Exploring Statistical and Computational Analysis Techniques for Comparison of Beat Perception Test Batteries

Final Progress Report

Neeraja Murali Dharan

Bioinformatics Thesis Project CS4460Z

Department of Computer Science

Western University

April 6th, 2018

Dr. Mark Daley & Dr. Jessica Grahn

Prof. Nazim Madhavji

Abstract - A beat is an underlying, recurring pulse [1]. Humans possess innate cognitive abilities to perceive and process beat in music [1]. Yet, there exists widespread differences in individual beat perception abilities [2]. Identification of such differences is an important topic of study in neuroscience and music research. Such investigations can help us better understand neural underpinnings of beat processing and how various neurological disorders affect these abilities [2]. Thus, researchers have developed numerous beat perception test batteries to classify beat-processing abilities of individuals. Previous studies [2] have highlighted the validity of each of these test batteries individually using factors such as sensitivity measures and performance distributions. However, tests from these batteries have not been compared to measure how scores on different tasks relate to each other. Here we use several commonly used test batteries to investigate how they compare to each other. Moreover, we use machine learning to understand whether these ten tests capture variability in individual beat perception abilities. For this project, ten subtests from four commonly used beat perception test batteries are compared. A behavioural study was conducted with twenty participants. Each participant was asked to perform all ten beat perception tasks in a single session. Our findings highlight that some of these task pairs showed no bias of one task over the other. There were no significant correlations found between these ten tasks. However, they captured variability present in individual beat perception abilities when used in combination. Studies examining individual differences in beat perception abilities use different test batteries, which makes it challenging to compare relevance of results. Findings from this study can help understand how such comparisons can be performed and whether measurements from these tests can contribute to classification of beat perception abilities.

1. INTRODUCTION

A beat is perceived as an underlying, recurring pulse in most Western music [1]. When we listen to a beat, we often engage in spontaneous, timely movements such as head nodding or feet tapping. Rhythmic ability varies widely in the general population and is related to several factors such as short-term memory (STM) capacity, levels of music training, and beat sensitivity [1]. Identification of individual differences in beat-processing and influence of various factors on this ability are important topics of research, as it can help scientists understand roles played by different parts of the brain in response to rhythm [1]. Such research is also highly applicable to different areas of study. For instance, new methods for diagnosing neurological disorders can be developed by studying how areas of the brain involved in perceiving beat are affected by these diseases [3].

It is crucial for researchers to use tools that provide accurate measurements of beat perception abilities when conducting research. For this reason, several beat perception test batteries have been developed. Four commonly used test batteries are the Harvard Beat Assessment Test (H-BAT), Battery for the Assessment of Auditory Sensorimotor and Timing Abilities (BAASTA), the Goldsmiths Musical Sophistication Index (Gold-MSI) and the ‘Speeding Up and Slowing Down’ test from the paper titled ‘Neural bases of individual differences in beat perception’ by Grahn and McAuley [3]. A study often uses one or more of these tests to measure beat perception abilities based on suitability and needs. For instance, a study on behavioural and neuroimaging investigations to determine differences in rhythmic ability used the BAT to assess beat sensitivity and the rhythm reproduction subtest from BAASTA to measure rhythm reproduction ability of individuals [4].

Even though researchers are given the flexibility of choosing from several test batteries for their studies, no study has ever investigated how all of these batteries relate to each other. The variable batteries used across studies lead to discrepancies and it becomes a challenge to compare between studies. Moreover, it is difficult to classify with accuracy which of these tests measure beat perception abilities. This is because researchers have developed these test batteries based on highly educated guesses of factors that measure beat perception. Currently, a test battery that has been proven to measure beat perception with certainty does not exist. Furthermore, these tests have not yet been analysed to determine whether they measure beat perception or other distinguishing abilities of an individual such as working memory. Therefore, performance in beat perception tasks must be compared with performance in tasks measuring other individual abilities to determine whether any correlations exist. Lastly, it is important to analyze whether these tests capture the variability present in individual beat perception abilities. In this paper, we aim to compare four commonly used beat perception test batteries, the BAT, H-BAT, BAASTA and ‘Speeding Up and Slowing Down’ test using an array of statistical and computational analysis methods. The results from this study will provide an understanding of how performances across these tests compare as well as an indication of the generalizability of results from these tasks.

This paper is divided into eight sections: Section 2 outlines some background research associated with the development and use of each test battery. General research objectives of this study as well as their significance are discussed in Section 3. Section 4 defines some important key concepts and terms that will be used throughout this paper. In Section 5, the experimental work carried out in this study is outlined. Section 6 and 7 present the findings of this study and implications of research results. The conclusion is presented in Section 8. Finally, Section 9 highlights some future work that can be conducted as an extension of this study.

2. BACKGROUND WORK

In this section, four beat perception test batteries used in this study (H-BAT, BAASTA, Gold-MSI and Speeding-Up-Slowing-Down test) are discussed in detail, with references made to papers published on each battery.

A. H-BAT

In 2013, Shinya Fujji and Gottfried Schlaug [2] developed the Harvard Beat Assessment Test (H-BAT) battery to assess beat perception and production abilities. This study also aimed to understand whether individuals’ beat perception and production skills are dissociated. It consists of four subtests: music tapping test (MTT), beat saliency test (BST), beat interval test (BIT) and beat finding and interval test (BFIT). MTT is designed to measure degree of tapping synchronization with the beat of music whereas BST, BFIT and BIT measure perception and production thresholds via psychophysical adaptive staircase methods (See Section 4) [2].

This study uses the BST, BFIT, and BIT. The design of BST references research indicating that individuals with difficulty detecting or synchronizing to the beat of musical rhythm also have difficulties in perceptual discrimination of duple and triple meter [2]. The stimuli for this subtest is created using woodblock tones with intensity accents every two or three tones to make either a duple or triple meter [2]. In the first part of BST, participants are asked to discriminate duple and triple meter perceptually. In the second part, they identify a duple or triple meter after tapping along to woodblock tones. The BIT and BFIT subtests are very similar. Both are designed to investigate an individual's ability to detect tempo change in sequences. Participants are asked to indicate whether patterns of tones are speeding up or slowing down by either listening to sequences or by tapping along to them. In BIT, the tones begin isochronously, then speed up or slow down. In BFIT, a repeating rhythmic pattern that is either slowing down or speeding up is presented. BFIT is designed based on findings that musicians are better at identifying beat in a pattern of time intervals compared to non-musicians [2]. A perception and production threshold were calculated for each of the beat perception tasks, according to the staircase paradigm (See Section 4). The paper revealed no significant differences between perception and production thresholds of BST [2]. However, the production thresholds of BIT and BFIT were significantly lower than their corresponding perception thresholds [2]. The authors also found that the lower the production threshold in BST, the lower the perception thresholds in the BIT and BFIT [2]. Lastly, there were no correlations found between the perception and production thresholds of BIT and BFIT [2]. Overall, H-BAT is proposed as a quick and efficient test battery to assess beat perception and production abilities.

