RESEARCH

Open Access



Feasibility and reliability of an online version of the beat alignment test in neurotypical adults and people with stroke

Sarah Gregor^{1,2,4}, Avril Mansfield^{3,5,6}, George Mochizuki^{3,7}, Joyce Chen^{4,6,8} and Kara K. Patterson^{3,4,5*}

*Correspondence: Kara.patterson@utoronto.ca

¹ School of Rehabilitation Science, McMaster University, Hamilton Canada ² Centre for Advancing Collaborative Healthcare & Education, University of Toronto, Toronto, Canada ³ KITE-Toronto Rehabilitation Institute, University Health Network, Toronto, Canada ⁴ Rehabilitation Sciences Institute, University of Toronto, Toronto, Canada ⁵ Department of Physical Therapy, University of Toronto, 160-500 University Ave, Toronto, ON M5G1V7, Canada ⁶ Evaluative Clinical Sciences. Hurvitz Brain Sciences Program, Sunnybrook Research Institute. Toronto, Canada ⁷ School of Kinesiology and Health Science, York University, Toronto, Canada ⁸ Faculty of Kinesiology and Physical Education, University of Toronto, Toronto, Canada

Abstract

Background: Rhythm-based rehabilitation interventions are gaining attention and measuring their effects is critical. With more clinical care and research being conducted online, it is important to determine the feasibility of measuring rhythm abilities online. However, some tools used to measure rhythm abilities, in particular the beat alignment test (BAT), have not been validated for online delivery. This study aims to determine the feasibility, reliability, and learning effects for online delivery of the BAT in adults with and without stroke.

Methods: Neurotypical adults and adults with chronic stroke completed the BAT online three times, with testing sessions separated by 2 to 4 days. The BAT includes a perception task (identifying whether tones overlayed on music matched the beat of the music) and a production task (tapping to the beat of music). Feasibility was evaluated with completion rates, technical challenges and resolutions, participant experience via exit guestionnaire, and test duration. Reliability was measured using inter-class correlations and standard error of measurement, and learning effects were determined using a repeated-measures ANOVA.

Results: Thirty-nine neurotypical adults and 23 adults with stroke participated in this study. More a priori feasibility criteria for the online BAT were met with neurotypical adults than people with stroke. Most components of the online BAT were considered reliable based on an ICC = 0.60 cut-off, except for perception in the neurotypical group, and production asynchrony in the stroke group. There was notable variability in performance, but no learning effects in either group.

Conclusions: Online administration of the BAT is more feasible for neurotypical adults than people with stroke. Challenges with online administration for people with stroke may be partly related to the delivery platform. The BAT is a reliable tool with no learning effects and therefore is a promising way to assess for rhythm abilities online with careful consideration of user interface for people with stroke.

Keywords: Stroke, Music therapy, Musical rhythm, Perception, Test-retest reliability, Feasibility studies



© The Author(s) 2025. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http:// creativecommons.org/licenses/by/4.0/

Background

Stroke is a leading cause of adult disability in Canada and worldwide [1]. One promising approach to address this disability is using music as a tool to enhance neurological rehabilitation [2, 3]. For example, rhythmic auditory stimulation (RAS) for gait post-stroke involves walking to a rhythmic cue provided via music or a metronome [4]. Large effect sizes for speed, cadence, and stride length were confirmed by a meta-analysis of RAS interventions post-stroke [5]. RAS capitalizes on the extensive connectivity between the motor and auditory systems, which facilitates entrainment between the rhythmic auditory cues and motor responses while walking [6]. However, despite favorable changes in gait with RAS observed at the group level, there is variation in responsiveness to RAS at the individual level, and rhythm abilities may play a role [7, 8].

