a motor sequence in which printed numbers are transformed into a sequence of keypresses. Meanwhile, the second result showed that pianists have a vertical-to-horizontal representation of space when asked to perform some stimulus-response mapping tasks, which can be transferred outside of the musical context.

4.3.3 Akiva-Kabiri and Henik (2012)

This study assessed the interference in musicians with absolute pitch (AP) and relative pitch (RP) through a Stroop task-like stimuli. They hypothesised that tone naming is an automatic process in AP possessors and their performance will differ from RP possessors. Sixteen participants were equally divided into an AP group and a RP group. Auditory and visual stimuli tasks were designed. For the auditory stimuli, a 1 second sound produced from a piano synthesizer (ranging from C4-B4) was recorded. In the auditory neutral condition, 1 second of white noise was used. Meanwhile, in the visual stimuli two tasks were designed: “musical notes and written words” (Akiva-Kabiri & Henki, 2012, p. 274). Each one had an experimental stimuli and control stimuli condition design. For the musical note tasks, one quarter note out of seven possible was showed in a treble clef. In the equivalent neutral condition, an empty staff was presented. Whereas in the written words tasks note names were presented written without a staff. The corresponding neutral condition showed an “XXX” stimuli.

Participants were presented with two tasks: tone naming and note naming. In tone naming, they were required “to respond to an auditory tone and ignore the visual note or word” (Alkiva-Kabiri & Henki, 2012, p. 272); while in note naming, participants were “asked to respond to the visual note and ignore the auditory tone” (Akiva-Kabiri & Henik, 2012, p. 272). During the experiment, both set of stimuli (auditory and visual stimuli) were presented simultaneously. The combined presentation of the auditory and visual stimuli could be set in congruent, incongruent or neutral arranges. When congruent, the musical notation corresponded with the reproduced sound; when incongruent, the notated stimuli was different from the audible sound; and, lastly, in the tone naming task a pitch note was paired with either a blank staff or “XXX” and in the note naming task a visual notation was paired with white nose. Participants were instructed to respond as accurately and quickly as possible while ignoring a particular irrelevant dimension. The experiment started after going through a practice block made of 21 practice trials. Subjects were showed with a white screen for 1.5 seconds, after which the simultaneous presentation of the visual and auditory stimuli was projected for a second. The visual stimuli, though, remained in the screen for up to 3 seconds. Response format was auditory: participants gave their answer on a microphone and researchers registered the response time by calculating the difference between the onset of the stimuli and the onset of the participant's reply (see figure 6). For this experiment, 42 incongruent trials, 42 incongruent trials and 42 neutral trials were designed; giving a total of a block of 126 trials. The 126 block were repeated twice, since the note conditions consisted of musical symbols or the written name of musical notes.

Given that there were two independent variables (AP vs. RP), two tasks (tone naming or note naming), two blocks of trials (musical notation or written musical names of notes), and three combination of stimuli (congruent, incongruent or neutral), this experiment is a 2x2x2x3 factorial design. The presentation of both the tone naming task and note naming task as well as the two 126-blocks were counterbalanced. Researchers calculated two sets of results: reaction times and error rate. A mixed-design four way ANOVA was used to contrast their hypothesis.

Two main findings were obtained: AP participants could not repress automatically labelling tones, even if their recognition was irrelevant for the task; and RP found this very same task demanding on its processing difficulty. Since AP possessors cannot refrain themselves from labelling pitch, it is important to control for this variable in this research since it is focused on interference due to acquired skills by means of practice and not due to an inherited characteristic.

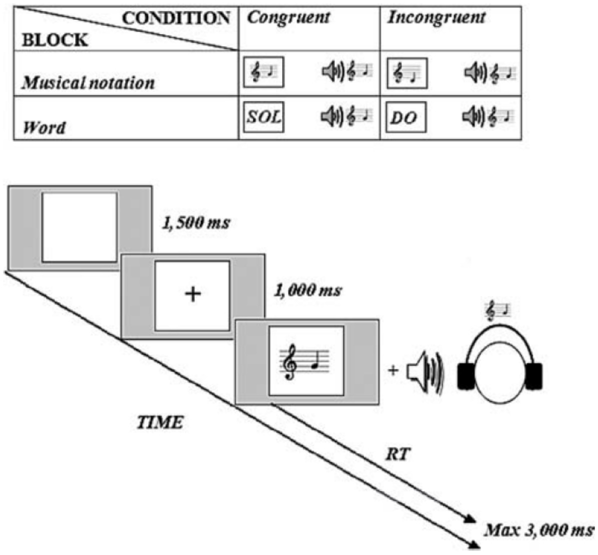

Figure 6. Experimental design of auditory music Stroop task. Auditory and visual stimuli could either be congruent or incongruent. Reproduced and adapted from Akiva-Kabiri, L., & Henik, A. (2012). A unique asymmetrical Stroop effect in absolute pitch possessors. *Experimental Psychology, 59*(5), 272-278. doi: 10.1027/1618-3169/a000153

4.3.4 Grégoire, Perruchet and Poulin-Charronnat (2013)

Grégoire et al. (2013) designed another music Stroop task based on Zakay and Glicksohn (1985) and Akiva-Kabiri and Henik (2012) works. They proposed an alternative way of studying the development of automatism which does not depend exclusively on reading. For them, studying automatism with reading tasks have practical and ethical limitations (e.g. children start reading at the same time that they develop many more cognitive abilities and it is ethically constraining to control reading learning exposure) which can be surpassed by a music Stroop task. Two experiments were designed to test the use of this music Stroop task.

For Experiment 1, two evenly divided group of undergraduate psychology students were recruited. One group had musical training and played an instrument for at least 5 years, while the other did not. They designed three experimental stimuli (which are illustrated in Figure 7) for three conditions: congruent, incongruent and out-of-context condition. On the one hand, both the congruent and incongruent conditions consisted of a treble clef where a musical note was positioned. Notes could range from C4 to A5. Inside this note, a word was printed which could correspond with the musical note location or not. Specifically, in the congruent condition, the musical note and the word printed inside matched (e.g. a note in the A4 position had printed inside “La [A]”). Whereas in the incongruent condition the musical note and the word printed inside did not match (e.g. a note in the A4 position had printed inside “Si [B]”). On the other hand, the out-of-context condition presented only names of notes without the staff but still positioned in any of the 13 possible spatial locations, “as if they were correctly positioned on a virtual staff” (p. 271).

Both groups of participants had to press a key when the word presented in the screen was a note name, but to refrain if it was not (hence, it was a go/no-go task). Grégoire et al. (2013) hypothesized that a music Stroop effect would occur when musicians faced a printed note name that was incongruent with its location on the staff. This Stroop effect would be observed in a delayed response time of the go/no-go task. For this experiment, an extra set of words was created. These were CE, JE, TU, NI, TA, VU and PAR. The seven stimuli could either appear inside a note located in the staff (in-context condition) or in an out-of-context condition.

Participants were instructed to press a space bar if they saw a note name displayed on the screen and to refrain if they did not. If no response was made, the next stimuli was showed after 1.2 seconds. Stimuli could be displayed in four possible locations in the screen without an immediate repetition of the same location. In this way, researchers controlled for the influence of iconic memory on the processing of the next stimuli. Finally, between stimuli, a centred fixation cross was displayed for one second. For each of the conditions (congruent, incongruent, note names out-of-context, words in-context and words out-of-context) six different stimuli were showed in 13 possible locations, adding a total of 390 trials for the whole experimental session. These trials were segmented into 10 blocks and pseudorandomized, thus avoiding any immediate repetition of note locations, non-note words and note names. After the experiment, participants fill in a survey. Response times for space bar hits were calculated and out-of-context condition was used to calculate a baseline. A mixed-design ANOVA was used to test their hypothesis.

Larger response times were obtained for the music group in the incongruent conditions in contrast to the congruent conditions. However, the out-of-context response time was shorter in contrast to both, incongruent and congruent conditions for both participant’s groups. This either indicated that musical expertise was irrelevant for this effect, or that reading a word inside a complex background as a staff was harder than reading a word with an empty background.

Experiment 2 examined three questions raised from Experiment 1 results. First, whether the music Stroop effect was still present even if the go/no-go task was replaced with a reading task. Second, test whether the longer response time obtained in the in-context conditions for both musicians and non-musicians was due to the visual complexity of the stimuli display. Third, whether the slowness of non-musicians was due to “the categorical membership decision required in the no/no-go task” (p. 273). For this experiment, 34 new participants were recruited, half of which had music and instrumental training for at least 5 years, and half which did not. Both groups read aloud the printed words while ignoring their position on the staff. Response times were recorded with a voice key. Again, they were asked to respond as quickly and accurate as possible. Respond times were calculated for note names, and non-note words. Hypotheses were tested by a mixed-design ANOVA.

Results from Experiment 1 were replicated: musicians evidenced a music Stroop effect and a congruity effect was limited only to this group. Also, the faster reaction time of the out-of-context condition found on musicians and non-musicians was due to perceptual complexity of the stimuli.

Finally, Grégoire et al. (2013) propose a series of valuable recommendations for obtaining larger effect sizes. These are:

- Participants should differ only in their amount of musical training from the general population.
- Present many stimuli per condition.
- Counterbalance and pseudorandomize stimuli presentation.
- Item-by-item mode of presentation should be used.
- Only incongruent and congruent trials could be presented, since additional conditions did not had an effect.
- Smaller samples of note positions is preferable.

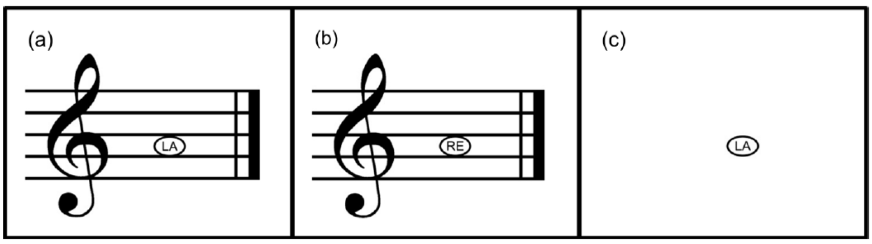

Figure 7. Example of a matching (a), non-matching task (b), and control task (c). Reproduced and adapted from Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect: Opening a new avenue to research on automatism. *Experimental Psychology, 60*(4), 269-278. doi: 10.1027/1618-3169/a000197

4.3.5 Controversies on Grégoire et al. (2013)

Grégoire et al.'s (2013) study raised several peer review commentaries which are worth presenting in this section. The purpose of doing so is to raise some methodological consideration noted by different authors and that will be taken into account for the experimental design of this research. Akiva-Kabiri and Henik (2014) noted some drawbacks in Grégoire's et al. (2013) music Stroop task. Basically, they pointed at two issues: heterogeneity of musical training and little importance to musical note naming. Akiva-Kabiri and Henik (2014) suggested that, for instance, musical education can greatly vary in the amount of practice that different instruments demand. Also, depending on the instrument that is being learned, notation systems might vary. Moreover, even if we would choose to pick only traditional western notation, musicians can vary on the kind of clef they can master. Concerning the second issue, Akivar-Kabiri and Henik (2014) claimed that explicit note naming training is hardly a common event in musical education. For them, musical education can start without needing to name notes (as in Suzuki system) or music notation is related to certain complex motoric movements needed to produce a sound in an instrument. Hence, when designing the stimuli it is important to take into account the training of each instrument, as well as their way of dealing with notation. Moeller and Frings (2014) pointed out at least two observations. The first one is a low ecological validity: notes are usually processed as motoric responses and not as orally reading names of notes. That is, musicians will respond to a note by producing a tone in his or her instrument; moreover, it is even possible that to name a note is irrelevant or impossible in order to produce the tone (e.g. a trumpeter cannot play any note while verbalizing it). The second one concerns the unknown influence of expertise in the music Stroop effect. According to Moeller and Frings (2014) Grégoire et al. (2013) missed to include a more heterogeneous sample of musicians with different levels of expertise. They included either non-musicians or musicians with greater familiarity with musical notation. However, it was not possible to determine if the influence of training followed a U-shaped since Grégoire et al. (2013) did not control for different levels of expertise. Nevertheless, recently Grégoire et al. (2015) addressed this limitation by testing whether the music Stroop effect interference increases as the level of practice does and the music Stroop effect follows a quadratic function as a consequence of years of practice. After testing non-musician children and children from different music education years, they reported a linear relationship between practice and interference effect just as the classical Stroop task literature suggests but failed to replicate an inverted U-shaped curve.

Consequently, including a measure of musical training time could describe the development of interference.

Zakay (2014) also raised some observations to Grégoire et al. (2013). Just like Moeller and Frings (2014), he considered that the experiment presented a problem of ecological validity: "whereas reading of words is a natural behaviour of most people, reading names of notes appearing on a staff is an unusual behaviour, even for musicians" (p. 78). Surely, musicians hardly ever encounter scores in which the name of notes are included inside the musical notation. Moreover, notes trigger complex psychological and motor behaviours that result in the execution of sounds in an instrument or voice (Zakay, 2014). Additionally, for Zakay (2014) the key for obtaining a real Stroop effect lies in obtaining a stimuli which interrelates two conflicting perceptual dimensions in the strongest way possible. According to him, in the music Stroop task proposed by Grégoire et al. (2013) the outcome of processing the name of a note and its movement-related-symbolic representation in a staff is not the same as the outcome of the classical Stroop task, in which the two perceptual dimensions (reading the words and naming the colour) produced the same outcome, i.e. the name of a colour. As a result of this, a stimuli which triggers motor responses would meet Zakay's (2014) observation and increase ecological validity.

Grégoire, Perruchet and Poulin-Charronnat (2014) defended their methodology by pointing out several points. Concerning Zakay (2014) observations, they claim that music Stroop task is not merely another Stroop-like test, in fact is a reverse Stroop-task because "reading is involved, but as the object, rather than the source, of interference" (Grégoire et al., 2014, p. 80). Moreover, their objective was not to replace the classical Stroop task but to provide a new method for studying the effect of practice on Stroop interference over time. According to a more recent research by Grégoire et al. (2015) previous Stroop research in the reading area has showed that the Stroop effect follows an inverted U-shaped curve as reading skills are acquired through schooling. They point at the fact that this result could either be due to the reading training received at school or the natural biological development in which crucial executive control-related brain areas are developed (Castellanos et al., 1999). Also, they clarify that by referring to the automaticity of note naming in musicians they are not implying that musicians experience an "irrepressible need to name aloud the note" (Grégoire et al., 2014, p. 81). Still, they acknowledged that note naming is an important part of musical education at least at its basics.

4.3.6 A new music Stroop task proposal

Grégoire et al. (2013) designed a music Stroop task as an alternative way of studying the development of automatism which did not depend exclusively on reading. However, different authors highlighted several limitations of the design which could compromise its ecological validity (see Section 4.3.5). Yet, Grégoire et al. (2013) showed that musicians evidenced a music Stroop effect. Additionally, they listed a series of suggestions to improve their design (see Section 4.3.4). Therefore, sufficient basis exist to try this design to study inhibitory control and transposition. In the Methodology section a new music Stroop task will be fully described. It will incorporate the suggestions of Grégoire et al. (2013), elements from previous music Stroop task designs (Akiva-Kabiri & Henik, 2012; Stewart et al., 2004; Zakay & Glicksohn, 1985) and the main methodological critiques and advices presented in Section 4.3.5.

5 THE CURRENT STUDY

Previous research has documented the plastic nature of EF across the life span (Diamond, 2013) and one way in which this has been tested is through music. More specifically, differences between musicians and non-musicians in EF related tasks has been reported elsewhere (Bialystok & DePape, 2009; Bugos et al., 2007; Joret et al., 2016; Moreno et al., 2014; Zuk et al., 2014). Furthermore, neuroplastic-related changes associated to music making has also been extensively documented (Bangert & Schlaug, 2006; Herholz & Zatorre, 2012). Therefore, there is a solid case for the argument that music making can have distinguishable behavioural and neurological effects. These differences are attributed to the highly demanding multimodal nature of music making (Bangert & Schlaug, 2006).

One clear example of a highly complex music task is transposition. As described in previous sections, transposition is a demanding cognitive task where a mismatch between the observed notes and the produced sounds seems to parallel the characteristic mismatch of the Stroop task paradigm. Therefore, it is plausible to assume a connection between transposition and inhibitory control tasks.