B. BAASTA

The BAASTA is a task designed to assess rhythm perception and auditory-motor coupling while still maintaining sensitivity to individual differences and impairments [5]. This is achieved through the development of a series of perceptual and sensorimotor synchronization tasks [5]. This study uses the duration discrimination task from BAASTA. In this task, participants hear two 1 KHz pure tones, one with a standard duration of 600 ms, and the other with a duration ranging from 600 ms to 1000 ms [5]. They are then asked whether the second tone lasted longer than the first. An algorithm termed the maximum likelihood procedure (MLP) is used to adjust the length of the second tone based on real time participants' responses, as defined in Section 4 [5]. An estimate of the perceptual threshold, expressed as a percentage of interonset interval (IOI), is obtained from this task [5].

C. Gold-MSI

Gold-MSI is a test battery designed to measure musical sophistication in a comprehensive way and capture individual differences in the general population [6]. Many musical skills are acquired through repeated and focused musical engagement [6]. As a result, analysis of individual beat perception skills must consider processes of implicit learning that occur during enculturation with Western music [6]. The authors argue that existing test batteries use artificially created experimental stimuli with very little ecological validity and resemblance to real music [6]. Thus, they do not consider music experience and listening expertise of participants [6]. The Gold-MSI test battery is specifically designed to assess musical skills closely related to real-world skilled listening behaviours [6]. In addition, a wide range of musical styles are incorporated when creating the stimulus items to obtain a fair evaluation of skilled engagement in different Western music styles [6]. Two subtests from Gold-MSI are used in this study: the melody memory task and the beat alignment (BAT) perception task [7]. In the melody memory task, novel melodies are created by shuffling distributions of pitch intervals and rhythmic tone durations of widely known folk or popular melodies [7]. In each trial, participants are presented two melodies, an original and a transposed variant. They must then identify whether the two patterns are the same or different despite using different pitches [7]. The BAT perception task uses twelve musical excerpts with a superimposed beep-track. In some of the excerpts, the beep-track coincides with the beats while in others, the superimposed beep-track is consistently ahead or behind the beat. Participants are asked to determine whether the beep-track is on or off the beat. Two performance measures are calculated, an accuracy score and a mean confidence score indicating the confidence levels of participants when answering the presented questions [7].

D. Speeding Up Slowing Down

In their study on neural bases of individual differences in beat perception, Grahn and McAuley (2009) introduce a 'speeding up or slowing down' test to demonstrate differences in sensitivity to an implied beat [3]. For this test, two types of sequences are used (see Appendix A). The first is a five-tone test sequence, consisting of three initial tones marking two 300-ms intervals, followed by two tones specifying a variable final interval [3]. In this sequence, the initial tones imply a periodic 600-ms beat. The next sequence is a four-tone control sequence, consisting of two tones that explicitly specify a 600-ms interval followed by two tones marking the same variable final interval [3]. Participants are presented both sequences and must judge whether they are speeding up or slowing down at the end [3]. The study showed that when presented control sequences, participants generally compared the duration of the final interval to the initial 600-ms interval [3]. However, when presented with sequences containing the implied 600-ms beat, some individuals had the tendency to compare the final interval to the 300-ms referent marked by the first three tones [3]. As a result, these individuals perceived that all sequences were slowing down (See Appendix A). In contrast, individuals who were able to pick up on the implied beat responded that sequences with the same final intervals were speeding up [3]. The aim of this study is to measure individual differences in sensitivity to implied beat. An estimate of individual beat perception strength, w, ranging from 0 to 1 was obtained [3]. Larger values of w indicate greater reliance on the implied 600-ms beat (see Section 4) [3].

See Appendix B for a table showing the summary of beat perception tasks used in this study.

3. RESEARCH OBJECTIVES

The following are research objectives for this study:

O1: *Design a behavioural study to obtain performance measures on the ten beat perception tasks presented in Appendix B*

O2: *Summarize and present preliminary results for tasks from each beat perception test battery*

Various statistical analysis techniques such as Tukey's Mean Difference plots and Pearson correlation coefficients are used to understand how these different tasks compare to each other.

O3: *Implement and use Principal Component Analysis to emphasize variations and bring out patterns in the dataset from ten beat perception tasks*

O4: *Implement and use t-distributed stochastic neighbor embedding (t-SNE), a machine learning algorithm to reduce the dimensions and better visualize the dataset collected using the beat perception tasks*

O3 and O4 will be used to determine the generalizability of measures obtained from the ten beat perception tasks. Ideally, these tasks should successfully capture variability present in beat perception abilities across the general population.

4. TERMS AND CONCEPTS

In this section, key concepts and terms used in this paper are defined:

A. Decibel (dB): A unit to measure the intensity of a sound

B. Inter-Stimulus Interval (ISI): The time interval between the termination of one stimulus/tone and the beginning of another

C. Interonset Interval (IOI): The time interval between the perceived beginning of one note and that of the next note. It is generally considered to be the strongest contributor to accent and pulse perception

D. Inter-Beat Interval (IBI): Time between perceived beats in the stimulus, regardless of where the physical tones are present

E. Maximum Likelihood Procedure (MLP): MLP is an adaptive approach to estimate perceptual thresholds. This procedure is used in the duration discrimination task from BAASTA to set the duration of the second tone in real time when performing each trial. [4]. The MLP algorithm is used because it allows quick and reliable estimation of psychophysical thresholds [4].

F. Principal Component Analysis (PCA): PCA is a dimensionality-reduction tool that transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. PCA is useful for visually assessing similarities and differences between datasets.