Rhythm is the pattern of silence, sound and emphasis found in music. The beat is the steady, regular pulse perceived in music [7]. When we feel the impulse to move to music, we usually tap to the beat rather than the rhythm [7]. This spontaneous synchronized movement is thought to arise from the processing of the musical beat in motor areas of the brain, specifically the supplementary motor area and basal ganglia (even when no movement is produced) [9]. An individual's baseline rhythm abilities, and more specifically beat perception abilities mediate their response to rhythmbased rehabilitation interventions like RAS [8, 10]. For example, young neurotypical adults with weak beat perception walk with shorter steps and slower speed with RAS compared to those with strong beat perception [10]. This has implications for the use of RAS in people with stroke because their ability to perceive a beat is impaired compared to neurotypical older adults [7]. Furthermore, people with stroke who exhibit worse temporal gait asymmetry with RAS, had weaker beat perception compared to those who did not exhibit a worsening gait pattern with RAS [8]. Therefore, measuring beat processing abilities may help determine who is best suited for a particular music- or rhythm-based intervention. By identifying patients with weak beat perception and production abilities, a therapist can rule out RAS as an option and spend valuable [11] rehabilitation time on interventions that will be more effective for that patient.

The beat alignment test (BAT) is a test of musical beat perception and synchronization [12]. The BAT was designed to be a "naturalistic and simple, yet comprehensive" test of beat processing abilities in the general population [12]. Thus the BAT is easily understood by people with limited music background [13] and has been used with younger adults [14], older adults [7], and people with neurological diagnoses including Parkinson's disease [11] and stroke [7]. Thus, the BAT holds promise as a targeted test of beat perception and synchronization to screen for the appropriateness of rhythm- and music-based rehabilitation interventions. There are two main components of the BAT: (1) a perception task, in which participants are asked to determine if tones superimposed on the music are congruent with the beat of the music; and (2) a production task, in which participants are asked to tap to the beat of the music [13]. Other tools that have been used to measure perceptual and sensorimotor timing and rhythm abilities include the Battery for the Assessment of Auditory Sensorimotor and Timing Abilities (BAASTA) [15], the Harvard Beat Assessment Test [16], and the computerized adaptive beat alignment test (CA-BAT) [17]. All these rhythm assessments include perception and production tasks similar to the BAT but, also include more complex tasks or a greater number of variations than the BAT. For example, compared to the single task of tapping to music stimuli in the BAT, the BAASTA has multiple tapping tasks including as tapping without a pacing stimulus, tapping to a metronome, tapping to music, synchronization continuation (start tapping to a pacing stimulus and continue tapping after the stimulus stops), and adaptive tapping (adapt tapping to the changing tempo in the pacing stimulus) [15]. Several factors limit the utility of these other assessments in the rehabilitation setting. For instance, the BAASTA can take 2.5–3 h to complete, the Harvard Beat Assessment requires specialized equipment, and the CA-BAT is adaptive, meaning that the difficulty of the tasks changes as the person progresses through the test making it harder to compare across participants. Therefore, it is likely that the BAT is the rhythm assessment most easily used for stroke rehabilitation research.

Any patient assessment, including tests of rhythm abilities, needs to adapt to a changing healthcare system. For example, there is a growing interest in the use of technology for the remote delivery of stroke rehabilitation (i.e., telerehabilitation) [18, 19]. Telerehabilitation is a feasible, effective, and acceptable alternative to in-person rehabilitation [20]. If rhythm-based interventions may be administered through telerehabilitation, and assessing rhythm abilities is important to understand who may respond best to such interventions, then it is also important to understand the feasibility of administering rhythm assessments online. The BAT is typically delivered in person with a researcher sitting beside the participant to answer any questions. Additionally, if the BAT is used as an outcome measure in research or clinical care, it is important to understand the psychometric properties and the potential practice effects when it is administered multiple times. Furthermore, components of validity and reliability have been assessed for the CA-BAT [17] and BAASTA [21], but not for the BAT.

The aim of this study is to (1) determine the feasibility of delivering the BAT online; (2) evaluate the reliability of an online version of the BAT in neurotypical adults and adults with stroke; and (3) evaluate practice effects of the online version of the BAT in neurotypical adults and adults with stroke.