One possible way of exploring whether inhibitory control is a necessary cognitive function required for transposing would be by comparing transposing musicians, non-transposing musicians and non-musicians on their performance on inhibitory control tasks. However, the main drawback of this approach is that musicians could develop transposition skills independently of the type of instrument they play. Daily musical demands could expose a non-transposing musician into developing transposition skills (e.g. a piano accompanist). Conversely, a musician who plays a transposing instrument might never find herself in the need to transpose if the music context has not demanded it. Therefore, an objective measurement of transposition skills is required to overcome this heterogeneity of context-dependent transposition skills.

However, to explore the possible relationship between transposition skills and inhibitory control, a solid experimental paradigm designed to measure inhibitory control is needed. Although multiple ways of measuring inhibitory control exist (Diamond, 2013), the Stroop paradigm has shown a long-lived validity in studying inhibitory control (MacLeod, 1991). Additionally, relatively recent music adaptations of this task (see Section 4.3) can be potentially useful to explore in deeper extent the effects of musical training on the development of executive functions. Although there are empirical reasons to justify the selection of the Stroop paradigm, its suitability as an inhibitory control test for this particular study should be empirically tested.

Consequently, the first research question to be addressed is whether the music Stroop task and the classic Stroop task are measuring inhibitory control. To investigate this question, the following hypothesis will be tested:

- **Non-musicians will not differ on their performance of congruent and incongruent conditions on the music Stroop task.** A necessary condition to elicit an inhibitory control response is that the ongoing process should be automatized or easily executed by an individual (Banich & Depue, 2015; Diamond, 2013; Stroop, 1935). Therefore, since non-musicians have not automatized music reading, they should not experience any inhibition during the incongruent trials. Hence, the performance in the incongruent trials of the music Stroop task should be similar or equal to its congruent counterpart.
- **Musicians will differ on their performance of congruent and incongruent conditions on the music Stroop task.** Following the previous reasoning, musicians should have developed and automatized music reading. Therefore, when presented with incongruent stimuli an inhibitory control reaction should be provoked. Thus, the performance in the incongruent trials of the music Stroop task should be lower (i.e. take longer time and evidence more mistakes) than its congruent counterpart.
- **Non-musicians and musicians will differ on their performance of congruent and incongruent conditions on the classic Stroop task.** Since both groups of participants should be capable of reading literal texts, it is expected that the necessary automatization for eliciting an inhibitory control response should be present (Banich & Depue, 2015; Stroop, 1935). Therefore, the performance of the incongruent trials of the classic Stroop task should be lower (i.e. take longer time and evidence more mistakes) than its congruent counterpart.
- **Musicians will differ from non-musicians on global measurements of inhibitory control.** Global measurements of both Stroop tasks will combine the accuracy and speed parameters and be sensitive to their trade-off. That is, it will control for cases in which a participant had a slower response but made fewer mistakes; as well as for cases in which a participant had a faster response but made more errors (see Section 6.5.7 for further details). Additionally, global scores of both Stroop tasks should be significantly correlated if they measure inhibitory control. Further support would be given to this assumption if musicians and non-musicians differ on the performance of the linear

combination of the Music Stroop and the classic Stroop global scores of inhibitory control.

The second research question will explore whether existing transposition skills impact upon inhibitory control task performance. To answer this question, the following hypotheses will be tested:

- **Musicians with higher transposition skills will outperform musicians with lower transposition skills in the music Stroop task.** This hypothesis rests on two assumptions: inhibitory control is necessary for music transposition to occur; and inhibitory control is affected by training. There is previous evidence that shows that inhibitory control can be trained and change with experience (Diamond, 2013). Therefore, if both assumptions are true, then it is expected that musicians with greater transposition skills will have trained their inhibitory control to a greater extent than their less experienced colleagues. Thus, the effects of music transposition on inhibitory control should translate into differences in the performance of the music Stroop task. Specifically, the higher the transposition skills of the participants, the faster and more accurate they will be in the music Stroop task. Conversely, the lower the transposition skills of the participants, the slower and less accurate they will be in the music Stroop task.
- **Musicians with higher transposition skills will outperform musicians with lower transposition skills in the classic Stroop task.** This hypothesis also rests in the assumption that the effects of music transposition on inhibitory control should translate into differences in the performance of, in this case, the classic Stroop task. Nonetheless, this hypothesis has bolder implications. It would suggest that specialized musical training can transfer to other cognitive domain functions such as conventional reading (i.e. reading texts instead of notes).
- **Musicians with higher transposition skills will outperform musicians with lower transposition skills in global inhibitory control scores.** Finally, this hypothesis is also grounded in the assumption previously explained. However, this hypothesis will be tested by combining the accuracy and reaction time scores for each Stroop task.

6 METHODOLOGY

6.1 Participants recruitment

Participants in all experiments here reported were recruited by mailing list, social media websites, QR-code flyers and face-to-face invitations. A webpage containing the description of the experiment was distributed. This web site was created using Google Sites and through it participants could also choose an appointment date, fill in an initial questionnaire and signed a consent form. Appointments were organized using Doodle and the initial questionnaire was designed in Google Forms. Three 50€ Sokos gift-cards were raffled within the participants of the study. After participating, subjects were debriefed on the real objective of the study.

6.2 The present experiment

In this section a general description of the experimental design will be presented. Later, every section of the process will be thoroughly described. Participants filled in a questionnaire before attending to the experiment. Based on the responses, participants were assigned to the group of “non-musicians” or to the group of “musicians”. Non-musicians continued directly to respond to the music Stroop and the classic Stroop tasks. The order of presentation was randomly counterbalanced. In contrast, musicians answered an additional questionnaire designed to identify their most familiar clef and notation system. Next, they performed a transposition task before continuing with the music Stroop and the classic Stroop tasks. The transposition task had the purpose of measuring their transposition skills. This measurement was later used to predict their performance in the inhibitory control tasks (music Stroop and classic Stroop tasks). Once the transposition task was completed, music participants responded to the remaining two tasks in randomly counterbalanced order. A diagram of the experimental procedure is depicted in Figure 8.

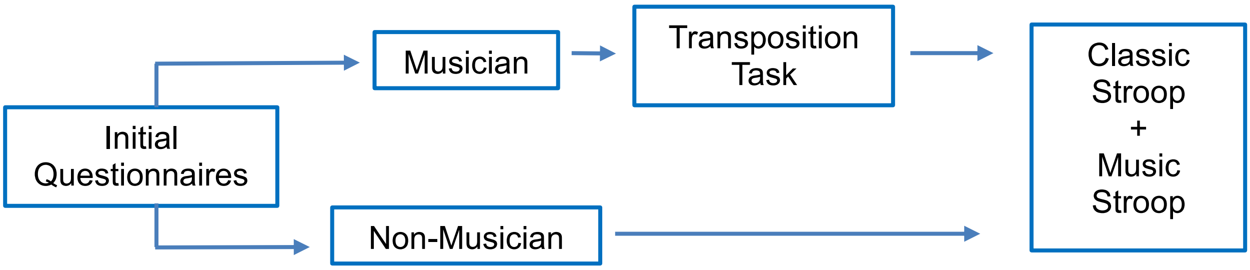

Figure 8. Diagram of experimental design.

6.3 Initial questionnaires

Two questionnaires were designed and used at the beginning of the experiment. First, a 13-item online-questionnaire gathered relevant information from the participants. The most relevant questions for this study were whether participants self-identified as musicians or not, what their mother tongue was, and what their primary music instrument was. Participants classified as musicians if they rated positive in two out of the three following questions of the online-questionnaire: “Do you self-identify as a musician?” “Is music your main source of income?” and “Do you read music?” Therefore, even if a participant had not studied a music-related degree, they could still be considered as musicians. Non-musicians who could read music were excluded from the study. The questionnaire also controlled for participants with colour blindness and dyslexia. An example can be found in Appendix B.

Second, in order to determine the most familiar clef and music reading system for music participants, a single A4 size page document containing the bass, alto and treble clefs was designed. One “B” whole note was printed in every staff. Participants were asked two questions. First, “on which of the following clefs do you feel more comfortable reading at?” Second, “how do you call this note (pointing at the B whole note)?” The reading system was determined by means of the B note because of the different names it can adopt. Therefore, if a participant responded: “B”, then he or she would be answering the English notation system; if a participant responded: “H” then he or she would be answering the German notation system; and, lastly, if a participant responded: “Si”, then he or she would be answering the Latin notation system (see Figure 9).

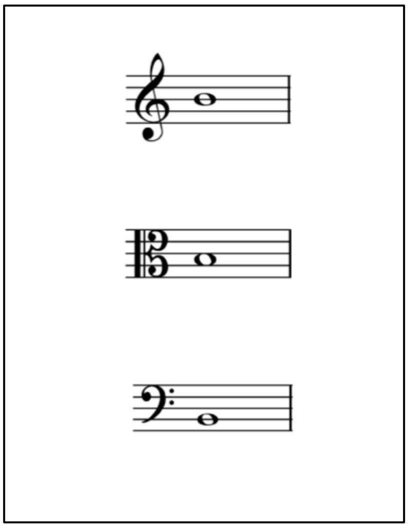

Figure 9. Method for determining the participant's most familiar notation system.

6.4 Transposition task

The transposition task was designed to measure the transposition skills of music participants. Musicians were asked to bring their main instrument to the meeting or they were provided to them by the researcher. After determining their most familiar musical reading system, they were audio recorded playing a series of quarter notes. Musicians were asked to play the notes either as written or transposed to a different interval. A general overview of this design is presented in Figure 10. Next, a more detailed description of each stage will be provided.

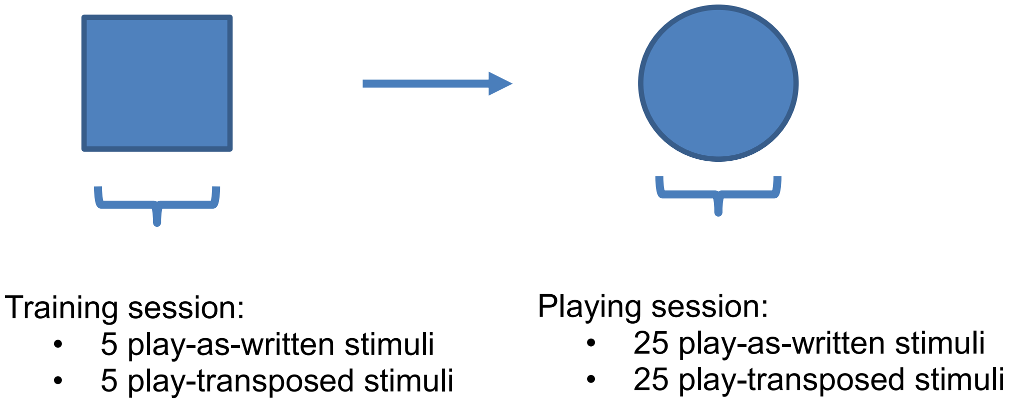

Figure 10. Overall design of the transposition task.

6.4.1 Participants in transposition task

Thirty-four musicians (female = 20; *M* age = 28.57; *SD* = 7.77; minimum = 20; maximum = 44) participated in this task. On average, participants have been playing their instruments for 18.59 years (*SD* = 7.41). Of the total sample size, 15 musicians were currently studying a music degree.

6.4.2 Stimuli design

Music stimuli were designed using the Sibelius music notation software (version 8.7). A total of 15 quarter notes were created. These quarter notes ranged from the second line of the staff (counting from the bottom) up to the fourth line. The notes could be located in a bass, alto or treble clef. The Sibelius file was exported to PDF format (version 1.4, Acrobat 5.x). The document was then zoomed to 600% of its standard size. Every music note was print-screened and edited in Paint (Microsoft Windows version 1709) until reaching the same dimensions (864 pixels width; 604 pixels height). Text was presented over a white background in black Arial font and on the default PsychoPy (Peirce, 2007, 2009) font size.

6.4.3 Task design

The transposition task was implemented in PsychoPy2 Experiment Builder on its version 1.85.2 (Peirce, 2007, 2009). Since three different clefs were used (i.e. bass, alto and treble) three versions of the task were created. Additionally, each was divided into two sections: a “training session” and a “playing session”.

6.4.4 Training session

During the training session participants were presented with the following introductory instructions:

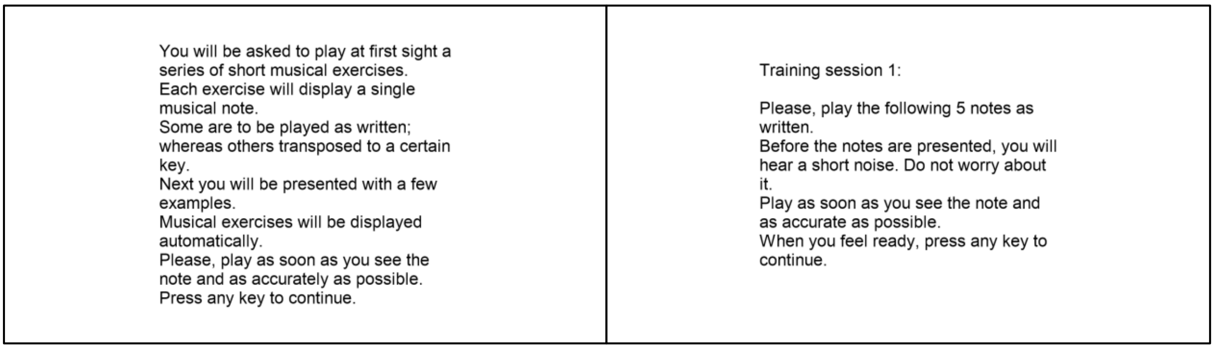

Figure 11. Initial instructions of the training session (left) and instructions of the “play-as-written” condition (right).

A 1.5 seconds noise was included to later calculate the reaction time of musicians. Immediately a quarter note together with a clarifying message were displayed on a white screen for 6.5 seconds. The five notes were randomly presented. See Figure 12.

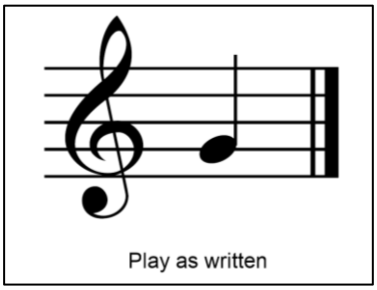

Figure 12. Training session example of the “play-as-written” condition.

The transposed training session started with the following instructions:

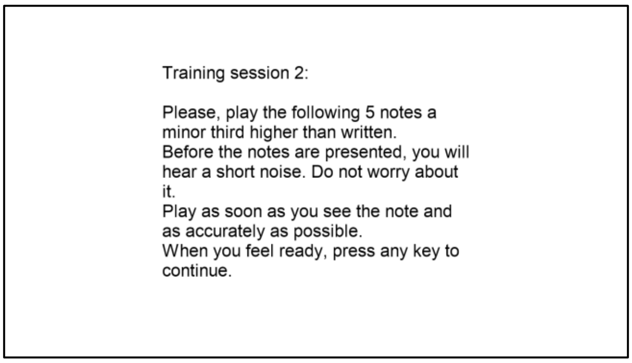

Figure 13. Instructions of the “play-transposed” condition.

Equally to the play-as-written training condition a 1.5 seconds noise was presented. A quarter note with a reminder note were also presented. The five notes were randomly displayed. See Figure 14 for an example.

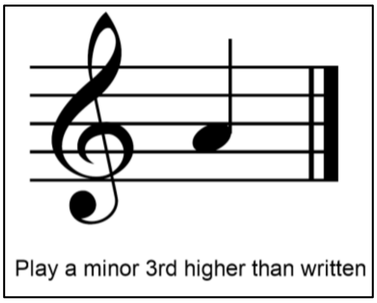

Figure 14. Training session example of the “play-transposed” condition.

A final set of instructions was presented to announce the beginning of the experimental session. Instructions read as follows:

> Now, you will be asked to play at first sight another series of short musical exercises.
> Once again, some are to be played as written; whereas others are transposed to different keys.
> No clarifying message will be displayed this time.
> As before, musical exercises will be displayed automatically.
> Please, play as soon as you see the note and as accurately as possible.
> Press any key to continue.