*G. Stair-Case Paradigm***:** The stair-case paradigm is an adaptive two-alternative forced-choice discrimination paradigm used to measure perception and production thresholds for the BST, BIT and BFIT tasks in H-BAT [2]. The parameter for relative-intensity difference starts at 20 dB in the BST [2]. In BIT and BFIT, the *d* parameter used in Equations (1) and (2) starts at 20 ms [2]. The parameter is halved when the pattern of the stimulus matches the participant's responses twice consecutively and doubled otherwise [2]. Every time the direction of the parameter changes, a turnaround point is obtained [2]. Each task is run until six turnaround points are collected. The threshold value is calculated as the average across six turnaround points [2].

*H. t-Distributed Stochastic Neighbour Embedding (t-SNE)***:** t-Distributed Stochastic Neighbour Embedding (t-SNE) is a machine learning algorithm for dimensionality reduction developed by

Geoffrey Hinton and Laurens van der Maaten. It is a technique that visualizes high-dimensional data by giving each data point a location in a two or three-dimensional map. t-SNE is designed to minimize divergence between two distributions: a distribution that measures pairwise similarities of the input objects and a distribution that measures pairwise similarities of the corresponding low-dimensional points in the embedding.

I. w -value from Speeding-Up-Slowing-Down Task: w is obtained by fitting a model to data collected from five-tone test sequence trials [3]. The model assumes that a participant used a weighted combination of the 300-ms and 600-ms intervals as reference when judging the final interval of the sequence. Therefore, the *w* value provides the proportion of the two references in predicting pattern of responses [3]. If *w* is closer to zero, the participant only used a 300-ms reference, and did not pick up on the implied beat [3]. On the other hand, if *w* is closer to one, then the participant used the 600-ms reference and is a strong beat perceiver according to the model [3].

5. METHODOLOGY

A. Participants

Twenty participants (10 males, 10 females) with no history of hearing impairments performed the study. The mean age was 19 ± 1.12 years (range: 18-21). There were three participants who were left-handed. Participants possessed a wide range of musical experience: some had no experience playing musical instruments while others had up to 13 years of formal musical training. Most participants with musical experience played the piano or drums. Other types of training included voice training. The mean duration of musical training was 5.50 ± 5.52 years (range: 0 – 16).

B. Apparatus

All tests used in this study were implemented as scripts on a laptop (Windows 7) using Matlab and Psychophysics Toolbox Version 3, which interfaces between Matlab and computer hardware. Matlab functions were used to accurately play controlled auditory stimuli and collect participants' responses. In addition, a Graphical User Interface (GUI) was implemented to act as a common interface for all tests. An electric drum pad was used to record tapping responses for production tasks. Noise cancelling headphones were used by participants to listen to the auditory stimuli. The Python programming language and Python libraries such as NumPy, Matplotlib and scikit-learn were used for data analysis.

C. Tasks

The study consisted of ten beat perception tasks. In addition, participants completed hearing and memory tests. Before beginning each task, detailed instructions and practice trials were provided. Participants performed the tasks in randomized order. For production tasks, participants tapped on the drum pad with the index finger of their dominant hands. A description of each task is provided below:

a) Hearing Test

In each trial, participants were shown either a downward or upward-pointing arrow. If a downward-pointing arrow was shown, participants heard a loud pure tone that gradually became softer. They pressed any key on the keyboard when the pure tone was no longer audible. When shown an upward-facing arrow, a soft pure tone gradually became louder. Participants pressed any key as soon as they started hearing the pure tone. Once this test was completed, a hearing threshold value was obtained, expressed in decibels (dB). This threshold was used to ensure that the rest of the study's auditory stimuli were presented at the same perceived loudness to the participant.

b) BST Perception (H-BAT)

Participants heard tone sequences that consisted of one pure tone to cue participants to the onset of the task and 21 woodblock tones that contained either a duple or triple meter (the first tone of either a group of 2 or 3 was accented with increased intensity). The relative intensity difference between accented and unaccented tones started at 20 dB and was manipulated using the stair-case paradigm (see Section 4). In each trial, participants had to identify the meter by pressing either 'F' (duple) or 'T' (triple).

c) BST Production (H-BAT)

This task was similar to the BST perception task, with the exception that participants had to tap in synchrony with each sequence before identifying whether it contained a duple or triple meter. Participants were asked to modulate tap amplitudes to accommodate the meter of each sequence.

d) BIT Perception (H-BAT)

Participants were presented tone sequences that consisted of one pure tone to cue participants to the onset of the task and 21 woodblock tones. The tempo of these tone sequences either slowed down or sped up gradually. Equation (1) was used to change the interstimulus interval (ISI) between the woodblock tones in each sequence [2]

$$\left.\begin{aligned} ISI_{i+1} &= ISI_i + d \text{ (slower)} \\ ISI_{i+1} &= ISI_i - d \text{ (faster)} \end{aligned}\right\} \quad (1)$$

where the first ISI is 500 (ms) and *d* is a constant [2]. The value of *d* started at 20 ms and was manipulated according to the stair-case paradigm (see Section 4) [2]. Participants decided which temporal-change pattern they heard by pressing 'F' (slower) or 'T' (faster) keys on the keyboard.

e) BIT Production (H-BAT)

BIT production was similar to BIT perception. However, participants tapped in synchrony to the woodblock tones by adjusting their tapping rate to each stimulus. After tapping, participants identified the temporal change pattern (slower or faster)

f) BFIT Perception (H-BAT)

Tone sequences consisted of one pure tone to cue participants to the onset of the task and 27 woodblock tones. Each sequence was created by repeating one quarter note, two eighth notes, one dotted-quarter note, and one eighth note. The inter-beat interval (IBI) of the sequences was slowed down or sped up using Equation (2):

$$\left.\begin{aligned} IBI_{i+1} &= IBI_i + d \text{ (slower)} \\ IBI_{i+1} &= IBI_i - d \text{ (faster)} \end{aligned}\right\} \quad (2)$$

where the parameter *d* started from 20 ms and was manipulated using the stair-case paradigm (see Section 4) [2]. After tapping, participants indicated whether the rhythmic pattern was speeding up or slowing down e by pressing ‘F’ (slower) or ‘T’ (faster) on the keyboard.

g) BFIT Production (H-BAT)

BFIT production was similar to BFIT perception, with the exception that participants produced the beat of the woodblock tones by tapping on the drum pad before identifying the temporal change pattern (slower or faster).

h) Duration Discrimination (BAASTA)