Results

Participants

Demographic information for all participants, including musical background can be found in Table 1. Thirty-nine neurotypical adults and 23 adults with chronic stroke consented to participate in our study. Overall, a larger proportion of participants in the stroke group were considered older adults (≥ 60 years old): neurotypical 13% (5/39), stroke 40% (9/22). Both groups had primarily women with 56% (22/39) and 83% (19/23) women in the neurotypical and stroke group, respectively. More participants in the neurotypical group reported formal music training (19/39, 49% neurotypical, and 10/23, 44% stroke), and informal music training (16/39, 41% neurotypical, and 6/23, 26% stroke). Informal music training was mostly described as participants teaching themselves a musical instrument. One person with stroke did not report their music experience.

	Neurotypical (n = 35) ^a	Chronic stroke (n=23)	
Age (n, %)			
18–29 years	16 (46)	1 (4)	
30–39 years	7 (20)	1 (4)	
40–49 years	3 (9)	4 (17)	
50–59 years	4 (11)	7 (30)	
60-69 years	3 (9)	7 (30)	
≥70 years	2 (6)	2 (9)	
NR	0 (0)	1 (4)	
Gender (<i>n</i> , %)			
Man	13 (37)	4 (17)	
Woman	22 (63)	19 (83)	
Stroke characteristics			
Time post-stroke, (months, SD)	N/A	110(135)	
Side of paresis (n, %)			
Left		11 (48)	
Right		6 (26)	
Both		2 (9)	
Unknown		2 (9)	
NR		2 (9)	
Type of stroke (n, %)			
Ischemic		13 (57)	
Hemorrhagic		5 (22)	
Unknown		4 (17)	
NR		1 (4)	
Reported music experience ^a			
Formal in school (<i>n</i> , %)	13 (37.1)	10 (45.5)	
Duration of experience (years, SD)	1.3 (2.2)	1.9 (3.2)	
Formal outside of school (n, %)	15 (42.9)	3 (13.6)	
Duration of experience (years, SD)	2.8 (4.0)	1.3 (3.4)	
Informal music experience (n, %)	16 (45.7)	6 (27.3)	
GSI (score/7, SD) ^a			
Active engagement	3.7 (1.3)	3.3 (1.2)	
Perceptual abilities	5.0 (0.8)	4.8 (0.9)	
Musical training	3.4 (1.7)	2.2 (1.2)	
Singing abilities	3.7 (1.0)	3.2 (1.1)	
Emotions	4.9 (1.3)	4.8 (1.2)	
General sophistication	3.8 (1.2)	3.2 (0.9)	

Table 1 Online BAT participant demographics

Values presented are means with standard deviations in parentheses (continuous variables) or counts with percentages in parentheses (categorical variables)

GSI, Goldsmith sophistication index; NR, not reported; N/A, not applicable; SD, standard deviation

^a 35 neurotypical participants completed the demographics and music background, 1 participant with stroke did not answer questions related to their music background

Feasibility

Completion rates

Figure 1 illustrates the completion of the BAT test at each session for neurotypical adults and adults with chronic stroke. Twenty-eight neurotypical adults (28/39, 72%)



Fig. 1 Flowchart of completion of BAT by participants at each session. Participants lost to follow-up had no recorded attempt of the BAT and did not respond to the investigator's communication attempts. "Technical issues" indicates participants that attempted the BAT but no BAT data was obtained. Complete BAT scores indicates participants for whom all 3 scores were obtained, and partial BAT scores indicates participants for whom at least 1 score was obtained for that session

and 14 (14/23, 61%) people with stroke attempted the BAT with some scores collected at all three sessions.

Technical issues and resolutions

Of the participants who consented (n=39 neurotypical, n=23 stroke), 14/39 (39%) neurotypical adults and 21/23 (91%) adults with stroke experienced technical issues and/ or required assistance to complete the BAT. This rate of technical issues did not meet the a priori feasibility criterion of < 25% of participants. Four (80%) of 5 the neurotypical adults > 60 years old and 9 (100%) of the 9 adults with stroke >60 years old required assistance. Of the participants who required assistance at any point during the study

(n = 14 neurotypical and n = 21 stroke), the issues were either fully resolved so that all 3 BAT scores were recorded or partially resolved so that at least one of the 3 BAT scores were recorded for 5/14 (36%) of the neurotypical adults and 12/21 (57%) of the adults with stroke. This did not meet the a priori feasibility criterion of 75% resolution rate for technical difficulties. Technical challenges often resulted in a longer interval between BAT sessions (range 1–6 days); however, these data were included in the main analysis as a prolonged interval does not influence test–retest reliability [22].