Figure 15. Instructions of the transposition task.

6.4.5 Playing session

Participants were randomly assigned to the order of presentation of the play-as-written or played-transposed conditions (see Figure 16). Random assignment was performed using the online random list generator (Haahr, 2006). Sets were interlayered. Every experimental trial lasted for a total of 8.00 seconds. The following intervals were used: major 2nd, major 3rd, perfect 4th, perfect 5th, and minor 6th. These particular intervals were chosen to control the level of transposition difficulty. Since musicians could vary on their transposition skills, it was important to present a variety of not so common intervals.

Please, play the following 5 notes as written. Play as soon as you see the note and as accurate as possible. When you feel ready, press any key to continue.	Please, play the following 5 notes a major 3rd higher than written. Play as soon as you see the note and as accurately as possible. When you feel ready, press any key to continue.

Figure 16. Playing session's instructions for the play-as-written (left) and played-transposed conditions (right).

The experimental session contained 10 sets of stimuli containing 5 musical notes each. The 10 sets were divided into 5 played-as-written and 5 played-transposed tasks.

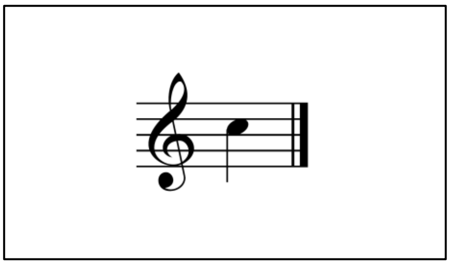

Figure 17. Stimulus musical note example.

6.4.6 Transposition skill calculation

Participants were audio recorded using a Zoom H4n handy recorder. Audio files were processed in MATLAB using the MIRToolbox (Lartillot, Toiviainen, & Eerola, 2013). Every audio file contained 10 sets of stimuli each one preceded by one instruction set. Before calculating the transposition skill of the participants, three were excluded from the analysis. One participant was excluded because the noise recorded in the audio file made the content unintelligible. A second participant could not be considered inside the final sample because the non-tempered nature of the instrument did not allow to play accidental notes present in the task. Finally, one last participant did not know what intervals were, making the task impossible to be accomplished. Therefore, the final sample consisted of 31 musicians (female = 20; *M* age = 28.65; *SD* = 7.96; minimum = 20; maximum = 44).

The transposing skill level was calculated through the following procedure: First, every audio file was processed through the rhythmic feature extraction function "mironsets" or "the computation of an onset detection curve, showing the successive bursts of energy corresponding to the successive pulses. A peak picking is automatically performed on the onset detection curve, in order to show the estimated positions of the notes" (Lartillot, 2013, p. 88). The exact temporal location of each onset was obtained through the function "mirgetdata". An audio file example is presented in Figure 18.

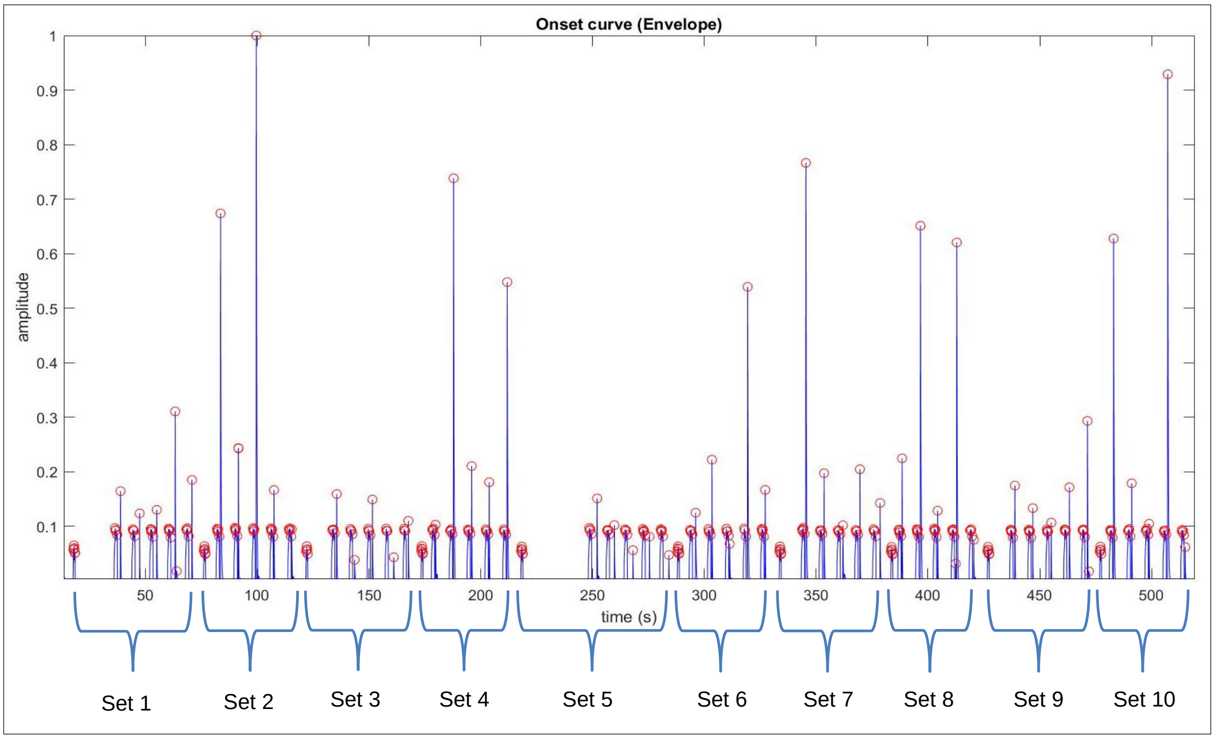

Figure 18. Example of audio file analysed with mironsets. Blue colour lines represent the audio signal. Red circles represent onsets.

Second, the time difference between the note onset and the last onset of the white noise was calculated for every stimuli. Whenever the mironsets function did not registered the last peak of a white noise (as in the last two white noise peaks in Figure 19), the "data cursor" option was selected and used to pin the time location. A graphical representation of this process is illustrated in Figure 19.

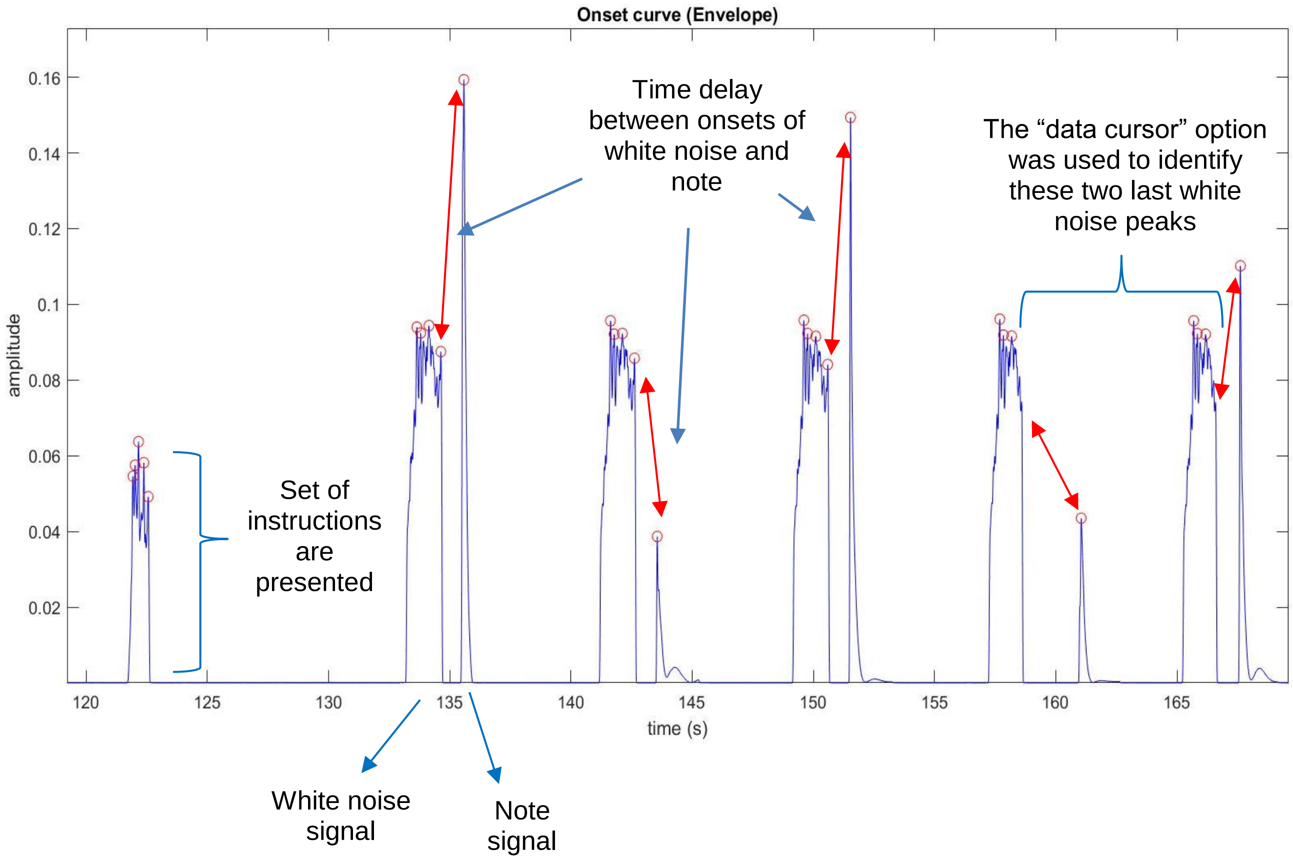

Figure 19. Example of a stimuli set.

Third, the five reaction times of every set were summed up. This process enabled identifying a play-as-written condition time and a play-transposed condition time. Then, these values where average by the amount of play-as-written (i.e. 5) and play-transposed conditions (i.e. 5). The difference between the latter and the former was calculated. The reason why it was decided to average the play-as-written and play-transposed scores is because otherwise one more participant would have been eliminated. A technical error resulted in the collapse of the software, losing the last play-as-written trial. To solve this, one play-transposed condition was randomly eliminated, leaving this participant with a set of 8 sets. To keep the scores of this participant for the posterior analysis, it was decided to calculate the average transposing skill level of participants instead of their absolute value.

Lastly, the resulting transposition skill scores were used for two purposes. The first one was to create a predictor variable for the classic and music Stroop task performance. The second one was to divide the musicians into three levels of transposition: high (n = 11), middle (n = 10) and low (n = 10) levels. The lower the score, the higher the transposing skill level.

6.5 Experiment 1: Music Stroop task

The music Stroop task was designed for measuring the inhibitory control level of musician and non-musician participants. Its design has closely followed Grégoire et al. (2013) model, but with some modifications. The design also took into account Steward et al. (2004), Zakay and Glicksohn (1985), Akiva-Kabiri and Henik (2014), Grégoire et al. (2014), Moeller and Frings (2014), and Zakay's (2014) methods and discussions.

Participants were showed a music score in which a whole note contained a note written name. The note written name could coincide with the whole note's location on the pentagram or not. Participants were asked to choose the note written name appearing inside a whole note at all times. To increase ecological validity of the experiment, familiarity with musical clefs as well as music reading system were accounted for. A general overview of this design is presented in Figure 20.

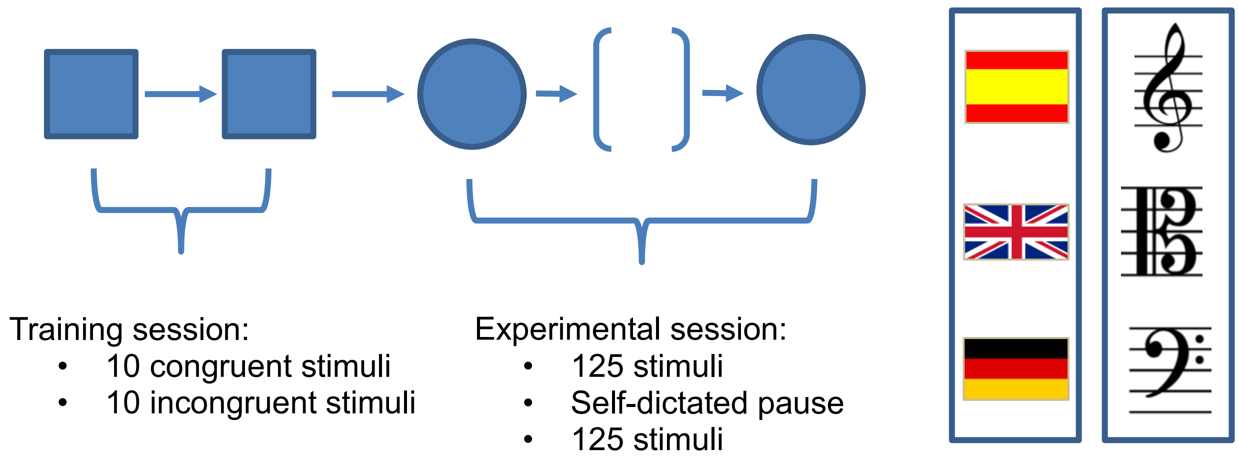

Figure 20. Overall design of the music Stroop task.

6.5.1 Participants in Experiment 1: Music Stroop task

A final sample of 64 participants was collected. Thirty-four were musicians (female = 21; *M* age = 28.57; *SD* = 7.77; minimum = 20 years; maximum = 44 years), and 30 were non-musicians (female = 22; *M* age = 28.05; *SD* = 7.25; minimum = 20 years; maximum = 44 years). Participants were recruited following the procedure described on Section 6.1.

6.5.2 Stimuli design

Music stimuli were designed using the Sibelius music notation software (version 8.7.). A total of 225 whole notes were created. These were the result of designing 25 notes per each music clef (bass, alto and treble) and per each music note reading system (German, English and Latin). Whole notes ranged from the second line of the pentagram (counting from the bottom) up to the fourth line. The Sibelius file was exported to PDF format (version 1.4, Acrobat 5.x). The document was then zoomed to 600% of its default size. Every music note was print-screened and edited in Paint (Microsoft Windows version 1709) until reaching the same dimensions (864 pixels width; 604 pixels height). Written note names were included inside every whole note. Independently of the note reading system, chosen font was Arial 20. Only the name "Sol" was set on font size 17 in order to get a better fit inside the whole note. Written names could either match the note location or do not match it. Text was presented over a white background in black Arial font and on the default PsychoPy font size. See Figure 21 for an example and Appendices B and C for examples of every clef and reading system.

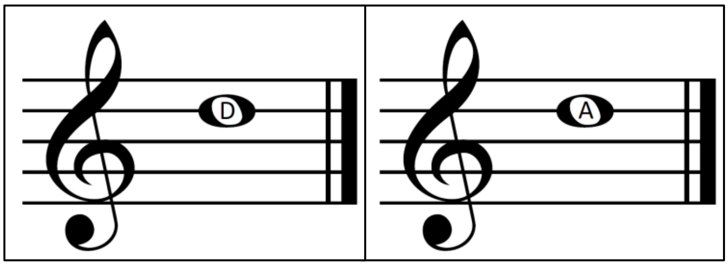

Figure 21. Example of congruence between note written name and its location in the staff (left) and of incongruence between the note written name and its location in the staff (right).

6.5.3 Task design

The music Stroop task was implemented in PsychoPy2 Experiment Builder on its version 1.85.2 (Peirce, 2007, 2009). One version of this task was created for each clef used (i.e. bass, alto and treble) and for each note reading system (German, English and Latin). Within every task two sections were designed: a "training session" and an "experimental session". The selection of the musical clef as well as the musical naming system was determined through one of the initial questionnaires (determine the most familiar notation system). To determine the note naming system of non-musicians, they were asked "how are you used to call the notes in your country?" By default, all participants in this group were presented with the treble clef.

The congruent stimuli featured five locations ranging from the second line of the pentagram (counting from the bottom up) up to the fourth line. The note location was determined by a whole note and the name of the note by a letter (e.g. A) or word (e.g. La) inside the whole note. Every note name and its corresponding whole note where repeated 25 times on every location. In other words, “G” was showed 25 times in the G line (if presented on the treble clef), “A” was showed 25 times in the A space, etc.