In each trial, participants heard two 1 KHz pure tones. The duration of the first tone was 600 ms. The second tone was either 600 ms or longer (up to 1000 ms). Participants judged whether the second tone lasted longer than the first. The duration of the second tone was set in real time using the MLP algorithm (refer to Concepts and Terms). Participants pressed the ‘S’ key on the keyboard if the second tone was the same as the first, and the ‘L’ key if it was different.

i) Melody Memory (Gold-MSI)

The stimuli were taken from the Gold-MSI test battery [7]. This task consisted of melodies and their variants, created by shuffling distributions of pitch intervals and rhythmic tone durations. In each trial, participants heard the original melody and its variant. The variant was either an identical or a different version. Participants pressed ‘F’ if the two melodies were the same and ‘T’ if they were different. They also gave their confidence in judgement on a three-point scale.

j) BAT Perception (Gold-MSI)

The stimuli were taken from the Gold-MSI test battery [7]. In this test, twelve instrumental excerpts were chosen from three distinct genres (rock, jazz, pop orchestral). In each trial, participants heard a single audio clip from the stimulus set. Half of the trials consisted of clips with superimposed beep-tracks that were on-beat while the other half consisted of superimposed beep-track were off-beat due to tempo or phase alteration. Participants pressed ‘F’ if the beep-track was on-beat and ‘T’ if the beep-track was off the beat. They also gave their confidence in judgement on a three-point scale.

k) Speeding-Up-Slowing-Down (Grahn & McAuley)

Participants were presented one of two sequences, a five-tone test sequence, consisting three initial tones marking two 300-ms intervals or a four-tone control sequence, consisting of two tones specifying the 600-ms interval, followed by two tones specifying a variable final interval (see Appendix A). In each trial, participants had to determine whether the variable interval at the end sped up or slowed down relative to the initial interval, by pressing ‘F’ (slowed down) or ‘T’ (sped up) key on the keyboard.

l) Forward digit span (Memory Test)
Participants heard recordings of random sequence of numbers and were asked to reproduce the sequence by typing the numbers on a keyboard. Sequences were randomized for each participant. The first sequence consisted of three digits. After each successful trial, one more digit was added to the next trial. After each incorrect trial, a digit was removed on the next trial. The task ended after three errors were made. The digit span was calculated as the longest digit string length that participants accurately reproduced.

6. RESULTS

In this section, results from each beat perception test battery are presented. Statistical analysis methods were used to compare results of the ten tasks. This section also provides details on Principal Component Analysis (PCA) and t-distributed stochastic neighbor embedding (t-SNE) algorithm, which were used to further understand how these tests collectively demonstrated variability in beat perception abilities of individuals. The tapping data from the H-BAT tests was not analyzed for this paper.

A. H-BAT Scores
H-BAT scores from twenty participants are summarized in Table 1, including the mean, standard deviation (SD), minimum and maximum of the perception and production thresholds for each task.

Table 1. Scores in the Harvard Beat Assessment Test (H-BAT) (n = 20)

Task	**BST (dB)**		**BIT (ms)**		**BFIT (ms)**	
	Perception	**Production**	**Perception**	**Production**	**Perception**	**Production**
Mean	2.21	7.64	0.59	0.46	0.46	0.56
SD	1.14	6.00	0.69	0.27	0.34	0.50
Minimum	0.86	0.23	0.06	0.12	0.06	0.08
Maximum	5.63	20.00	3.00	0.94	1.33	2.33

B. BAASTA Scores
The threshold value for each participant, measured in percent of interonset interval (IOI), is calculated using the slope of the psychometric function that best fit the data. The mean threshold obtained in the duration discrimination task is 12.85 ± 7.22 % IOI. The minimum threshold value was 6.6×10^{-4}% IOI and the maximum threshold value was 25.04 % IOI.

C. Gold-MSI Scores
Results from melody memory and BAT perception tasks are presented in Table 2. For both tasks, mean, SD, minimum, and maximum values of correct response percentage and the confidence score were obtained.

Table 2. Scores for the melody memory and BAT perception tasks from Goldsmiths Musical Sophistication Index (Gold-MSI) (n = 20)

Task	**Melody Memory**		**BAT Perception**	
	Accuracy (% of correct responses)	**Confidence Score**	**Accuracy (% of correct responses)**	**Confidence Score**
Mean	0.55	0.77	0.67	0.76
SD	0.15	0.37	0.16	0.20
Minimum	0.23	0.46	0.29	0.38
Maximum	0.85	1.89	1.00	1.21

D. Speeding-Up-Slowing-Down Task Scores

Two measures were obtained for this task. The first measure is a threshold value expressed in percentage of inter-onset interval (IOI). The second measure is the *w* value, which provides a measure from 0 (weak beat perceiver) to 1 (strong beat perceiver) (refer to Section 4). The mean, standard deviation (SD), minimum and maximum for each measure is provided in Table 3**.**

Table 3. Scores from the speeding-up-slowing-down task (n = 20)

Task	**Speeding-Up-Slowing-Down**	
	Threshold (% of IOI)	***w***
Mean	10.602	0.629
SD	9.078	0.395
Minimum	0.000	0.000
Maximum	30.925	1.000

E. Forward Digit Span Task Scores

The mean for digit span across twenty participants was 6.05 ± 1.40 digits. The highest digit span length was 8 digits and the lowest was 4 digits.

F. Preliminary Analysis

a) Tukey Mean-Difference Plots

In this study, Tukey Mean-Difference (TMD) plots were used as the first step to visualize observed differences between beat perception tasks with identical measurements. To create a TMD plot for two beat perception tasks, the difference between data points from both tasks were plotted on the vertical axis against their average (mean) on the horizontal axis. A reference line for mean difference and a line indicating ± 1.96 times SD from mean difference (lines of agreement) were also plotted. TMD plots were developed for each of the BFIT, BIT and BST perception and production task pairs, the melody memory and BAT perception tasks, as well as for duration discrimination and speeding-up-slowing-down tasks. All five plots are shown in Appendix C. The TMD plots suggested no consistent bias of one task over another between duration discrimination and speeding-up-slowing-down tasks as well as the BAT perception and BAT melody memory tasks (See Appendix C1, C2) because most points on these plots were well dispersed and within the lines of agreement. The plots for BFIT and BIT perception and production task pairs were inconclusive because points on the TMD plots were not evenly distributed and there were a few outlines present in the plots (See Appendix C4, C5). However, there was bias between the BST perception and production tasks (See Appendix C3). The plot for BST perception and production tasks showed that perceptual thresholds from the production task were consistently greater than those from the perception task, as most points were below a difference of zero (total agreement) with mean difference of -5.43, suggesting bias of the production task over the perception task.