Common technical difficulties reported were challenges downloading the BAT experiment, or difficulties returning to the JotForm page to complete the study after finishing the BAT. Common feedback reported about the BAT in the open-ended responses included issues with the volume, as one neurotypical participant reported "I found the beats in the [perception] test so loud that I could not hear the music clearly". Issues with volume were reported by more people in the stroke group (n=3) than in the neurotypical group (n=1). One recommendation for the production component of the BAT was to improve the clarity about how they should tap to the beat "I just realized with this second session of the production exercise that my tapping varies. Sometimes, I am tapping 1-2-3-4-1-2-3-4; sometimes, I am tapping $1-\ldots-3-\ldots-1-\ldots-3-\ldots$ [...] I assume that you are taking this into account". Other suggestions for the BAT were related to using different music clips for the practice trials than the experiment and having the option to re-do a trial (for example if the participant was distracted). Five participants in the neurotypical group and 6 participants in the stroke group reported not being confident that they did the experiment correctly.

Technical issues that were not resolved and the number of participants who experienced them included: (1) issues with the graphics display of the BAT on the computer screen (likely due to an older incompatible version of Windows software)—2 neurotypical adults, 5 people with stroke; (2) attempting the study on an incompatible computer (i.e., Apple)—1 neurotypical adult, 1 person with stroke; and (3) technical issue not reported—1 person with stroke.

One final issue was completing the same component of the BAT multiple times in a single study session (n=2). One neurotypical participant completed the test twice at time point 1 and three times at time point 3. A second neurotypical participant completed the BAT twice at time point 2. This may be attributed to participants being unaware they repeatedly clicked the same link. In these cases, the scores from the first attempt were used in the analysis.

User experience and BAT duration

Table 2 summarizes the ratings for the 7 items on the exit questionnaire for each group. Ratings for all items met the a priori feasibility criterion (median rating = 4/5) except for Item 1 ("I found the rhythm perception test easy to do online) in the stroke group. The rating of BAT duration as "just right" also met the feasibility criterion in both groups.

Reliability

Average BAT scores over time as well as the ICC, SEM, and MDC for each measure of the BAT in both groups are presented in Table 3. In the neurotypical group, the ICC

Table 2 Responses to exit questionnaire

Exit questionnaire item	Feasibility threshold criterion	Neurotypical	Chronic stroke
1. I found the rhythm perception test easy to do online	Median rating≥4/5	4 (1–5)	3 (1–5)
2. I found the rhythm production test easy to do online		4 (2–5)	4 (2–5)
3. I found the rhythm perception test instructions clear		5 (2–5)	4 (2–5)
4. I found the rhythm production test instructions clear		5 (2–5)	5 (4–5)
5. I found this test fun to do		5 (1–5)	4 (1–5)
6. I think this test accurately measured my rhythm perception abilities		4 (1–5)	4 (2–5)
7. I think this test accurately measured my rhythm production abilities		4 (1–5)	4 (2–5)
Rating of BAT duration (<i>n</i> , %) ^a	\geq 75% rated the BAT to be the right		
Too short	length	2/32 (6%)	2/17 (12%)
Too long		3/32 (9%)	1/17 (6%)
Right length		27/32 (84%)	14/17 (82%)

Feasibility of the online BAT. Variables are presented as medians with ranges in parentheses or as frequency counts with percentages in parentheses

^a Neurotypical n = 32, stroke n = 17 completed the exit questionnaire

Group	Time 1 score	Time 2 score	Time 3 score	ICC (3,1) ^b	SEM ^a	MDC
Perception ac	curacy (% correct)					
NT	64.27 (35.29–88.23)	63.18 (23.43–94.12)	66.74 (29.41–94.12)	0.50	6.42	17.79
Stroke	57.56 (35.29–82.35)	55.88 (23.53–94.12)	53.92 (29.41–88.24)	0.80	3.56	9.86
Production as	ynchrony (msec)					
NT	100.02 (82.07–137.01)	96.51 (71.31–132.01)	99.11 (78.41–133.76)	0.73	3.12	8.64
Stroke	99.13 (76.77–121.47)	99.48 (78.69–118.97)	102.50 (81.03–126.46)	0.47	5.16	14.30
Production va	riability (coefficient c	of variation, %)				
NT	10.77 (4.16–41.58)	11.49 (3.95–48.73)	9.47 (3.63–31.11)	0.89	0.78	2.16
Stroke	12.46 (4.89–21.35)	12.58 (4.37–23.76)	14.94 (5.83–46.97)	0.80	1.68	4.64