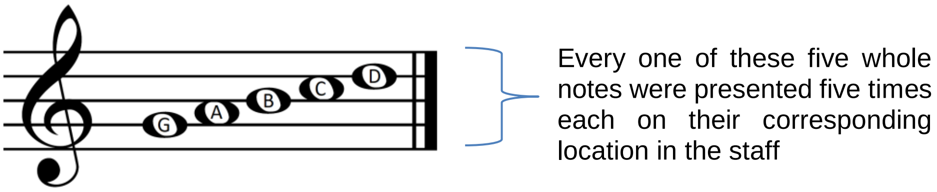

Figure 22. Conceptualization of the congruent stimuli design process.

Meanwhile, the incongruent stimuli featured the same five locations ranging from the second line of the pentagram up to the fourth line. The note location was determined by a whole note and the name of the note by a letter or word inside the whole note. The five whole notes were displayed 25 times in each possible location. However, the options of names of notes that could be featured inside the musical symbol were reduced by one. That is, if a whole note was displayed in the G line of the treble clef, then A, B, C or D could be showed inside the whole note. Therefore, only 4 incongruent notes were left to be displayed inside the note.

This presented a problem for keeping a parallel design with the congruent trials, since five incongruent note name options were required per each location in the staff. Therefore, every note name was repeated twice until meeting every location on the staff. That is, for the G line the five note names were A, B, C, D and A. For the second set of the G line the next five note names were B, C, D, A and B. For the third set of the G line the next five note names were C, D, A, B and C. For the fourth set of the G line the next five note names were D, A, B, C and D. Finally, for the last set of the G line the next five note names were once again A, B, C, D and A. This solution was systematically repeated on the remaining four locations in the staff. Since there was no interest on knowing the effect of every specific note name on the interference experience, it did not matter to have every incongruent note name set repeated twice.

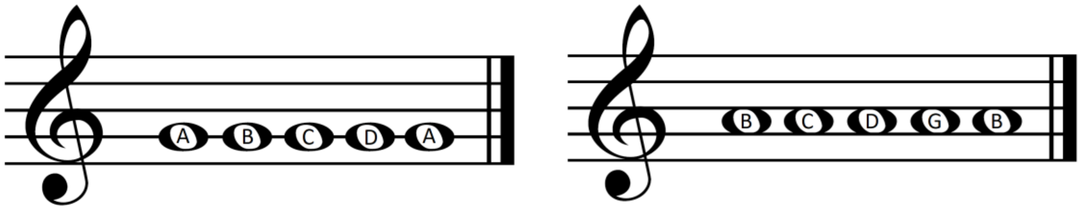

Figure 23. Conceptualization of the incongruent stimuli design process. Every line and space on the staff had five notes including a repetition of the first of each series.

Following Grégoire et al. (2013), the staff was presented at one of the four possible positions in the screen to "prevent the iconic memory of the staff to influence the processing of the following note" (p. 272). The written name of the note, the note symbol and the location of the staff on the screen were pseudorandomized to avoid immediate repetition by using the software Mix (van Casteren & Davis, 2006). Stimuli remained on the screen until the participant elicited a response.

6.5.4 Training session

The training session consisted of 10 congruent stimuli trial and a 10 incongruent stimuli trial. Each one was preceded by a set of instructions. During the first part of the training session participants were asked to choose the note written name when the note location and its written name matched. An example of this set of instructions is presented in Figure 24.

You will be presented with a series of
musical notes.
The note location and the written name
of that note match each other.
Please, choose the note written name.
1 = G
2 = A
3 = B
4 = C
5 = D
Please, answer as fast and accurately
as possible.
Press any key to continue.

Figure 24. Congruent trial instruction.

Participants were then presented with 10 example stimuli with clarifying instructions. Before the presentation of the stimuli, a fixation cross was located in the middle of the screen and displayed for 0.5 seconds.

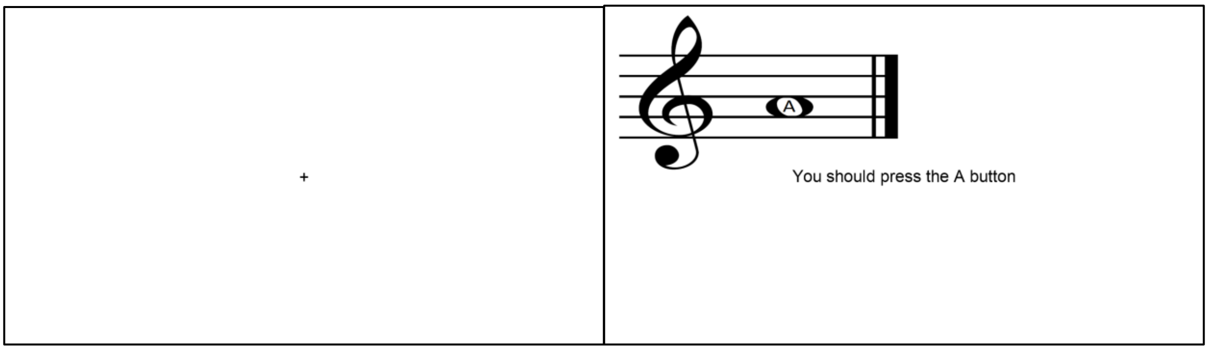

Figure 25. Example of congruent training stimulus.

Instructions for the incongruent trial read as follows:

You will be presented with a series of musical notes in a staff.
Now, the note location and the written name of that note DO NOT match each other.
Please, choose the note written name.
1 = G
2 = A
3 = B
4 = C
5 = D
Please, answer as fast and accurately as possible.
Press the "N" key to continue.

Figure 26. Incongruent training instruction.

Participants were then presented with the 10 example stimuli with a clarifying instruction. Before the presentation of the stimuli, a fixation cross was located in the middle of the screen and displayed for 0.5 seconds.

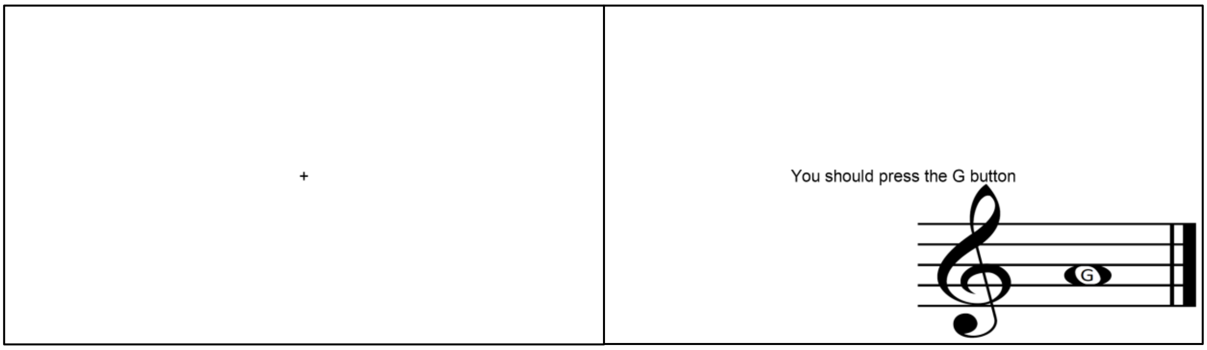

Figure 27. Example of incongruent training stimulus.

After finishing the training session, participants were asked if the instructions were clear. Then the researcher indicated to start the experimental session after he left the room.

6.5.5 Experimental session

The experimental session consisted of a 125 congruent stimuli and a 125 incongruent stimuli trial. Both were presented in a pseudorandomized order to avoid immediate repetition of note location, note naming, staff location and congruence or incongruence. There was a single set of instructions and a self-dictated pause in the middle of the trial. An example of the instructions is presented in Figure 28.

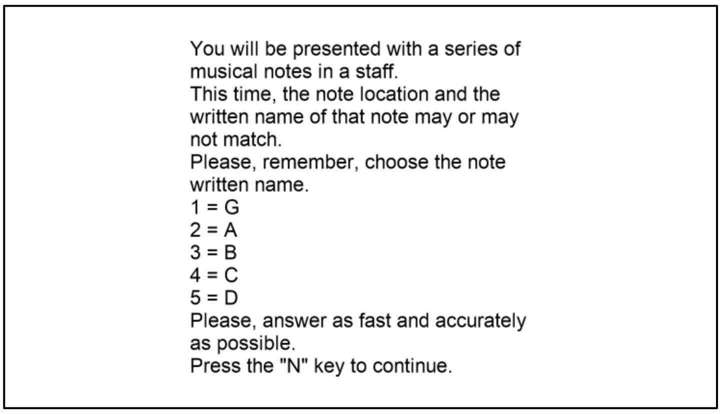

Figure 28. Experimental session instructions.

Participants were then presented with the 250 pseudorandomized stimuli. These were equally divided in 125 congruent stimuli and 125 incongruent stimuli. Before the presentation of the stimuli, a fixation cross was located in the middle of the screen and displayed for 0.5 seconds.

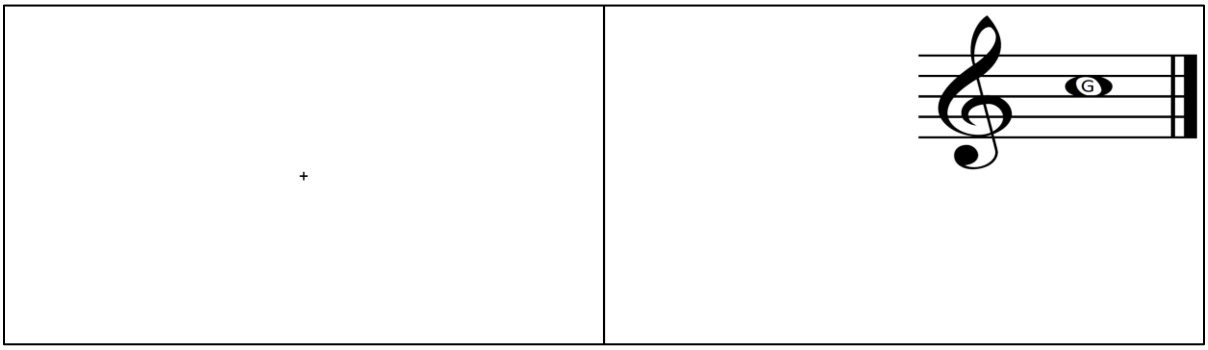

Figure 29. Example of experimental session stimulus.

6.5.6 Response interface design

Participants responded to the stimuli in a laptop keyboard placed in front of them based on Stewart et al.'s (2004) response setup. Five music note names were printed as tags and pasted in the first five numerical keys of the laptop. Letter font Calibri 18 was used for all the tag names. Tags were changed for each participant according to their most familiar clef and note reading system.

6.5.7 Rating of the scores of the music Stroop task

Based on Scarpina and Tagini (2017), three scores were calculated for each participant: accuracy, speed and a global interference score. Accuracy was defined as the amount of correct answers on both the congruent and incongruent trials. Therefore, scores could vary from 0 to 125. Speed was defined as the total amount of time that a participant took to answer both the congruent and incongruent trials. Finally, the global interference score was determined by the following equation based on Scarpin and Tagini (2017):

$$Global\ interference\ score = \left(\frac{a}{b}\right) - \left(\frac{c/d}{2}\right)$$

a = number of right answers in the incongruent trials

b = total reaction time of the incongruent trials

c = number of right answers in the congruent trials

d = total reaction time of the congruent trials

Global interference scores were calculated to account for the trade-off between accuracy and speed. That is, to control for cases in which a participant had a slower response but more accuracy; as well as for cases in which a participant had a faster response but less accuracy.

Therefore, global interference scores will provide a comprehensive account of inhibitory control performance. The lower the global interference score, the lower the participant's performance on the music Stroop task.

6.6 Experiment 2: Classic Stroop task

A motor variation of the classic Stroop task was selected as a complementary measurement of inhibitory control level of musician and non-musician participants. A motor response version of this task was chosen for two reasons: to keep a parallel response format as the music Stroop task and to preserve the more natural motor behaviour characteristic of instrumental playing. Previous literature supports motor implementations of the Stroop task (MacLeod, 1991).
Participants were presented with written names of colours which could be displayed in a congruent colour (the word RED showed in red colour) or incongruent colour (the word RED showed in yellow colour). The task consisted in ignoring the written word and choose the perceived colour. To increase ecological validity of the experiment, the words were presented in the participants' mother tongue. A general overview of this design is presented in Figure 30.

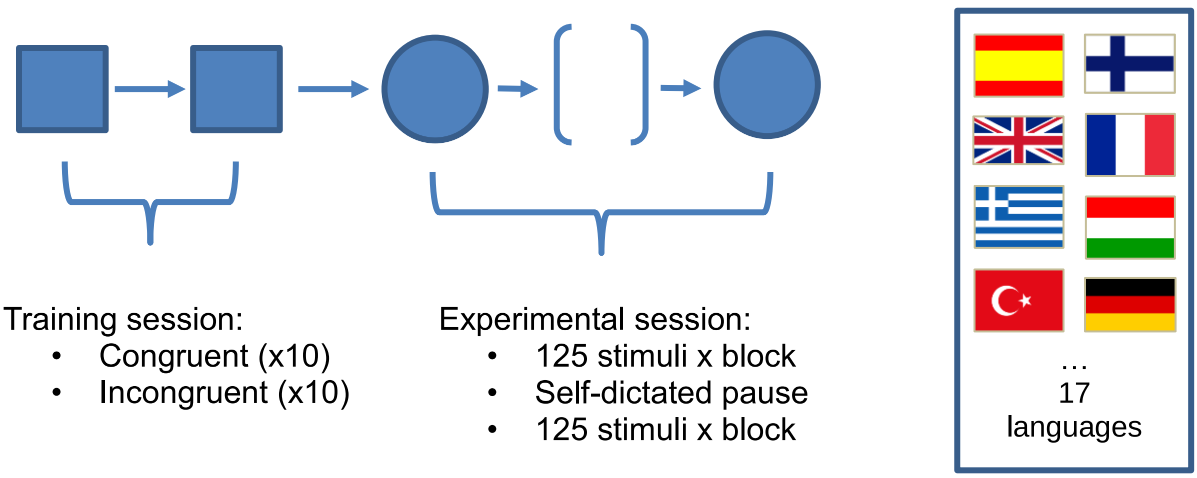

Figure 30. Overall design of the classic Stroop task.

6.6.1 Participants in Experiment 2: Classic Stroop task

Subjects participating in this experiment were the same as the ones described in Section 6.5.1. Seventeen different languages were identified. The majority corresponded to Finnish (36.9%) and English (12.3%).

6.6.2 Stimuli design

Based on Stroop (1935) five colour names were used: red, blue, green, yellow and purple. Although Stroop originally included brown, it was replaced by yellow to facilitate colour contrast in the screen. Words were presented over a grey background in the middle of the screen in Arial font and on the default PsychoPy font size.

Figure 31. Example of congruence between colour written name and its ink (left) and of incongruence between the colour written name and its ink (right).

6.6.3 Task design

The classic Stroop task was implemented in PsychoPy2 Experiment Builder on its version 1.85.2 (Peirce, 2007, 2009). A classic Stroop task was created for every mother tongue identified through the online-questionnaire. Within every task, two sections were designed: a "training session" and an "experimental session".

6.6.4 Training session

The training session consisted of 10 congruent stimuli trial and a 10 incongruent stimuli trial. Each one was preceded by a set of instructions. During the first part of the training session written names of colours and their ink matched. Participants were asked to choose the "ink of the colour" and ignore the written word. An example of this set of instructions is presented in Figure 32.

Figure 32. Congruent trial instruction.

Participants were then presented with the 10 example stimuli with a clarifying instruction. An empty space of 0.5 seconds preceded the appearance of every colour word.