b) Pearson's Correlation Coefficient

Pearson's correlation coefficients (r) were calculated across 11 behavioural variables to understand the relationship among tasks from the four test batteries as well as the memory test. Only accuracy measures from the Gold-MSI tasks and thresholds (% IOI) from the speeding-up-slowing-down task were used for this calculation. A significance level at $p < 0.05$ was used for correlation analysis. The Benjamini-Hochberg procedure was used to lower the expected proportion of type I errors or False Discovery Rate. The false discovery rate or q value chosen was 0.1. See Appendix D for the correlation matrix across 11 behavioural variables from ten beat perceptions tasks and one memory task. The correlation matrix revealed no significant correlation between performance on the beat perception tasks and participants' short-term memory, which was measured using the forward digit span task. Moreover, there were no significant correlations found between any of the beat perception tasks, even for tasks belonging to the same test battery.

G. Principal Component Analysis

Each participant's beat perception abilities can be described through his or her performance on the ten beat perception tasks. Principal Component Analysis (PCA) is a mathematical technique that was used to reduce the dimensionality of these variables, while retaining the variation in beat perception abilities possessed by individuals (refer to Section 4). The leading two principal components accounted for 94% of variance in participants' beat perception abilities, Principal Component 1 with explained variance of 71% and Principal Component 2 with explained variance of 22%. The data was projected onto a two-dimensional plot, with Component 1 on the horizontal axis and Component 2 on the vertical axis (Fig. 1A). When analysing the PCA plot, clusters are formed if subsets of participants have beat perception abilities that are highly similar. As shown in Fig. 1A, no clear formation of clusters was found, indicating that variability of participants' beat perception abilities was captured by the ten beat perception tasks. The results of the PCA plot is consistent with findings that individual beat perception abilities vary in the general population and this was captured by dataset from the tasks.

H. Analysis using t-Distributed Stochastic Neighbor Embedding

Like PCA, t-Distributed Stochastic Neighbor Embedding (t-SNE) technique is well suited for visualization of high-dimensional datasets, as defined in Section 4. However, t-SNE is a probabilistic technique, not a mathematical one. In this study, t-SNE is used to reduce the dataset consisting of behavioural variables from ten beat perception tasks to two dimensions. Participants are plotted using the t-SNE dimensionality reduction step to understand how representative their results were in capturing variability in beat perception. On the two-dimensional plot, participants with similar beat perception abilities are modeled by nearby points and those with varying abilities are modelled by distant points (Fig. 1B). Like the PCA plot, the t-SNE plot did not reveal any clusters, indicating that participants possessed variable beat perception abilities. Similar results were produced after running the t-SNE algorithm multiple times.

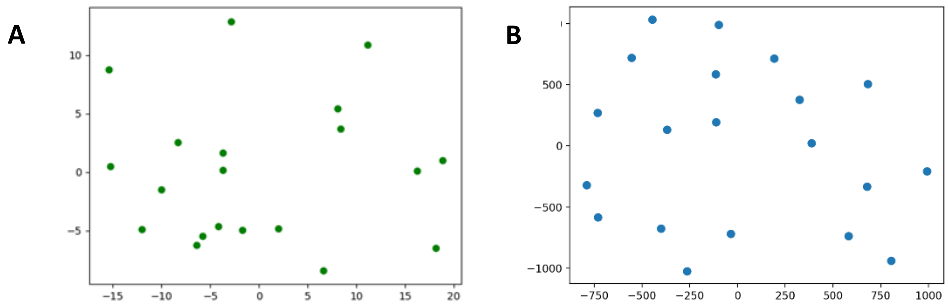

Figure 1. Individual participants are plotted using **(A)** PCA and **(B)** t-SNE to visualize variation in beat perception abilities measured by ten beat perception tasks. For PCA, PC1 is plotted on the x-axis and PC2 is plotted on the y-axis. The t-SNE plot axes are dimensionless

8. DISCUSSION

TMD plots were used as the first step to study agreement between pairs of beat perception tasks with identical measures. There was no bias shown between the duration discrimination (BAASTA) and speeding-up-slowing down tasks as well as the BAT melody memory and BAT perception tasks because points on the plots were well-distributed and within lines of agreement (see Appendix C1 and C2). Agreement between these task pairs implied they can potentially be used interchangeably. The plots for BFIT and BIT perception and production task pairs were inconclusive because points on the TMD plots were not as evenly distributed due to the presence of potential outliers (See Appendix C4 and C5). To analyse the presence of bias between these two task pairs, it is necessary to collect more participant data. However, there was disagreement between the BST perception and production tasks. The production thresholds for BST were consistently greater than the perception thresholds, indicating the presence of bias in one task over the other (see Appendix C3). As a result, the BST perception and production tasks should not be used interchangeably to measure individual ability to discriminate duple and triple meters. To obtain an accurate measure, participants must perform both tasks.

Pearson’s correlation coefficients calculated for the ten beat perception tasks and for the forward digit span task (short-term memory test) showed no correlations, suggesting that participants’ performance on beat perception tasks are unrelated to their short-term memory. Moreover, no significant linear correlations were found across the ten beat perception tasks, which shows that participants’ performance on these tasks were not distributed in the same way. Thus, each of these tasks could be measuring different factors that contribute to variability in individual beat perception skills. The results show that performance of an individual in one task cannot be used to estimate his or her performance in another task, even if both claim to measure beat perception abilities.

Even though the ten beat perception tasks were not correlated, they should still capture variability present in individual beat perception abilities when analyzed in combination. PCA and t-SNE techniques were used to ensure that this was the case by reducing the dimensionality of the dataset. Both PCA and t-SNE plots showed a single population, meaning that the beat perception tasks used in this study successfully captured variability present in the population when measuring beat perception abilities. The absence of clustering in plots indicated that there were no subsets of participants with highly similar beat perception skills. Therefore, these tasks are representative of the variability present in beat perception of the general population.