Table 3 Reliability of the BAT

Values for rhythm perception and production scores for each group at each time point are presented as means with ranges in parentheses, with ICC, SEM, and MDC values also provided

NT, neurotypical; ICC, intra-class correlation; SEM, standard error of measurement; MDC, minimal detectable change;

%, percent; ms, millisecond

^a Average of 3 time points

^b ICC calculated based on time points 1 and 2

for production component of the BAT (both the asynchrony and variability scores) exceeded the 0.60 minimum threshold for clinical tests. In the stroke group, the ICC for the perception component and the variability score for the production component exceeded the minimum threshold for clinical tests.

Learning effects

For the BAT perception task score (perception accuracy) there was no significant within-subject effect for time (F(2,70)=0.73, p=0.49, $\eta_G^2=0.0053$). There was a significant main effect for group (F(1,35)=10.02, p=<0.01, $\eta_G^2=0.1650$) but no significant group*time interaction (F(2,70)=0.73, p=0.49, $\eta_G^2=0.0053$). For the BAT production task asynchrony score there was no significant within-subject effect for time (F(2,78)=0.65, p=0.53, $\eta_G^2=0.0035$). There was also no significant main effect of group (F(1,39)=0.36, p=0.55, $\eta_G^2=0.0184$) nor was there a significant group*time interaction (F(2,78)=1.26, p=0.38, $\eta_G^2=0.007$). For the BAT production task variability score there was no significant within-subject effect of time (F(2,78)=0.10, p=0.90, $\eta_G^2=0.0008$), no significant effect of group (F(1,39)=1.97, p=0.17, $\eta_G^2=0.05$) and no significant group*time interaction (F(2,78)=1.45, p=0.24, $\eta_G^2=0.0078$).

Discussion

This study assessed the feasibility, test–retest reliability, and learning effects of an online version of a commonly used test of rhythm abilities: the beat alignment test (BAT). Our study produced several interesting findings. First, online remote administration of the BAT did not meet all the feasibility criteria for neurotypical adults and was even less feasible for people with chronic stroke. This was mostly attributable to technical issues associated with the specific platform used to deliver the test. Second, there were no learning effects for the BAT in either neurotypical adults or adults with stroke, which facilitates its use as a repeated outcome measure. Third, as a clinical test, only one of the three BAT components, the variability score for the BAT production task, met the minimum ICC criterion (i.e., 0.60) in both neurotypical adults and adults with stroke. Other components of the BAT met the clinical criterion for only one of the two groups: the perception task for the stroke group and the production asynchrony score for the neurotypical group).

There are a few factors that may have contributed to the observed lower feasibility in the stroke group. First, based on proportions of participants in each age group, the stroke group was older. Previous work has noted that older adults (>54 years old) have more challenges navigating telehealth websites than younger adults [23], even though older adults show similar interest and satisfaction in using technology in healthcare settings [24]. The BAT protocol in the present study required participants click on/close multiple windows, which may have caused difficulties for older participants in the stroke group. Moreover, all but one participant over the age of 60 in either group required help from the researchers to complete this study. Therefore, the present results suggest that age contributes to feasibility of testing online.

A second factor that could have led to lower feasibility is stroke-related cognitive impairment. Telemedicine is more successful with people with stroke when it is easier to use and matches an individual's functional abilities (i.e., motor function and cognition) [25]. Cognition was not assessed in our study, as previous work found no relationship between BAT scores and cognition after stroke [7]. However, 40–79% of people with stroke have mild cognitive impairment [26, 27]. Therefore, it is possible that some participants with stroke in the present study had mild cognitive impairment. These

individuals may have experienced difficulty navigating through the BAT online, even though cognition likely had limited direct impact on BAT scores.