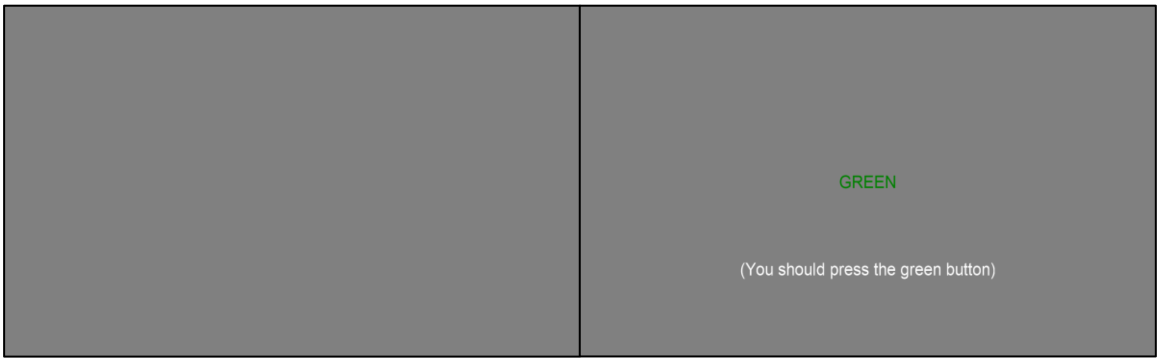

Figure 33. Example of congruent training stimulus.

Instructions for the incongruent trial read as follows:

Figure 34. Example of incongruent training stimulus.

Participants were then presented with the 10 example stimuli with a clarifying instruction. An empty space of 0.5 seconds preceded the appearance of every colour word.

Figure 35. Example of incongruent training stimulus.

After finishing the training session, participants were asked if the instructions were clear. Participants were encouraged to keep the written words under the scope of their sight to reduce the possibility that they could use their peripheral vision. The researcher indicated to start the experimental session after he left the room.

6.6.6 Experimental session

The experimental session consisted of a 125 congruent stimuli and a 125 incongruent stimuli trial. Both were presented in a pseudorandomized order to avoid immediate repetition of colour written name and colour ink. There was a single set of instructions and a self-dictated pause in the middle of the trial. An example of the instructions is presented in Figure 36.

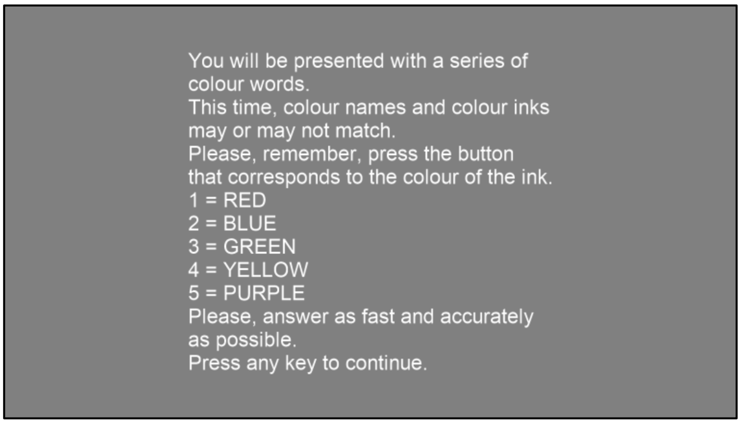

Figure 36. Experimental session instructions.

Participants were then presented with the 250 pseudorandomized stimuli. An empty space of 0.5 seconds preceded the appearance of every colour word.

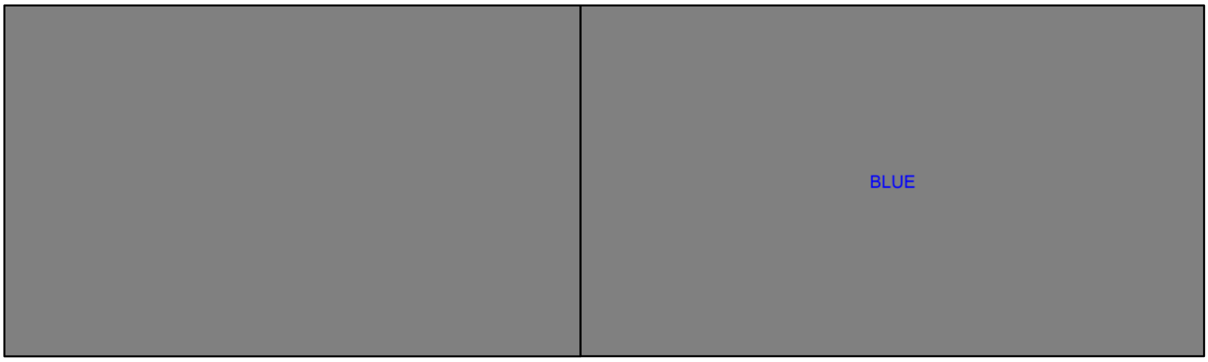

Figure 37. Example of experimental session stimulus.

6.6.7 Response interface design

Participants responded to the stimuli in a laptop keyboard placed in front of them based on Stewart et al.'s (2004) response setup. The five colour tags were pasted in the first five numerical keys of the laptop.

6.6.8 Rating of the scores of the classic Stroop task

The calculation of the accuracy, speed and global interference scores of the classic Stroop task were also based on Scarpina and Tagini (2017) and defined as in the Music Stroop Task section.

6.7 Statistical analysis

6.7.1 Normality assumptions

Normality assumptions were tested with Shapiro-Wilk test.

6.7.2 Detecting and modifying outliers

Percentiles for the variables to be analysed were calculated. Values falling below the 5th percentile and above the 95th percentile were Winsorized following Aguinis, Gottfredson and Joo (2013). That is, values exceeding the cutting-points were transformed to the values of the 5th and 95th percentiles.

6.7.3 Hypothesis testing

Hypothesis were evaluated through one-way ANOVA, one-way mixed ANOVA and one-way MANOVA tests. Significance value was stablished at $p < .05$.

6.7.4 Statistical processing

IBM SPSS Statistics on its 24th version was used to process and analyse experimental data.

7 RESULTS

Results will be presented by research question. Every question presents a normality assumption analysis, specific statistical analysis driven by each question, and a specific discussion.

7.1 Are the music Stroop task and a classic Stroop task measuring inhibitory control?

7.1.1 Non-musicians will not differ on their performance of congruent and incongruent conditions on the music Stroop task.

Preliminary analysis of the music Stroop task show similar mean values among the accuracy (i.e. number of correct answers in a condition) and reaction time (i.e. total time taken to respond on each condition). Because the accuracy dimension showed significant differences from normality, only the reaction time dimension will be included for the next statistical tests. A summary of these results are presented in Table 2.

Table 2

Normality tests of music Stroop task accuracy and reaction time of non-musicians

Music Stroop Measurements	*M*	*SD*	Shapiro-Wilk		
			Statistic	*df*	*p*
Correct answers in incongruent trials	122.90	2.76	.76	30	< .001***
Correct answers in congruent trials	123.20	1.74	.84	30	< .001***
Reaction time in incongruent trials	125.60	16.20	.97	30	.58
Reaction time in congruent trials	123.70	16.30	.97	30	.60

***p < .001.

A one-way repeated-measures ANOVA showed that reaction times on the music Stroop task did not differ significantly between the incongruent (M = 125.60, SD = 16.20) and congruent (M = 123.70, SD = 16.30) conditions, $F(1, 29) = 3.80$, $p = .06$, $r = .34$, $1 - \beta = .48$.

7.1.2 Musicians will differ on their performance of congruent and incongruent conditions on the music Stroop task.

Preliminary analysis of the music Stroop task show similar mean values among the accuracy (i.e. number of correct answers in a condition) but not on reaction time (i.e. total time taken to respond on each condition) among musicians. Because the accuracy dimension showed significant differences from normality, only the reaction time dimension will be included for the next statistical tests. A summary of these results are presented in Table 3.

Table 3

Normality tests of music Stroop task accuracy and reaction time of musicians

Music Stroop Measurements	*M*	*SD*	Shapiro-Wilk		
			Statistic	*df*	*p*
Correct answers in incongruent trials	121.59	3.60	.85	34	< .001***
Correct answers in congruent trials	123.15	2.84	.66	34	< .001***
Reaction time in incongruent trials	131.49	15.23	.98	34	.73
Reaction time in congruent trials	118.96	15.39	.95	34	.13

***$p < .001$.

A one-way repeated-measures ANOVA showed that incongruent ($M = 131.49$, $SD = 15.23$) and congruent ($M = 118.96$, $SD = 15.39$) reaction times on the music Stroop task differ significantly, $F(1, 33) = 76.47$, $p < .001$, $r = .84$, $1 - \beta = 1.00$.

7.1.3 Non-musicians and musicians will differ on their performance of congruent and incongruent conditions on the classic Stroop task.

Preliminary analysis of the classic Stroop task show similar mean values among the accuracy values on both groups of participants. Further normality tests showed that all measurements of accuracy deviate significantly from normality. A summary of these results are presented in Table 4.

Table 4

Normality tests of classic Stroop test accuracy of non-musicians and musicians

Group	Condition	*M*	*SD*	Shapiro-Wilk		
				Statistic	*df*	*p*
Non-musicians	Correct answers in incongruent trials	122.53	2.36	.83	30	.001***
	Correct answers in congruent trials	123.41	2.45	.68	30	.001***
Musicians	Correct answers in incongruent trials	121.97	3.04	.86	34	.01**
	Correct answers in congruent trials	123.41	1.84	.80	34	.001***

**$p < .01$. ***$p < .001$.

The reaction time dimension of the classic Stroop task showed differences between the congruent and incongruent trials. Across all participants, congruent trials present themselves as lower than incongruent. Despite the fact of violation of normality assumption on the group of musicians, it was decided to carry on a mixed repeated-measures ANOVA. This decision is based upon the central limit theorem (Field, 2009).

Table 5

Normality tests of classic Stroop test reaction time of non-musicians and musicians

Group	Condition	*M*	*SD*	Shapiro-Wilk		
				Statistic	*df*	*p*
Non-musicians	Reaction time in incongruent trials	108.59	13.74	.97	30	.61
	Reaction time in congruent trials	99.16	12.08	.98	30	.78
Musicians	Reaction time in incongruent trials	105.20	15.47	.93	34	.03*
	Reaction time in congruent trials	96.34	15.18	.88	34	.01**

*$p < .05$. **$p < .01$.

A one-way mixed repeated-measures ANOVA showed that incongruent (M = 106.89, St. error = 1.84) and congruent (M = 97.75, St. error = 1.73) reaction times on the classic Stroop task differ significantly, $F(1, 62) = 254.40$, $p < .001$, $r = .90$, $1 - \beta = 1.00$. Performance in the reaction time dimension of the classic Stroop task did not differ significantly when comparing non-musicians and musicians, $F(1, 62) = 0.24$, $p = .63$, $\eta p^2 = .004$, $1 - \beta = .08$. However, overall musicians were a few seconds faster than non-musicians (see Figure 38).

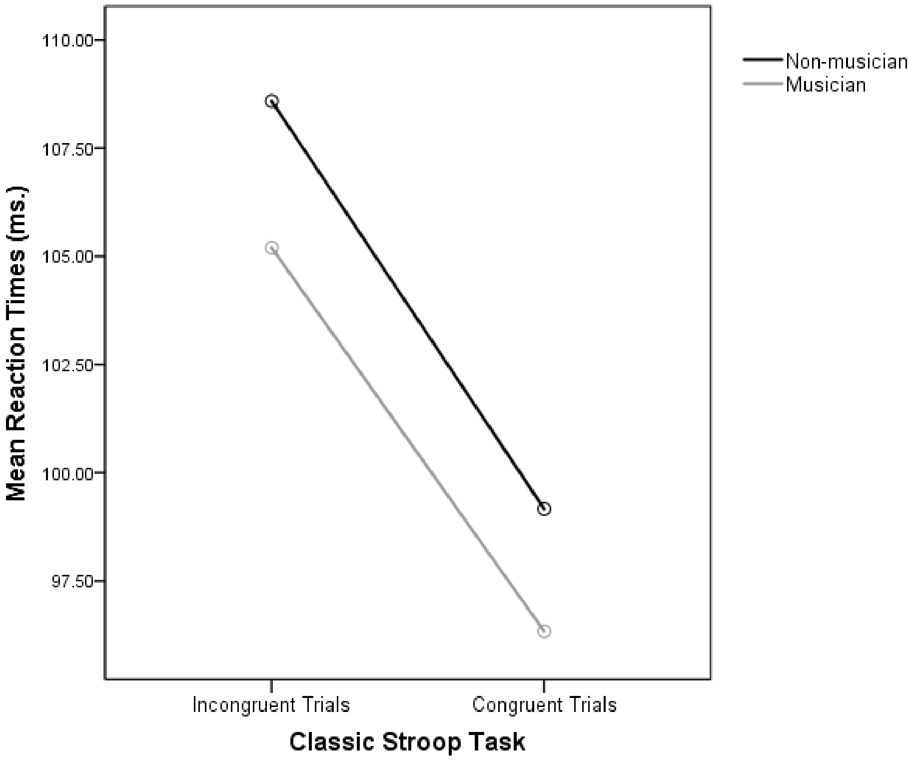

Figure 38. Interaction effect of reaction times according to group of belonging.

7.1.4 Musicians will differ from non-musicians on global measurements of inhibitory control

Previous measurements of inhibitory control have focused only in the reaction time dimension. However, it is an empirical possibility that a combined score of accuracy and reaction time will detect differences between musicians and non-musicians. To explore this assumption a MANOVA test will be conducted.

Since sample sizes were different within each group (musicians = 34; non-musicians = 30), 4 musicians were randomly excluded from the analysis, resulting in a total sample of 60 participants (musicians = 30; non-musicians = 30). Normality tests revealed that mean scores of the music Stroop global interference score and the classic Stroop global interference score derive from a population whose distribution is normal.

Table 6

Normality test of global interference scores of non-musicians and musicians

Group	Global Score	*M*	*SD*	Shapiro-Wilk		
				Statistic	*df*	*p*
Non-musician	GIM	.49	.07	.97	30	.56
	GIC	.74	.18	.99	30	.93
Musician	GIM	.42	.07	.97	30	.56
	GIC	.53	.09	.94	30	.10

Note: GIM = Music Stroop global interference score. GIC = Classic Stroop global interference score.
$^*p < .05$.

A correlation analysis was conducted between the global scores of the music Stroop task and the classic Stroop task. A significant large correlation was obtained, $p < .001$, $r = .57$, thus supporting the decision to continue with a MANOVA test.

Using Pillai's trace, being a non-musician or a musician had an effect on the linear combination of the music Stroop global interference and the classic Stroop global interference, $V = 0.37$, $F(2, 57) = 16.87$, $p < .001$, $\eta p^2 = .37$, $1 - \beta = 1.00$. Therefore, the mean global score differences presented in Table 6 are significantly different. The group of non-musicians had a better performance in inhibitory control when compared to musicians when ignoring their linear combination.

This analysis was followed up with a discriminant analysis for two reasons. First, to explore whether there is a common variate which explains the group differences of the global interference scores of the music and classic Stroop tasks. Second, to explore whether this variate can accurately distinguish musicians from non-musicians. The analysis revealed one discriminant function which explained 100% of the variance, $\eta p^2 = .37$. This discriminant function significantly differentiated the group of non-musicians from the group of musicians, $\Lambda = .63$, $\chi^2(2) = 26.50$, $p < .001$. The correlation between the outcomes and the discriminant function revealed that the classic Stroop global interference scores, $r = .97$, and the music Stroop global interference score, $r = .65$, loaded heavily onto the same variate. Classification results showed that for non-musicians, there were 24 occasions when they were correctly predicted as belonging to the non-musicians' group (80% of cases correctly classified). Conversely, for musicians, there were 28 occasions when they were correctly predicted as belonging to the musicians' group (93.3% of cases correctly classified). Therefore, the variant (presumably inhibitory control) can be used to discriminate musicians from non-musicians effectively.

7.1.5 Discussion of Research Question 1

The first research question aimed at evaluating if a music Stroop task and a classic Stroop task could discriminate the inhibitory control performance of musicians and non-musicians. Since new adaptations of previous inhibitory control tasks were designed specifically for this study, it was necessary to test their psychometric properties before addressing the second research question.