One potential explanation for the lack of correlation across the ten tasks is related to the different key aspects that govern beat perception [8]. For instance, previous studies have shown that beat perception engages motor areas of the brain, can be transmitted to the brain via different modalities, is hierarchical and is flexible across a wide range of tempi [8]. Most of these tests may be developed to account for different combinations of such factors, resulting in valid but varying measures of beat perception. However, it is not necessarily the case that one task is better than another. Therefore, when used in combination, they are still able to capture the variability existing in the population in terms of beat perception skills. It is also possible that none of these tasks are actually measuring beat perception. All or some of these tasks could be providing measures of individual variability in distinguishing abilities that are unrelated to beat perception, such as the ability to identify intensities and time durations. Researchers are still trying to understand what beat perception is and how it can be measured. Even though it would be ideal to have a gold standard test to measure beat perception skills, this does not exist yet. Therefore, the tasks explored in this study are currently the best available options for behavioural measures of beat perception.

9. CONCLUSION

Ten beat perception tasks from four different test batteries were analyzed in this study using an array of statistical and computational techniques. Data was collected from twenty participants who took part in a behavioural study. The goals of this research project were to understand how these beat perception tasks compared to each other and understand whether performance in these tasks was related to other distinguishing abilities such as short-term memory by comparing performance of participants in beat perception tasks and a memory test. We also aimed to visualize how these tasks captured variability in individual beat perception when used together. The TMD plots revealed that there was disagreement between the BST perception and production tasks. As a result, they should not be used interchangeably. However, TMD plots showed agreement for the melody memory and BAT perception tasks as well as the speeding up slowing down and duration discrimination tasks, indicating the absence of bias between these task pairs. TMD plots for all other task pairs were inconclusive. There were no significant correlations found between any of the ten beat perception tasks and forward digit span task. The lack of correlation indicated that performance on these tasks was not related to short term memory of individuals. There were no significant correlations found across the ten beat perception tasks. This could be the result of each task measuring different factors that contribute to variability in beat perception. Even though all tasks may measure beat perception abilities, each of them might have taken a different approach to do so. Lastly, PCA and t-SNE plots for individual participants revealed variability in beat perception abilities measured by the ten tasks. Therefore, measurements of these tasks are reflective of variability in the general population with regard to beat perception abilities.

The results highlight complexity associated with understanding and measuring how humans perceive beat. There are many factors that must be taken into consideration when attempting to measure beat perception of individuals. However, test batteries may only consider a subset of factors associated with beat perception in their measurements. As a result, even measures from subtests of the same test battery may not be correlated. Even though beat perception tasks from this study were not correlated, they might still be valid measures of beat perception. In contrast, it could also be the case that these tasks are measuring unrelated abilities of individuals such as the ability to identify intensities or time durations in rhythms, which are unrelated to their abilities to measure beat. Even though variability across population would still be demonstrated, these measures are unrelated to how individuals perceive beat. Identifying the underpinnings of beat perception and accurately measuring how we perceive beat is still an ongoing topic of research. Therefore, it is important to understand that these tests may or may not accurately measure beat perception, but they are currently the best available tools for behavioural measures.

10. FUTURE WORK AND LESSONS LEARNED

In this section, future work that will be conducted as an extension of this study and important lessons learned from this research project are discussed.

A. Analysis of Tapping Data

During the behavioural study, participants were asked to tap on the drum pad for tasks that involved beat production. Tapping data was collected for these tasks but was not analyzed in this study. Future work will involve analyzing the tapping data that was collected. Valuable insight can be obtained by comparing results of the tapping data with the perceptual data because studies have shown that motor actions such as tapping are common spontaneous manifestations of beat perceptions [3]. This suggests that strong beat perceivers will show greater supplementary motor activity when listening to auditory sequences [3]. As a result, performing comparisons between perceptual and tapping data can provide information about participants' performance changes when they are able to use motor actions.

B. Analysis of Non-Linear Relationships

In this study, Pearson correlation coefficients were calculated to analyse linear relationships that may exist between beat perception tasks. However, it is also important to determine the presence of any non-linear relationships that might exist in the dataset. Analysing non-linear relationships can potentially reveal more complex patterns among the different beat perception tasks.

C. Comparison with EEG Measurements of Beat Perception Abilities

A recent study used electroencephalogram (EEG) recordings to explore spontaneous building of the beat and meter that was hypothesised to emerge from selective entrainment of neuronal populations at the beat and meter frequencies [9]. It was revealed that when EEG was recorded as participants listened to various rhythms, multiple steady state-evoked potentials (SS-EPs) were elicited [9]. These SS-EPs were observed in the EEG spectrum at frequencies corresponding to the rhythmic pattern envelope [9]. The results of the experiment showed that neural responses to beat and meter reflected the spontaneous emergence of an internal representation of beat, through a mechanism of selective neuronal entrainment within a resonance frequency range [9]. Further work associated with this study would be to compare existing EEG measures of beat perception with behavioural measures of beat perception tasks from this research project to understand how they compare.

D. Lessons Learned

The complexity and limitations associated with accurately understanding and measuring beat perception abilities of individuals as well as the difficulty associated with classifying beat perception is reflected in this project. This study is only the first step in understanding how different tests contribute to measuring beat perception and opens doors to a wide range of potential research topics. There are still more questions to be answered and more progress to be made in this area of research, but this study was an essential first step.

This project also highlights the importance of interdisciplinary research, as solving the problem explored in this study required the expertise of members from multiple disciplines. While working in this study, input from members from the Computer Science department and Brain and Mind Institute at Western University were equally important in understanding the basis of beat perception and accurately analysing the behavioural data. Integration of information, techniques, perspectives, and concepts from the two fields was essential to conduct this study successfully.