A third factor that may have contributed to lower feasibility in the stroke group is "hidden or invisible" disabilities due to impaired visual and/or auditory perception. Vision impairments affect up to 60% of people with stroke [28], which could have impacted the ability of participants in the present study to complete the BAT. For example, visual field loss or visual inattention could have prevented them from properly scanning the computer screen and locating the correct window to click, resulting in incomplete BAT sessions. Furthermore, central auditory processing disorder is the most common type of hearing impairment in people with stroke (40–55%) [29]. Participant comments in the present study related to volume of the stimuli and difficulty distinguishing the tones from the music clip during the perception task may be partly attributed to this invisible disability. Future work should investigate how cognitive function and invisible visual and auditory disabilities contribute to the success of online testing of rhythm abilities after stroke. In addition, applying signal detection theory to the BAT perception score may provide additional insight into beat processing in people with stroke by separating the behavior into sensitivity and bias [30].

Overall, our ICC analysis showed that the reliability of the various scores generated by the BAT ranged from poor to good test-retest reliability in both groups, with some scores meeting the threshold for use as a clinical test. There was notable variability for each measure within both groups (based on the SEM analysis), especially rhythm perception and production asynchrony. The perception task specifically required the participant to pay attention to the auditory stimulus and only respond at the end of each trial. Research shows that younger participants commonly multi-task with online study surveys [31], and therefore may have been multitasking and distracted during our study. This may explain the high variability observed in the rhythm perception task, especially in the neurotypical group, which included younger participants.

Test-retest reliability in the present study was lower than other measures of rhythm abilities assessed in neurotypical populations. For example, a previous test-retest reliability study on the BAASTA (which used different musical stimuli than the BAT), reported perception score ICC=0.94, rhythm production accuracy ICC=0.77, and consistency ICC = 0.97 in 20 neurotypical older adults [21]. It is important to note that the perception score was calculated differently from the present study (number of hits (correct identification of tone off beat) / number of misses (inaccurate identification of a tone off beat)) [21]. This difference in test-retest reliability in neurotypical adults between the BAASTA and the BAT in the present study may be attributed to the BAASTA being a much longer test overall. A series of validation and test-retest studies on the CA-BAT found that increasing the length of the rhythm assessment improves test-retest reliability in young neurotypical adults (sample size ranged from n=71to = 223) [17]. Furthermore, the BAASTA was delivered in person. Rhythm testing in a laboratory setting produces better test-retest reliability because listening conditions are consistent between testing sessions [17]. In the present study, the online delivery of the BAT meant that beyond giving participants instructions, we had little control over the listening conditions. This also may have attributed to the lower reliability compared to the BAASTA.

Poorer reliability with online testing in the present study may be a result of participants not being able to ask clarifying questions in real-time. For example, participants reported not knowing if they should "tap to the whole note, half note, or quarter notes", with some participants reporting switching their approach throughout the testing. While our calculation of production asynchrony and variability takes these different approaches into account, the uncertainty in instructions may have influenced participants' confidence and therefore behaviors during the production task. Future work with online administration of the BAT should consider having research team members call participants while they complete the online test to help clarify instructions.

Recommendations by the Stroke Recovery and Rehabilitation Roundtable state that researchers developing and testing novel interventions should identify characteristics of individuals who respond best to ensure interventions are efficient and effective [32]. With rising interest in music- and rhythm-based interventions, and evidence that rhythm abilities may mediate who responds best to these interventions it is important to be able to accurately measure rhythm abilities. To our knowledge, this is the first study to evaluate the feasibility and test–retest reliability of a rhythm assessment for people with stroke in an online context. The online version of the BAT holds promise as a reliable test to measure rhythm abilities, however, work is needed to make the test more feasible for people with stroke. Future work should consider if completing the test online simultaneously with a live call with a researcher to provide step-by-step instructions and answer questions improves feasibility and increases test–retest reliability.