It was found that on the music Stroop task, no difference was found between congruent and incongruent conditions for non-musicians. Because in non-musicians music reading is not automated they do not need inhibitory control to "interrupt, or abort ongoing processes, especially when those processes are well engrained" (Banich & Depue, 2015, p. 17). This claim has been supported long before, even by Stroop himself although with a slightly different terminology. Stroop (1935) explained the effect he discovered on the grounds of the strength of association between stimuli and response: "the associations that have been formed between the word stimuli and the reading response are evidently more effective than those that have been formed between the color stimuli and the naming response" (pp. 659-660). Therefore, it was expected that non-musicians would not experience any inhibition by responding to the congruent and incongruent trials of the music Stroop task because they have not formed the associations needed for music reading. The support for this hypothesis adds discriminant support to the music Stroop task here designed.

Conversely, it was expected that musicians would differ on their performance of the congruent and incongruent conditions of the music Stroop task. Data supported this prediction, and one reason which could explain the large effect detected is the automatization of music reading resulting from extensive musical training. Since it was carefully controlled that non-musicians did not read music, the observed differences between the group of musicians and non-musicians could be attributed to familiarity with music notation. One more relevant consequence can be extracted. Although this study's design did not replicate Grégoire et al. (2013) design, it was heavily based on their music Stroop task. Therefore, the empirical support for the first two hypotheses previously discussed strengthens Grégoire et al.'s (2013) methodology. Altogether, the lack of significant differences in reaction time between congruent and incongruent trials on the group of non-musicians, and the opposite scenario with the group of musicians, reinforces the validity evidence of this study's methodology.

The next aim was to test the hypothesis which stated that non-musicians and musicians would differ on their performance of congruent and incongruent conditions of the classic Stroop task. This was partially rejected. Musicians were slightly faster than non-musicians on the reaction time difference between the congruent and incongruent trials. However, the difference did not reach statistical significance, thus failing to detect any of the documented advantage of inhibitory control of musicians over non-musicians (Bialystok & DePape, 2009; Janus et al., 2016; Joret et al., 2016). Nevertheless, just as the well-established literature on Stroop tasks suggest (MacLeod, 1991; Stroop, 1935), participants replied more slowly to incongruent trials when compared to congruent ones in the classic Stroop task. Interestingly, the effect was present even when using a motor response format for this task, instead of the more extended oral response modality. MacLeod (1991) and Zakay (2014) provide a tentative explanation for this negative result. By changing the response format from an oral to a motor modality then the close interrelation between two dimensions that clash perceptually was weakened. Therefore, the experimental design lost power to detect any relevant statistical significance. In fact, Figure 38 suggest that musicians did have a faster reaction time than non-musicians, although this difference between groups was not statistically significant. Therefore, future studies should test the oral response modality of the classic Stroop task.

So far, hypotheses were tested using reaction times as dependent variables, thus excluding the accuracy of participants from the analysis. The accuracy dimension was excluded given the biased behaviour of the data and because some conceptual inconsistencies were detected. That is, sometimes participants evidenced more mistakes in the congruent sessions than in the incongruent sessions. This deviance from the empirically expected results could be explained by one limitation in this research design. Although participants were required to answer as fast and accurately as possible, stimuli were presented without a time limit. Setting a time limit for the presentation of each stimuli could have changed the outcome of the accuracy dimension.

In order to have a clearer picture of the behaviour of inhibitory control, besides presenting separate results of accuracy and reaction time, a calculation of a compound score based on these two variables was included following Scarpina and Tagini recommendations (2017). The psychometric properties of the music and the classic Stroop tasks were tested once again by testing whether musicians differed from non-musicians on global measurements of inhibitory control.

This hypothesis was accepted, although with an unexpected twist which limits its validity. Mean differences on the linear combination of the classic and music Stroop compound score tasks were significantly different, meaning that being a musician or a non-musician affected the outcome of the combination of the compound scores. However, very unexpectedly, non-musicians had a better performance than musicians in these scores. This finding also contradicts previous EF literature which suggest that musicians present enhanced performance than other populations in EF tasks (Bialystok & DePape, 2009; Bugos et al., 2007; Janus et al.,2016; Joret et al., 2016; Moreno et al., 2014; Zuk et al., 2014). Perhaps calculating a compound score using the accuracy dimension decreased the reliability of the global scores of the music and classic Stroop tasks.

Nonetheless, the discriminant analysis that followed up this MANOVA gave relevant insights into the psychometric properties of the experiments designed. A single discriminant function or factor explained 37% of the changes in scores of the combination of both global interference scores. Based on the hypotheses previously tested and the existing literature there is sufficient grounds to name this factor as "inhibitory control". The discriminant analysis also revealed that the global interference score of the classic Stroop (r = .97) loaded more strongly than the global interference score of the music Stroop (r = .65) into the inhibitory control factor. Therefore, both tasks seems to be strong measurements of inhibitory control. More interestingly, the fact that the classic Stroop correlation was the strongest coincides with previous literature supporting the robustness of the classic Stroop task as a measurement of inhibitory control. Additionally, having found a strong correlation between the global score of the music Stroop task and the inhibitory control factor suggests that the former is indeed a solid measurement of the latter and legitimises its experimental design. Finally, because the discriminant analysis classified participants as musicians or non-musicians based on the linear combination of the composite scores with a high rate of accuracy, there is additional evidence that inhibitory control behaves differently depending on whether someone is a musician or not. However, it is worth remembering that the MANOVA model showed that non-musicians had a better inhibitory control performance than musicians, therefore it is possible that these unexpected and, up to some degree, contradictory findings are due to methodological limitations.

Fulfilled predictions regarding the music Stroop task could be attributed to modifications of the task which followed recommendations from previous authors. As suggested by Akiva-Kabiri and Henik (2014), different notation systems and clefs were introduced so that every participant, regardless of what their main instrument was, could find the task as familiar as possible, as Moeller and Frings (2014) correctly pointed out that notes are usually processed as complex motoric responses and not as oral reading of notes. Therefore, a motor response format of the music Stroop task was implemented with the aim of replicating as much as possible the daily life activity of musicians.

Zakay (2014) was particularly critical to Grégoire et al.'s (2013) music Stroop task design. In relation to Moeller and Frings' (2014) critique, Zakay stated that Gregoire et al.'s music Stroop task was not as good as the classic Stroop task, because a real Stroop task is defined by a close connection between two discordant perceptual dimensions. Thus, to overcome this critique and increase the chances of finding relevant experimental effects, both a music Stroop task and a classic Stroop task were included in this experiment. Scarpina and Tagini (2017) also pointed out the relevance of reporting accuracy, speed and composite scores when testing a Stroop task. This advice was gathered and used in both the music and classic Stroop tasks.

The experimental design of the music Stroop task used in this study also followed what previous music Stroop task studies had done. For instance, Zakay and Glicksohn (1985) gathered speed and accuracy measures. Stewart et al. (2004) implemented their response options in a computer keyboard to keep it as close as possible to the usual way in which musicians produce music. This approach was replicated in the present research. Stewart et al. (2004) also introduced training sessions' stimuli and used a middle fixation point. Pseudorandomizing of trials and calculations of response time and accuracy were also inspired by them. Akiva-Kabiri and Henik (2012) and Grégoire et al. (2013) also included measurements of accuracy and speed.

Grégoire et al. (2013) provided much of the ideas for the design of this study. Among the different elements, fixation cross on the centre of the screen, the immediate repetition of note locations, note names and display on the screen were included in this study.

Other recommendations from these authors were addressed. For instance, participants differed only in their amount of musical training from the general population, many stimuli were presented per condition, and stimuli presentation was counterbalanced and pseudorandomized. Additionally, item-by-item mode of presentation was used, only incongruent and congruent trials were presented, and smaller samples of note positions were used.

On the other hand, the classic Stroop task was adapted into a motor response format to keep the response format consistent with the music Stroop task. Although using a motor response format has been associated with less statistical power (MacLeod, 1991), results showed that the experimental manipulation did produce a noticeable effect. Nonetheless, an important limitation of the motor implementation is that some participants reported using their peripheral vision during the task, thus keeping the word out of their reading focus but still being able to perceive the colour in the screen. Therefore, matching the colour with the corresponding colour key on the keyboard became easier because no interference was produced. Participants were encouraged to keep the colour words presented on the screen at a readable height. Thus, this limitation could have diminished potential inhibitory control effects. Still, significant results were gathered despite the limitations, but future studies should keep the oral response format of the classic Stroop task.

In sum, the design of the music and classic Stroop tasks followed previous similar studies and tried to maximize the ecological validity of the tasks, taking into consideration the daily-life demands of music activities. Thus, there is strong evidence to suggest that performances on a music Stroop task, and on a classic Stroop task discriminate the inhibitory control performance of musicians and non-musicians. Therefore the use of the scores derived from both experimental tasks as a measurement of inhibitory control, are empirically supported. Meaning, any differences in inhibitory control detected will most certainly reflect actual changes in inhibitory control. Thus, after justifying the use of the Stroop tasks as measurements of inhibitory control, the next step is to predict whether transposition skills of musicians can predict changes in inhibitory control.

7.2 Do existing transposition skills impact inhibitory control task performance?

7.2.1 Stablishing transposition skill levels

Normality analysis of reaction times of the non-transposed playing condition, Shapiro-Wilk(31) = 0.89, p = .003, and the transposed playing condition, Shapiro-Wilk(31) = 0.88, p = .002, showed a significant difference from the normality assumption.

Table 7

Normality tests of non-transposed and transposed playing conditions

Playing condition	*M*	*SD*	Shapiro-Wilk		
			Statistic	*df*	*p*
Non-Transposed	4.74	2.05	.89	31	< .01**
Transposed	9.66	3.55	.88	31	< .01**

**p < .01.

However, based on the central limit theorem (Field, 2009), normality was assumed. A paired-sample *t*-test was conducted to compare the mean difference between the non-transposing task condition and the transposing task condition. There was a significant difference between the non-transposed condition (*M* = 4.75; *SD* = 2.10) and the transposed condition (*M* = 9.67; *SD* = 3.57), $t(30) = -12.26$, $p < .001$, $d = -1.67$, $1 - \beta = 1.00$.

A transposition level skill score for every participant was obtained by subtracting the time of the non-transposed condition out of the time of the transposed condition. Normality analysis of the transposition level skill suggest that scores originate from a population whose distribution is normal, Shapiro-Wilk(31) = 0.96, *p* = .26.

Table 8

Normality test of transposing skill mean score

	M	*SD*	Shapiro-Wilk		
			Statistic	*df*	*p*
Transposition skill	4.84	2.02	.96	31	.26

*$p < .05$.

Scores were sorted in an ascending order. Participants were divided into three transposing skill levels: high level (*n* = 10), mid-level (*n* = 11) and low level (*n* = 10). Given the uneven amount of musicians (*N* = 31), one was randomly assigned to one of the three groups.

Mean scores of transposition time for low transposing skill, *M* = 7.23, *SD* = 1.29, Shapiro-Wilk(10) = 0.85, *p* = .06, and for mid transposing skill musicians, *M* = 4.54, *SD* = 0.78, Shapiro-Wilk(8) = 0.95, *p* = .66, suggest that scores originate from a population whose distribution is normal. Nonetheless, mean scores of transposition time for high transposition skills significantly differed from normality, *M* = 2.81, *SD* = 0.71, Shapiro-Wilk(10) = 0.81, *p* = .02. However, normality was assumed based on the central limit theorem (Field, 2009).

Residuals of transposition time for low transposing skill, Shapiro-Wilk(10) = 0.85, *p* = .06, and for mid transposing skill musicians, Shapiro-Wilk(8) = 0.95, *p* = .67, suggest that scores originate from a population whose distribution is normal. However, residual scores of transposition time for high transposition skills significantly differed from normality, Shapiro-Wilk(10) = 0.81, *p* = .02. However, normality was assumed based on the central limit theorem (Field, 2009).

To test whether the transposition time differences among the three transposing skill levels proposed were significantly different to each other a one-way ANOVA was conducted. A Leven's test of homogeneity of variance revealed that variances between the three transposition skills were significantly different, $F(2, 28) = 3.73$, $p = .04$.

Differences in transposition time differed significantly among the three transposing skill levels, $F(2, 28) = 54.20$, $p < .001$, $\omega = .88$, $1 - \beta = 1.00$. A significant linear trend $F(1, 28) = 106.57$, $p < .001$, $\omega = .88$, $1 - \beta = 1.00$, indicating that as the level of transposition increased, reaction time in the transposition task decreased proportionately.

Table 9

Descriptive statistics of transposition skill level groups

Transposition skill	*N*	*M*	*SD*
High	10	2.82	0.71
Middle	11	4.54	0.78
Low	10	7.23	1.29

7.2.2 Musicians with higher transposition skills will outperform musicians with lower transposition skills in the music Stroop task

Two sets of normality test were undertaken. The first one included the accuracy scores of the music Stroop task. Only the group of low transposition skills showed evidences of normality both in the incongruent, Shapiro-Wilk(10) = .86, $p = .07$ and congruent trials, Shapiro-Wilk(10) = .84, $p = .05$. See a summary of results in Table 10.

Table 10

Normality tests of music Stroop accuracy according to transposing level

Accuracy	Transposing level	*M*	*SD*	Shapiro-Wilk		
				Statistic	*df*	*p*
Correct answers in incongruent trials	Low	123.70	1.34	.86	10	.07
	Middle	121.27	3.93	.85	11	.04*
	High	120.40	3.89	.82	10	.03*
Correct answers in congruent trials	Low	123.80	1.32	.84	10	.05
	Middle	123.45	2.98	.59	11	< .001***
	High	122.80	3.08	.74	10	< .01**

*$p < .05$. *$p < .01$. ***$p < .001$.

The second one included the reaction time scores of the music Stroop task. Normality was reported in the means of the three transposition level groups.

Table 11

Normality tests of music Stroop task reaction time according to transposing level

Reaction time	Transposing level	*M*	*SD*	Shapiro-Wilk		
				Statistic	*df*	*p*
Reaction time in incongruent trials	Low	138.27	19.06	.93	10	.48
	Middle	130.36	15.03	.91	11	.23
	High	130.07	7.86	.96	10	.77
Reaction time in congruent trials	Low	128.68	20.38	.90	10	.22
	Middle	113.76	13.49	.95	11	.68
	High	116.97	7.98	.91	10	.27

$^*p < .05$.

A one-way mixed repeated-measures ANOVA showed significant large effect of the type of trial (i.e. congruent vs. incongruent) on the reaction time in the music Stroop task independently of the level of transposing skill of participants, $F(1, 28) = 80.49$, $p < .001$, $r = .86$, $1 - \beta = 1.00$. There was a non-significant main effect of transposition level on the reaction time, $F(2, 28) = 1.96$, $p = .16$, $r = .26$, $1 - \beta = .37$.

7.2.3 Musicians with higher transposition skills will outperform musicians with lower transposition skills in the classic Stroop task

Two sets of normality test were undertaken. The first one included the accuracy scores of the classic Stroop task. All the group scores in the congruent trial differed significantly from normality. See a summary of results in Table 12.

Table 12

Normality tests of classic Stroop accuracy according to transposition level

Accuracy	Transposing level	*M*	*SD*	Shapiro-Wilk		
				Statistic	*df*	*p*
Correct answers in incongruent trials	Low	123.50	1.43	.89	10	.15
	Middle	121.45	3.20	.91	11	.23
	High	121.10	3.57	.92	10	.39
Correct answers in congruent trials	Low	123.90	1.20	.82	10	.03*
	Middle	123.36	2.25	.70	11	< .001***
	High	123.40	1.71	.81	10	< .02**

$^*p < .05$. $^{**}p < .01$. $^{***}p < .001$.

The second one included the reaction time scores of the classic Stroop task. Scores from the low transposition skills in the congruent trial differed significantly from normality, Shapiro-Wilk(10) = .78, p = .01.

Table 13

Normality test of classic Stroop task reaction time according to transposing level

Reaction time	Transposing level	*M*	*SD*	Shapiro-Wilk		
				Statistic	*df*	*p*
Reaction time in incongruent trials	Low	109.87	21.59	.89	10	.15
	Middle	100.04	11.91	.91	11	.23
	High	109.05	11.98	.92	10	.39
Reaction time in congruent trials	Low	102.62	22.51	.78	10	.01*
	Middle	91.96	11.96	.97	11	.87
	High	98.28	8.01	.96	10	.81

*p < .05.