APPENDICES

APPENDIX A. Illustration of the Slowing-Down-Speeding-Up Task

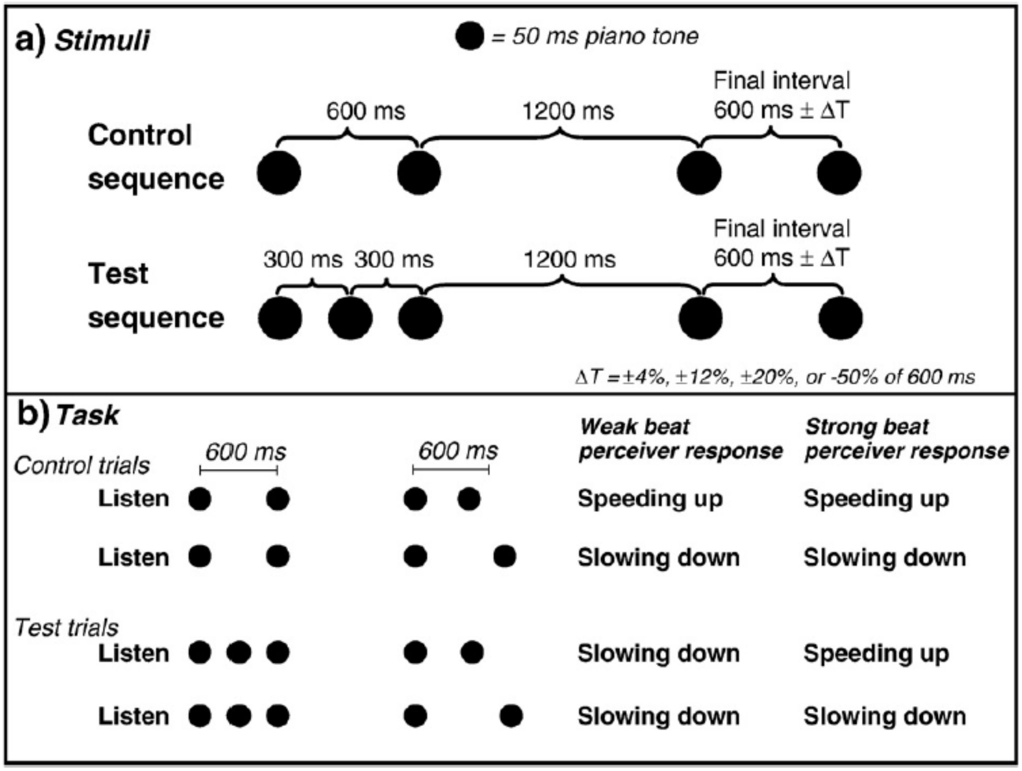

Fig. A1. Ambiguous tempo judgement paradigm: (a) Illustration of control and test sequences. (b) Task and general pattern of responses to control and test sequences by individuals who hear the implied beat (strong beat-perceivers) and by individuals who tend not to bear the implied beat (weak beat-perceivers).

APPENDIX B. A summary of each beat perception task used in this study and dependent measures of each.

Table B1. A summary of each beat perception tasks used in this study, the test battery that it belongs to and its measure

Beat Perception Task	**Test Battery**	**Measure**
Beat Saliency Test (BST) Perception	H-BAT	Perceptual threshold (dB)
Beat Interval Test (BIT) Perception	H-BAT	Perceptual threshold (ms)
Beat Finding and Interval Test (BFIT) Perception	H-BAT	Perceptual threshold (ms)
Beat Saliency Test (BST) Production	H-BAT	Production threshold (dB)
Beat Interval Test (BIT) Production	H-BAT	Production threshold (ms)
Beat Finding and Interval Test (BFIT) Production	H-BAT	Production threshold (ms)
Beat Alignment Test (BAT) Melody Memory	Gold-MSI	Mean accuracy score & mean confidence score
Beat Alignment Test (BAT) Perception	Gold-MSI	Mean accuracy score & mean confidence score
Duration Discrimination	BAASTA	Duration discrimination threshold (% of IOIs)
Speeding Up Slowing Down Test	Grahn & McAuley (2009)	Perceptual threshold (% of IOIs) and 'w' score

APPENDIX C. Tukey Mean-Different Plots for Pairs of Beat Perception Tasks with Identical Measures