Limitations

As participants completed this study in their homes, we were unable to control the environment (e.g., distractions, background noise) and technology (e.g., types of headphones) used during the study. For example, while we screened for hearing abilities as part of our eligibility criteria, we were unable to do a headphone test as recommended by Wood and colleagues for web-based auditory experiments [33]. Furthermore, we were unable to evaluate how much individuals paid attention to the task or whether there were any issues with sound quality that could have influenced BAT scores. Finally, demographic data were collected by self-report, and we were unable to confirm the diagnosis, type of stroke or lesion characteristics with medical records or imaging which are characteristics that may have influenced BAT performance.

Conclusions

Repeated online administration of the BAT exhibited no learning effects and test–retest reliability ranged from poor to good for neurotypical adults and adult with stroke.

Previous in-person administration of the BAT with people with stroke has been successful [34]. However, future work is needed to improve online administration of the BAT. The influence of cognitive and perceptual deficits on online BAT performance should be examined, possibly by including an in-person visit in the protocol to collect demographics and assess impairments. Modifications to online delivery (e.g., administered in real-time with a researcher via web conferencing) to improve feasibility and test–retest reliability and assessment of these modifications with a validated

usability measure (e.g., System Usability Scale [35]) are also required before clinical implementation is possible.

Methods and materials

Participants

This study included two groups of adult participants (over 18 years old): neurotypical adults and people with chronic (>6 months) stroke). Inclusion criteria were self-reported (1) access to a computer with Windows software; (2) access to speakers (including those built into the computers); and (3) ability to understand written English. Participants were excluded if they had: (1) more than mild hearing loss (based on self-report); (2) diagnoses of neurological disorders (other than stroke); and/or (3) completed the BAT or another rhythm assessment within the last year. Participants were recruited using online advertising through physiotherapy clinics specializing in neurorehabilitation, multiple social media outlets (i.e., Twitter, Facebook), and through Heart and Stroke Foundation advertisements. All participants provided informed consent prior to participation in this study. This study was approved by the University of Toronto Research Ethics Board.

Sample size calculation

Sample size calculation was derived from the formula of intra-class correlation coefficient (ICC) test as outlined by Bujang and coauthors [36]. When alpha and power are 0.05 and 80%, respectively, a minimum sample size of 15 is sufficient to detect an ICC of 0.60 when there are 2 observations [36]. We selected an ICC value of 0.60 because this has been described as the minimal acceptable ICC value for clinical evaluations [37]. Therefore, the minimal sample size for recruitment to power our ICC calculation is 15 people per group. We increased our recruitment goal by 50% to 23 per group to account for potential technical issues, missing data and participants withdrawal.

Study protocol

All components of the study, including consent, eligibility assessment, and questionnaires were completed using the online platform JotForm (2021 JotForm Inc., San Francisco, USA). Links to the BAT experiments, which were completed using E-prime Go software (Psychology Software Tools, Pittsburgh, USA), were also provided on the JotForm webpage.

Study participants completed the BAT three times, each separated by 2–4 days, to give flexibility for participants who completed the study in their homes. The 2–4 day timeline was chosen as research suggests that when a population is clinically stable, an interval between tests longer than 2 days does not significantly change test–retest reliability [22].

Demographics

Prior to completing the first session of the BAT, participants completed a demographic questionnaire to collect age, gender identity, and previous experience completing the BAT. If applicable, participants also answered close-ended questions about their stroke including type (i.e., what type of stroke did you have? Ischemic, hemorrhagic, unknown) and hemiparesis (i.e., what side of your body was affected? Right, left, both, unknown). Participants also completed a questionnaire about formal and informal music training

and completed the Goldsmith musical sophistication index (GSI) [38, 39]. Musical sophistication is defined as music behaviors, skills, experiences, and achievements of an individual [38]. The GSI includes questions that span 6 sub-categories (active engagement, perceptional abilities, musical training, singing abilities, emotions, and general sophistication), which explore how people engage with music in Western society [38].

BAT perception and production tasks

The BAT has 2 components: the perception task and the production task. Participants watched an instructional video that replaced the verbal instructions usually provided with in-person administration of the BAT. The instruction video was embedded into the JotForm page and included information regarding how to download the experiment from E-Prime Go, run the experiment, and get back to the JotForm page to continue with the study. These instructions were also provided in written text followed by the link to the E-Prime Go download page. Participants were instructed to complete the BAT in a quiet space using headphones.