There was a significant large effect of the type of trial (i.e. congruent vs. incongruent) on the reaction time in the classic Stroop task independently of the level of transposing skill of participants, $F(1, 28) = 126.92$, $p < .001$, $r = .91$, $1 – \beta = 1.00$. There was a non-significant main effect of transposition level on the reaction time variables, $F(2, 28) = 1.85$, $p = .18$, $r = .25$, $1 – \beta =.35$.

7.2.4 Musicians with higher transposition skills will outperform musicians with lower transposition skills in global inhibitory control scores

Similarly as in the first research question, previous analysis have focused on the reaction time dimensions of the music and classic Stroop tasks. However, composite scores which take into account both accuracy and reaction time could show different results. A MANOVA will be conducted to test this assumption.

One participant from the middle-level transposition skill was randomly excluded from the analysis in order to have equal sample sizes in the three groups. A test of normality was conducted on the mean scores of the global interference scores of the music and classic Stroop. All the group showed evidences of normality.

Table 14

Normality test for global interference scores according to transposing level

	Transposing level	*M*	*SD*	Shapiro-Wilk		
				Statistic	*df*	*p*
GIM	Low	.42	.06	.92	10	.32
	Middle	.39	.07	.96	10	.82
	High	.40	.04	.96	10	.72
GIC	Low	.53	.09	.88	10	.12
	Middle	.56	.06	.95	10	.63
	High	.49	.09	.96	10	.81

Note: GIM = Music Stroop global interference score. GIC = Classic Stroop global interference score.
$*p < .05$.

Using Pillai's trace it was showed that the level of transposition did not have an effect on the music Stroop global interference and the classic Stroop global interference, $V = 0.14$, $F(4, 54) = .99$, $p = .42$, $\eta p^2 = .07$, $1 - \beta = .29$.

7.2.5 Discussion of research question 2

The second research question was aimed at evaluating if existing transposition skills impact upon inhibitory control task performance. To investigate this, it was first necessary to test the suitability of the transposition task to discriminate different levels of transposition. Results suggest that the play-as-written and play-transposed conditions differed significantly in their scores. As expected, transposed trials took a longer time to be played than non-transposed trials. The additional ANOVA test supported the division of transposition skills into low, middle and high levels. Reported means followed the expected pattern: as reaction time decreased, transposition skills increased.

One limitation of this design is that there was no complementary measurement of transposition skills which could provide us with validity evidence based on the relation with other variables. However, at least conceptually it seems that the difference in reaction times between the conditions can be attributed to the transposition skills of the musicians. Having gathered evidence to support the use of the transposition task as a way of determining different transposing levels, and the use of the music and classic Stroop tasks to measure inhibitory control, the second research question could be fully addressed.

It was found that musicians with higher transposition skills did not outperform musicians with lower transposition skills in the music Stroop task. The accuracy dimension of the music Stroop task still showed statistical biases which could compromise any inferential analysis even when dividing the sample according to the level of transposition. Therefore, as with the first research question procedure, only the reaction time dimension was considered to test this hypothesis. Once again, through the music Stroop task it was possible to detect an overall difference between incongruent and congruent trails, thus adding empirical support to this task as a measure of inhibitory control. However, none of the transposition level groups distinguished from each other significantly. That is, participants of all transposing levels performed similarly in the music Stroop task.

A similar scenario was found when testing the second hypothesis. Musicians with higher transposition skills did not outperform musicians with lower transposition skills in the classic Stroop task. Again, a difference between the incongruent and congruent conditions was detected when ignoring the transposition level of participants. Nevertheless, the performance on the Stroop task was similar across all skill groups. That is, participants of all transposing levels performed similarly in the classic Stroop task.

Similarly, musicians with higher transposition skills did not outperform musicians with lower transposition skills in global inhibitory control scores. Not even the linear combination of the music and classic Stroop global interference scores all found an effect of the transposition level. Reasons that could explain the failure in predicted results are unclear. Perhaps the method for establishing the three transposition levels was not the most solid choice. Sorting the sample in ascending order and then assigning the group based on an equal division of the sample into three levels, might not have been the best approach for discriminating levels. An alternative might have been dividing the participants according to quartiles, but that would have resulted in fewer participants per group (and therefore reduced statistical power) and greater complexity in the interpretation of statistical tests.

Audio recording techniques may have been another contributor to measurement error. The “mironset” function of the MIRToolbox failed to recognize some onsets on the soundwaves due to recording issues such as the volume of the computer, the distance with the instrument, etc. Future studies should better control the audio recording settings to optimize data collection. Still, the introduction of the transposition task was an ambitious project which needs further empirical testing. Forthcoming psychometric research could test the properties of this task by correlating it with other existing, or to-be-designed transposing tasks and test different methods to undercover its measuring properties. Still, introducing this measurement paradigm is in itself a valuable contribution to music psychology research.

In sum, no sufficient evidence was found to state that existing transposing abilities impact inhibitory control task performance. Although the difference between incongruent and congruent conditions was present across all hypothesis testing, belonging to a particular transposition skill group did not translate into a significantly different performance of inhibitory control.

8 GENERAL DISCUSSION

This research investigated whether transposition skills had an impact on inhibitory control performance. To answer this question, two main empirical approaches were followed: first, to test the psychometric properties of the music and classic Stroop tasks; and second, to test whether existing transposition skills impacted inhibitory control task performance. Although strong evidence in favour of using the music and classic Stroop tasks as measurements of inhibitory control were gathered, no sufficient evidence was obtained to claim that transposition skill has an effect on inhibitory control performance.

Transposition was defined as "the notation or performance of music at a pitch different from that in which it was originally conceived, by raising or lowering the notes in it by the same interval" (the new Grove dictionary of music and musicians, 1980, p. 121). This complex cognitive process has not been investigated with the tools and lenses of EF literature and particularly with one of its core components, inhibitory control. Diamond (2013) defined inhibitory control as the executive function which "involves being able to control one's attention, behaviour, thoughts, and/or emotions to override a strong internal predisposition or external lure, and instead do what's more appropriate or needed" (p. 137). Banich and Depue (2015) added that to inhibit is to "interrupt, or abort ongoing processes, especially when those processes are well engrained" (p. 17).

Results from this study failed to gather enough empirical evidence that could connect music transposition and inhibitory control. However, arguably this research also sets a precedent of attempting to explore neuroplasticity and the flexible nature of EF within the context of a very specific musical process. Moreover, this attempt also introduced ambitious methodological designs holding ecological validity as a guiding principle. Naturally, several limitations were present and they have been addressed previously in greater detail. Some final more-general-scope limitations are worth mentioning before concluding.

Perhaps one of the first limitations with relevant implications for further research fields in music psychology is sorting out musicians from non-musicians. In this study an attempt to solve this problem was taken by classifying participants based on more than one discriminant criteria. This shows that identity as a musician cannot be limited to a single indicator (e.g. whether someone self-identifies as a musician, whether they play a music instrument, etc.). Figuring out what is music identity is crucial for future studies that wish to compare musicians against non-musicians.

Generalization of this study's results is limited. Given the non-randomized sampling used and the presence of deviations from normality in different measured variables, it is not possible to extend the findings of the study outside of the collected sample. However, the methodological design can provide some ideas worth further exploration. For example, would replicating the study with music participants of different music skills produce different outcomes? That is, will amateur musicians, professional music students and professional musicians with different transposition backgrounds evidence different results in inhibitory control? Would using different inhibitory control tasks to study transposition will shed new lights on neuroplasticity and plasticity of EF? How could the transposition task proposed here be improved in the future? Although not all predictions were fulfilled, this study opened a new research window in the field of EF. Namely, it explored the relationship between music transposition and inhibitory control using the methodological tools provided by psychology. What is more, this study exemplified the relevance that a highly interdisciplinary field such as music psychology has in understanding how our mind works.

9 CONCLUSIONS

- There are no enough evidences to claim that existing transposing abilities impact upon inhibitory control task performance.
- Performances on a music Stroop task and a classic Stroop task discriminate fairly well inhibitory control performance of musicians and non-musicians, although excluding accuracy measures.
- Musicians differed from non-musicians on global measurements of inhibitory control. Although it was found that non-musicians had better performance at this particular measurement when compared to musicians. Contradictory results could be due to methodological limitations.
- Stroop tasks should report measurements of accuracy, reaction time and global composite scores.
- The transposition task created for this study showed clear differences between play-as-written and transposed conditions. However, the way participants were sorted according to their transposing levels failed to predict expected results.

10 REFERENCES

Adler, S. (2002). *The study of orchestration* (3rd ed.). New York: W.W. Norton.

Aguinis, H., Gottfredson, R. K., & Joo, H. (2013). Best-practice recommendations for defining, identifying, and handling outliers. *Organizational Research Methods, 16*(2), 270-301. doi: 10.1177/1094428112470848

Akiva-Kabiri, L., & Henik, A. (2012). A unique asymmetrical Stroop effect in absolute pitch possessors. *Experimental Psychology, 59*(5), 272-278. doi:10.1027/1618-3169/a000153

Akiva-Kabiri, L., & Henik, A. (2014). Additional insights. Commentary on "the musical Stroop effect: Opening a new avenue to research on automatisms" by L. Grégoire, P. Perruchet, and B. Poulin-Charronnat (experimental psychology, 2013, vol. 60, pp. 269–278). *Experimental Psychology, 61*(1), 75-77. doi:10.1027/1618-3169/a000208

Ardila, A., Pineda, D., & Rosselli, M. (2000). Correlation between intelligence test scores and executive function measures. *Archives of Clinical Neuropsychology, 15*(1), 31-36. doi: 10.1016/S0887-6177(98)00159-0

Baddeley, A. D., & Hitch, G. J. (1994). Developments in the concept of working memory. *Neuropsychology, 8*(4), 485-493. doi:10.1037//0894-4105.8.4.485

Bangert, M., & Schlaug, G. (2006). Specialization of the specialized in features of external human brain morphology. *European Journal of Neuroscience, 24*, 1832-1834.

Banich, M. T., & Depue, B. E. (2015). Recent advances in understanding neural systems that support inhibitory control. *Current Opinion in Behavioral Sciences, 1*, 17-22. doi:10.1016/j.cobeha.2014.07.006

Bialystok, E., & DePape, A. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 565-574. doi: 10.1037/a0012735

Bugos, J. A., Perlstein, W. M., McCrae, C. S., Brophy, T. S., & Bedenbaugh, P. H. (2007). Individualized piano instruction enhances executive functioning and working memory in older adults. *Aging & Mental Health, 11*(4), 464-471. doi: 10.1080/13607860601086504

Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., . . . Shoda, Y. (2011). From the cover: Behavioral and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences, 108*(36), 14998-15003. doi:10.1073/pnas.1108561108

Castellanos, F. X., Blumenthal, J., Liu, H., Rapoport, J. L., Jeffries, N. O., Paus, T., . . . Giedd, J. N. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience, 2*(10), 861-863. doi: 10.1038/13158

Costa-Giomi, E. (2014). The long-term effects of childhood music instruction on intelligence and general cognitive abilities. *Update: Applications of Research in Music Education*, 1-7. doi: 10.1177/8755123314540661

Cragg, L. (2016). The development of stimulus and response interference control in midchildhood. *Developmental Psychology, 52*(2), 242-252. doi: 10.1037/dev0000074

Cragg, L., & Nation, K. (2008). Go or no-go? Developmental improvements in the efficiency of response inhibition in mid-childhood. *Developmental Science, 11*(6), 819-827. doi:10.1111/j.1467-7687.2008.00730.x

Cramon, D. Y. v., Brass, M., & Derrfuss, J. (2004). Cognitive control in the posterior frontolateral cortex: Evidence from common activations in task coordination, interference control, and working memory. *Neuroimage, 23*(2), 604-612. Retrieved from http://repository.ubn.ru.nl/handle/2066/139547

Davidson, M. C., Amso, D., Anderson, L. C., & Diamond, A. (2006). Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia, 44*(11), 2037-2078. doi:10.1016/j.neuropsychologia.2006.02.006

Diamond, A. (2013). Executive functions. *Annual Review of Psychology, 64*(1), 135-168. doi: 10.1146/annurev-psych-113011-143750

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica, 30*(1), 412-431. doi: 10.1016/0001-6918(69)90065-1

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143-149. doi: 10.3758/BF03203267

Eriksen, C. (1995). The flankers task and response competition: A useful tool for investigating a variety of cognitive problems. *Visual Cognition, 2*(2), 101-118. doi: 10.1080/13506289508401726

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.).London: Sage.

Gast, A. (2014). What is learned, and when? Commentary on “the musical Stroop effect: Opening a new avenue to research on automatisms” by L. Grégoire, P. Perruchet, and B.

Poulin-Charronnat (experimental psychology, 2013, vol. 60, pp. 269–278). *Experimental Psychology, 61*(1), 71-74. doi:10.1027/1618-3169/a000206

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2013). The musical Stroop effect. *Experimental Psychology, 60*(4), 269-278. doi:10.1027/1618-3169/a000197

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2014). Is the musical Stroop effect able to keep its promises? A reply to Akiva-Kabiri and Henik (2014), Gast (2014), Moeller and Frings (2014), and Zakay (2014). *Experimental Psychology, 61*(1), 80-83. doi:10.1027/1618-3169/a000222

Grégoire, L., Perruchet, P., & Poulin-Charronnat, B. (2015). How does Stroop interference change with practice? A reappraisal from the musical Stroop paradigm. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 41*(2), 417-425. doi: 10.1037/xlm0000044

Haahr, M. (2006). Random.org: True random number service [web resource]. Retrieved from https://www.random.org/

Herholz, S. C., & Zatorre, R. J. (2012). Musical training as a framework for brain plasticity: Behavior, function and structure. *Neuron, 76,* 486-502. http://dx.doi.org/10.1016/j.neuron.2012.10.011

Hofmann, W., Schmeichel, B. J., & Baddeley, A. D. (2012). Executive functions and self-regulation. *Trends in Cognitive Sciences, 16*(3), 174-180. doi:10.1016/j.tics.2012.01.006

Hommel, B. (2011). The Simon effect as tool and heuristic. *Acta Psychologica, 136*(2), 189-202. doi:10.1016/j.actpsy.2010.04.011

Hutton, S. B., & Ettinger, U. (2006). The antisaccade task as a research tool in psychopathology: A critical review. *Psychophysiology, 43*(3), 302-313. doi:10.1111/j.1469-8986.2006.00403.x

Janus, M., Lee, Y., Moreno, S., & Bialystok, E. (2016). Effects of short-term music and second-language training on executive control. *Journal of Experimental Child Psychology*, *144*, 84-97. doi:10.1016/j.jecp.2015.11.009

Joret, M., Germeys, F., & Gidron, Y. (2016). Cognitive inhibitory control in children following early childhood music education. *Musicae Scientiae,* 1-13. doi: 10.1177/1029864916655477

Kunert, R., Willems, R. M., Casasanto, D., Patel, A. D., & Hagoort, P. (2015). Music and language syntax interact in Broca's area: An fMRI study. *PLoS ONE*, *10*(11), e0141069. doi:10.1371/journal.pone.0141069

Lartillot, O., Toiviainen, P., & Eerola, T. (2011). A Matlab Toolbox for Music Information Retrieval. In C. Preisach, H. Burkhardt, L. Schmidt-Thieme, & R. Decker (Eds.), *Data Analysis, Machine Learning and Applications, Studies in Classification, Data Analysis, and Knowledge Organization* (pp. 261-268). Switzerland: Springer-Verlag.