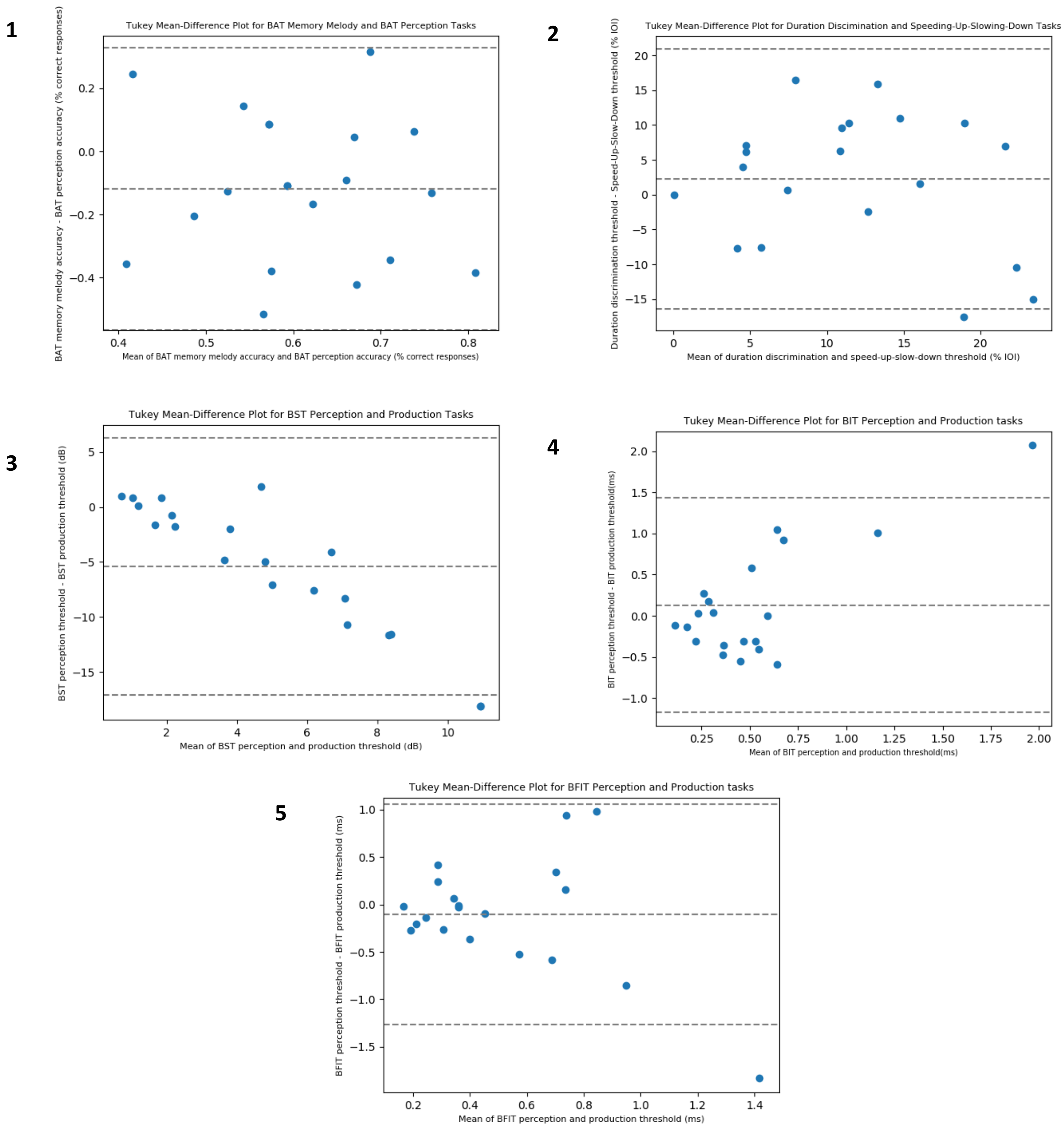

Fig D: Tukey Mean-Difference plots for (1) melody memory and BAT perception tasks, (2) duration discrimination and speeding-up-slowing down task, (3) BST perception and BST production tasks, (4) BIT perception and production tasks (5) BFIT perception and BFIT production tasks.

APPENDIX D. Pearson Correlation Coefficients across ten beat perception tasks and one memory task

Table D1. The correlations calculated across eleven behavioural variables from ten beat perception tasks and one memory test using Pearson's correlation coefficient (*r*). The values in parenthesis indicate the *p*-values.

	Digit Span	**BFIT Per.**	**BIT Per.**	**BST Per.**	**BFIT Prod.**	**BIT Prod.**	**BST Prod.**	**Melody Mem.**	**BAT Perc.**	**Duration Discr.**	**Speed-Up-Slow-Down**
Digit Span											
BFIT Per.	-0.2443 (0.2993)										
BIT Per.	-0.3816 (0.0970)	0.1226 (0.6067)									
BST Per.	0.5825 (0.0070)	0.2314 (0.3263)	0.2338 (0.2312)								
BFIT Prod.	0.1940 (0.4125)	0.0422 (0.8560)	0.4223 (0.0636)	0.1416 (0.5516)							
BIT Prod.	-0.3083 (0.1860)	0.2851 (0.2231)	0.2817 (0.2290)	-0.1510 (0.5250)	-0.1343 (0.5724)						
BST Prod.	-0.0006 (0.9979)	0.1470 (0.5362)	0.1580 (0.5058)	0.1256 (0.5977)	-0.0785 (0.7421)	0.1993 (0.4000)					
Melody Mem.	0.2793 (0.2330)	0.2751 (0.2436)	-0.2075 (0.3801)	-0.1133 (0.6343)	-0.1965 (0.4063)	-0.0885 (0.7106)	0.1040 (0.6626)				
BAT Perc.	0.3350 (0.1489)	-0.3828 (0.0958)	-0.4679 (0.0375)	0.1395 (0.5575)	-0.0980 (0.6811)	-0.2938 (0.2087)	-0.1467 (0.5372)	-0.0962 (0.6868)			
Duration Discr.	0.1740 (1.4633)	-0.3175 (0.1726)	-0.5369 (0.0147)	-0.3144 (0.1670)	-0.0885 (0.7107)	-0.0863 (0.7176)	-0.1777 (0.4535)	-0.2407 (0.3067)	0.1192 (0.6166)		
Speed-Up-Slow-Down	0.1981 (0.4025)	-0.4951 (0.0264)	-0.2497 (0.2883)	-0.3325 (0.1520)	-0.1654 (0.4859)	-0.5316 (0.0158)	-0.1683 (0.4781)	-0.0006 (0.9979)	-0.0813 (0.7332)	-0.0813 (0.7332)	

ACKNOWLEDGEMENT

I would first like to thank my project advisors, Dr. Mark Daley and Dr. Jessica Grahn of the Department of Computer Science and Department of Psychology (the Brain and Mind Institute) at Western University. I am deeply grateful to both professors for giving me the opportunity to join their groups. They consistently allowed this paper to be my own work but steered me in the right direction whenever I needed guidance.

I would also like to offer my thanks to Molly Henry and Avital Sternin for their guidance and support. I could not have completed this project without their help. Lastly, I would like to offer my gratitude to the Guryjot Johal and Shakir Ahamed from Grahn Lab for helping out with behavioural study.

REFERENCES

[1] M. Brett and J. A. Grahn, "Rhythm and beat perception in motor areas of the brain", *Journal of Cognitive Neuroscience*, vol. 19, no.5, pp. 893-906, 2007.

[2] S. Fujii and G. Schlaug. "The Harvard Beat Assessment Test (H-BAT): a battery for assessing beat perception and production and their dissociation", *Frontiers in Human Neuroscience*, vol. 7, 2013.

[3] J. Grahn and J. McAuley, "Neural bases of individual differences in beat perception", *NeuroImage*, vol.47, no. 4, pp.1894-1903, 2009.

[4] J. Grahn and D. Schuit, "Individual differences in rhythmic ability: Behavioural and neuroimaging investigations", *Psychomusicology: Music, Mind and Brain*, vol.22, no. 2, pp.105-121, 2012.

[5] S. Dalla Bella, N. Farrugia, C. Benoit, V. Begel, L. Verga, E. Harding and S. Kotz, "BAASTA: Battery for the Assessment of Auditory Sensorimotor and Timing Abilities", *Behavioral research methods*, vol. 49, no. 3, pp. 1128-1145, 2016

[6] D. Mullensiefen, B. Gingras, J. Musil, and L. Stewart, "The Musicality of Non-Musicians: An Index for Assessing Musical Sophistication in the General Population", PLoS ONE, vol. 9, no. 2, p. e8964, 2014.

[7] D. Mullensiefen, B. Gingras, J. Musil, and L. Stewart, "Goldsmiths Musical Sophistication Index (Gold–MSI) v1.0: Technical Report and Documentation Revision 0.3", 2013.

[8] M. Lense and E. Dykens, "Beat Perception and Sociability: Evidence from Williams Syndrome", *Frontiers in Psychology*, vol. 7, 2016.

[9] S. Nozaradan, I. Peretz and A. Mouraux, "Selective Neuronal Entrainment to the Beat and Meter Embedded in a Musical Rhythm", *Journal of Neuroscience*, vol.32, no. 49, pp. 17572-17581, 2012.