Levy, B. J., & Wagner, A. D. (2011). Cognitive control and right ventrolateral prefrontal cortex: Reflexive reorienting, motor inhibition, and action updating. *Annals of the New York Academy of Sciences*, *1224*(1), 40-62. doi:10.1111/j.1749-6632.2011.05958.x

Luna, B. (2009). Developmental changes in cognitive control through adolescence. *Advances in Child Development and Behavior, 37,* 233-278. doi: 10.1016/S0065-2407(09)03706-9

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin, 109*(2), 163-203. doi:10.1037//0033-2909.109.2.163

Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science, 244*(4907), 933-938. doi:10.1126/science.2658056

Mischel, W., Ebbesen, E. B., & Raskoff Zeiss, A. (1972). Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology, 21*(2), 204-218. doi: 10.1037/h0032198

Moeller, B., & Frings, C. (2014). How automatic is the musical Stroop effect? Commentary on "The musical Stroop effect: Opening a new avenue to research on automatisms" by L. Grégoire, P. Perruchet, and B. Poulin-Charronnat (experimental psychology, 2013, vol. 60, pp. 269–278). *Experimental Psychology, 61*(1), 68-70. doi:10.1027/1618-3169/a000204

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, . . . Heckman, J. J. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America, 108*(7), 2693-2698. doi:10.1073/pnas.1010076108

Moreno, S., Bialystok, E., Barac, R., Schellenberg, E., Cepeda, N., & Chau, T. (2011). Short-term music training enhances verbal intelligence and executive function. *Psychological Science, 22*(11), 1425-1433. doi: 10.1177/0956797611416999

Moreno, S. & Farzan, F. (2015). Music training and inhibitory control: A multidimensional model. *Annals of the New York Academy of Sciences, 1337*(1), 147-152. https://doi.org/10.1111/nyas.12674

Moreno, S., Wodniecka, Z., Tays, W., Alain, C., & Bialystok, E. (2014). Inhibitory control in bilinguals and musicians: Event related potential (ERP) evidence for experience-specific effects. *PloS One, 9*(4), e94169. doi:10.1371/journal.pone.0094169

Mullane, J. C., Corkum, P. V., Klein, R. M., & McLaughlin, E. (2009). Interference control in children with and without ADHD: A systematic review of flanker and Simon task performance. *Child Neuropsychology*, *15*(4), 321-342. doi: 10.1080/09297040802348028

Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience, 5*(3), 218-228. doi: 10.1038/nrn1345

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674-681. doi: 10.1038/nn1082

Patel, A. D. (2008). Music and the brain. In S. Hallam, I. Cross & M. H. Thaut, *Oxford handbook of music psychology*. Oxford University Press. doi: 10.1093/oxfordhb/9780199298457.013.0019

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13. https://doi.org/10.1016/j.jneumeth.2006.11.017

Peirce J. W. (2009). Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics. 2*(10). doi: 10.3389/neuro.11.010.2008

Pietschnig, J., Voracek, M., & Formann, A. K. (2010). Mozart effect-Shmozart effect: A meta-analysis. *Intelligence, 38*, 314-323. doi: 10.1016/j.intell.2010.03.001

Rauscher, F. H. (2009). The impact of music instruction on other skills. In S. Hallam, I. Cross & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 244-252). UK: Oxford University Press.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature, 365,* 611. doi: 10.1038/365611a0

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., . . . Taylor, E. (2001). Mapping motor inhibition: Conjunctive brain activations across different versions of go/no-go and stop tasks. *Neuroimage, 13*(2), 250-261. doi: 10.1006/nimg.2000.0685

Scarpina, F., & Tagini, S. (2017). The Stroop color and word test. *Frontiers in Psychology, 8*, 1-8. doi: 10.3389/fpsyg.2017.00557

Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science*, *14(*6), 317-320. https://doi.org/10.1111/j.0963-7214.2005.00389.x

Schellenberg, E. G. (2006a). Exposure to music: the truth about the consequences. In G. E. McPherson (Ed.), *The child as musician: A handbook of musical development* (pp. 111-134). UK: Oxford University Press.

Schellenberg, E. G. (2011).Examining the association between music lessons and intelligence. *British Journal of Psychology, 102*(3), 283-302. https://doi.org/10.1111/j.2044-8295.2010.02000.x

Sibelius Group (Version 8.7) [Computer software]. Ireland: Avid Technology Europe Limited.

Simon, J. R., & Rudell, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology, 51*(3), 300-304. doi: 10.1037/h0020586

Slevc, L. R., Davey, N. S., Buschkuehl, M., & Jaeggi, S. M. (2016). Tuning the mind: Exploring the connections between musical ability and executive functions. *Cognition, 152*, 199-211. doi: 10.1016/j.cognition.2016.03.017

Stanley, S. (Ed.). (1980). *The new Grove dictionary of music and musicians* (Vol. 19). Hong Kong: Macmillan Publishers.

Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception & Psychophysics, 66*(2), 183-195. doi: 10.3758/BF03194871

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18*(6), 643-662. doi: 10.1037/h0054651

Thorell, L. B., Lindqvist, S., Nutley, S. B., Bohlin, G., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*(1), 106-113. doi: 10.1111/j.1467-7687.2008.00745.x

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods, 38*(4), 584-589. Retrieved from http://www.mrc-cbu.cam.ac.uk/personal/maarten.van-casteren/vanCasteren_brm2006.pdf

Verbruggen, F., & Logan, G. D. (2008). Automatic and controlled response inhibition. *Journal of Experimental Psychology: General, 137*(4), 649-672. doi: 10.1037/a0013170

Zakay, D. (2014). Can the "musical Stroop" task replace the classical Stroop task? Commentary on “The musical Stroop effect: Opening a new avenue to research on automatisms” by L. Grégoire, P. Perruchet, and B. Poulin-Charronnat (experimental psychology, 2013, vol. 60, pp. 269–278). *Experimental Psychology, 61*(1), 78-79. doi:10.1027/1618-3169/a000211

Zakay, D., & Glicksohn, J. (1985). Stimulus congruity and S-R compatibility as determinants of interference in a Stroop-like task. *Canadian Journal of Psychology/Revue Canadienne De Psychologie, 39*(3), 414-423. doi: 10.1037/h0080069

Zuk, J., Benjamin, C., Kenyon, A., & Gaab, N. (2014). Behavioral and neural correlates of executive functioning in musicians and non-musicians. *PloS One, 9*(6), e99868. doi: 10.1371/journal.pone.0099868

11 APPENDICES

11.1 Appendix A: Description of main transposing instruments

English horn: The English horn sounds a perfect fifth lower (or a perfect fourth above) than the written note. The written range of the instrument extends from B3 to G6. Hence, the sounding range extends from E3 to C6. Consequently, the middle 5 notes range comprises C5 to G5.

Oboe d´amore: The oboe d´amore sounds a minor third lower than the written note. The written range of the instrument extends from B3 to E6. Hence, the sounding range extends from G#3 to C#6. Consequently, the middle 5 notes range comprises B4 to F5.

Clarinet: A Bb clarinet emits a sound a major second bellow the written note. Its written range extends from E3 to A6. Hence, the sounds it emits extends from D3 to G6. Consequently, the middle 5 notes range comprises A4 to E5.

Meanwhile, an A clarinet emits a sound a minor third bellow the written pitch. Its written range extends from E3 to A6. Hence, the sounds it emits extends from C#3 to F#6. Consequently, the middle 5 notes range comprises A4 to E5.

Saxophones: All saxophones have the same written range which comprises Bb3 to G6, although the better register is between D4 and D6. Hence, the middle 5 notes written range for any kind of saxophone will extend from B4 to F5. Next, I will list the different kind of Saxophones and their intervallic transposition range:

- Eb sopranino = a minor 3rd above written tone.
- Bb soprano = a major second below written tone.
- Eb alto = a major sixth below written tone.
- Bb tenor = a major ninth below written tone.
- Eb baritone = a major thirteenth below written tone.
- Bb bass = two octaves and a major second bellow written tone.

Valve horn: The valve horn is also known as F horn and it will sound a perfect fifth below the written tone. It can either be notated in the treble clef or the bass clef. If a hornist is presented with the former clef, then he or she will transpose a perfect fifth bellow the written tone; whereas if presented with the latter clef, he or she will transpose a perfect fourth above the written tone.

The written range expands from D2 to C6. Hence, the sounding range comprises G1 to F5. Consequently, the middle 5 notes range will extend from B3 to F4.

Valve trumpet: The most common valve trumpet´s found in contemporary use are the trumpet in C or trumpet in Bb. If playing a trumpet in C, then both written and real sound will match. But if playing a trumpet in Bb, then the real sound will be a major second bellow the written tone.

Regardless of the trumpet, its written range expands from F#3 up to D6. Hence, the sounding range comprises E2 to C6. Consequently, the middle 5 notes range includes G4 to D5.

Cornet: The cornet is also tuned in Bb, which means that it has the same range as the trumpet in Bb. Consequently, the middle 5 notes range includes G4 to D5.

Bass trumpet: The bass trumpet is not usually played by trumpeters, but by trombone players. There are bass trumpets in C, Bb, A, and Eb.

The Bb bass trumpet produces real sounds a major ninth below the written tone. Its written range expands from F#3 to C6. Hence, it sounds from E2 until Bb4. Consequently, its written middle 5 note range comprises F4 to C5.

The Eb bass trumpet produces real sounds a major sixth bellow the written tone. Its written range expands from F3 to C6. Hence, its sounding range comprises Ab2 until Eb5. Consequently, its written middle 5 note range includes F4 to C5.

The D bass trumpet produces real sounds a minor seventh bellow the written tone. Its written range expands from F3 until C6. Hence, its sounding range includes G2 to D5. Consequently, its written middle 5 note range includes F4 to C5.

Flugelhorn: The flugelhorn in Bb is transposed a major second bellow the written tone. Its written range expands from F#3 to C6. Hence, its sounding range includes E3 until Bb5. Consequently, its written middle 5 note range includes F4 to C5.

11.2 Appendix B: Initial questionnaire

Sex:
M ( ) F ( )

Age:

Mother language:

1.- Do you self-identify as a musician?
Yes ( ) No ( )

2.- Do you self-identify as bilingual (or multilingual)?
Yes ( ) No ( )

3.- Is music your main source of income?
Yes ( ) No ( )

4.- Do you have absolute pitch?
Yes ( ) No ( )

5.- Do you have colour blindness?
Yes ( ) No ( )

6.- Do you have dyslexia
Yes ( ) No ( )

7.- Do you read music?
Yes ( ) No ( )

8.- Are you currently studying a music degree? If yes, please specify which one.
Yes ( ) No ( ) Programme:

9.- Which is your primary/main musical instrument?
Instrument:
I do not play a musical instrument ( )

10.- At what age did you start to receive instrumental tuition on your primary instrument?
Age:
I have not received any instrumental tuition ( )

11.3 Appendix C

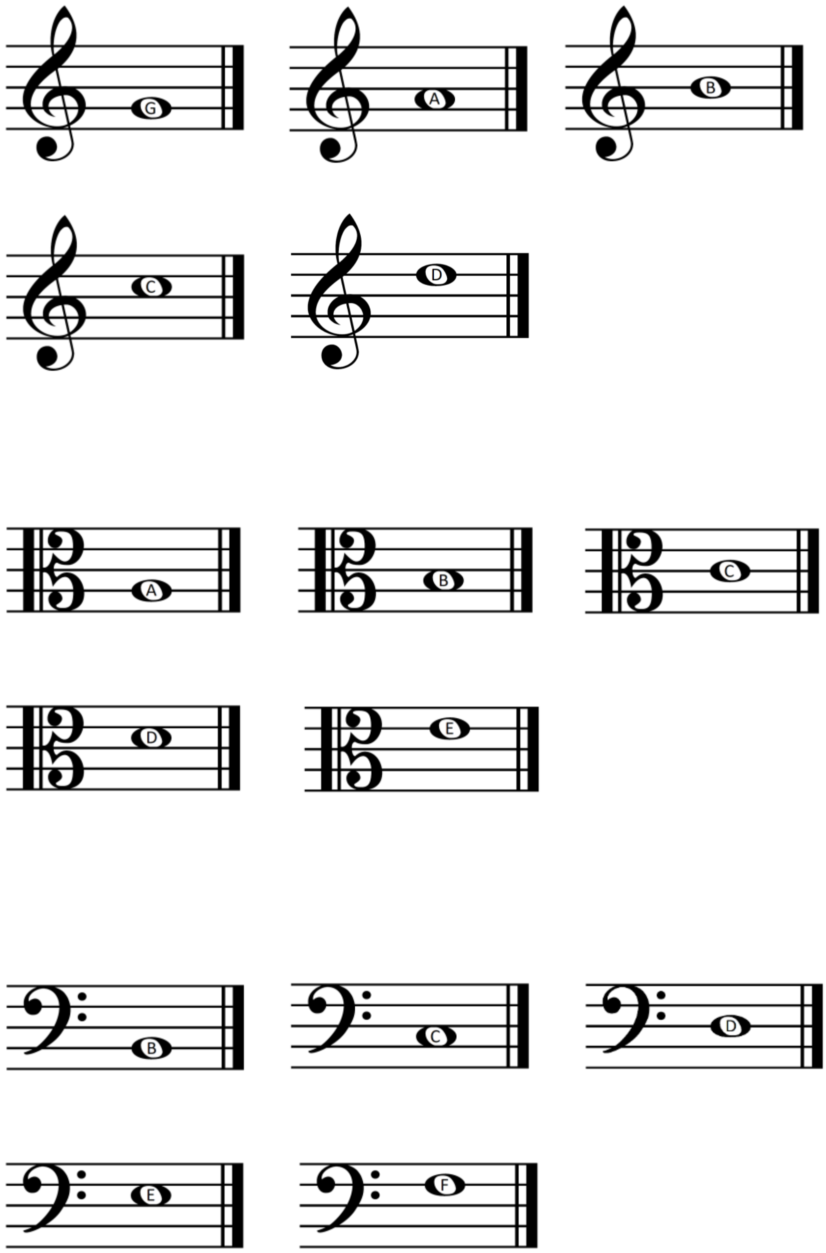

Figure C1. Examples of congruent trials in three clefs and English note naming system. Equivalent stimuli were created for the German and Latin note naming system on the same locations as the five displayed in this figure. Thus, the only difference were the name of the notes.

11.4 Appendix D

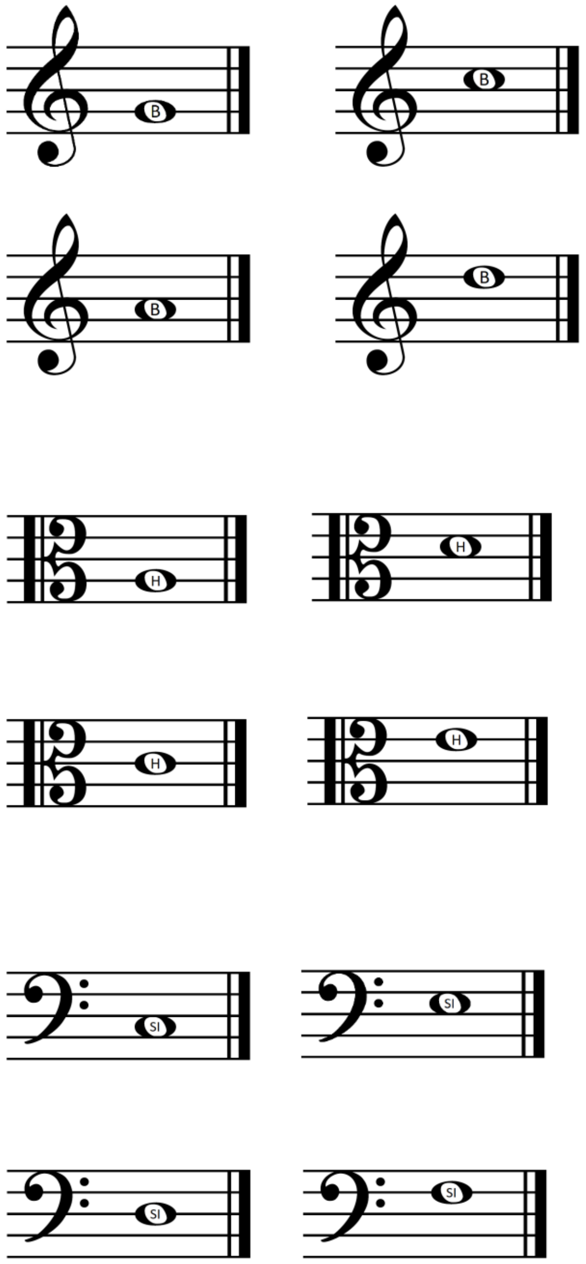

Figure D1. Examples of incongruent trials in three clefs and three note naming systems. Equivalent stimuli were created for the rest of the notes within the lower second line of the pentagram and the fourth line.