The online version of the BAT included the same instructions and stimuli typically used with in-person administration. To adapt for delivery online, we provided participants the opportunity to repeat practice trials if they felt they did not understand the experiment. The perception task of the BAT required participants to determine if tones overlaid on music were on or off the beat of the song [13]. For the off-beat trials, the tones were either (a) too fast/slow in tempo relative to the music (tempo error); or (b) out of phase with the actual beat of the music (phase error) [13]. The response options to the prompt "are the tones on the beat of the music?" were yes or no. There were 17 musical excerpts with an average duration of 16 s [13]. Perception was evaluated as accuracy of responses; the percentage of trials that participants accurately determined if the tone was on or off the beat of the music (perception accuracy (%)=number of trials correct/17 *100).

The BAT production task required participants to tap to the beat of a song with the keyboard space bar (stroke participants used their unaffected hand). The same musical excerpts were used for both perception and production tasks [13]. Rhythm production ability was quantified in two ways: (1) asynchrony—the difference in time between stimulus onset and the participant's taps in milliseconds, averaged across all trials; and (2) variability—the variability in inter-tap intervals (ITI) that is the timing between participant taps (coefficient of variation (%)=standard deviation ITI/mean ITI * 100).

Feasibility

The present study was guided by work on feasibility studies by Bowen and colleagues [40] and previous work that evaluated online arts-based therapies in people with stroke [41]. Bowen colleagues outlined the purpose of feasibility studies as: "to identify not only what—if anything—in the research methods or protocols needs modification but also how changes might occur" [40]. Of the 8 areas of focus addressed by feasibility studies, two applied to the current work: implementation and acceptability [40]. Domains within these two areas of focus that were identified as important for administering the BAT remotely were as follows:

Implementation (to what extent can a measure be successfully delivered to intended participants? [40]): BAT completion rate (percentage of consenting participants who completed all 3 test sessions), technical issues/requests for assistance (number and type), and resolution of technical issues (description of solution and resolve rate measured as number of issues/requests resolved/total number issues/requests). A participant was determined to have complete data at a session if all BAT data (i.e., 3 scores) were recorded and partial data if at least 1/3 BAT scores were recorded. These parameters were tracked throughout the study.

Acceptability (to what extent is a measure judged as suitable, satisfying or attractive to recipients? [40]): suitability and satisfaction were measured with an exit questionnaire administered after participants completed the BAT for the first time. The exit questionnaire consisted of 7 items related to suitability and satisfaction with the online BAT (e.g., clarity of instructions, ease of use, enjoyment, etc.). Participants rated their agreement with the 7 items using a 5-point Likert scale (1=strongly disagree to 5=strongly agree). Participants also rated the length of the BAT as too short, the right length, or too long. Finally, participants were requested to provide any other comments about technical difficulties or recommendations to improve this test online.

Statistical analysis

Demographics

Demographic data were presented with descriptive statistics, including mean and standard deviation for continuous data (i.e., time post-stroke, rhythm abilities) and frequency counts and percentages for categorical and ordinal data (i.e., age, group, and gender).

Objective 1: feasibility

Parameters for each of the feasibility domains were analyzed with descriptive statistics for both study groups. The threshold for feasibility was determined for each domain based on what we estimated to be clinically important: (a) completion rate: \geq 75% of participants complete all 3 sessions; (b) technical issues: \leq 25% of participants report technical issues; (c) technical issue resolution: \geq 75% of all issues are resolved; (d) user experience: all items on the exit questionnaire have a median rating of \geq 4/5; and (e) BAT length: \geq 75% of participants rate the BAT as the 'right length'. The proportion of older adults (i.e., \geq 60 years old, as defined by the United Nations [42]) that reported technical challenges during this study was also calculated. A qualitative descriptive analysis was used for any open-ended responses.

Objective 2, 3: reliability and learning effects

Data for neurotypical adults and people with stroke were analyzed separately. We followed the guidelines for the selection and reporting of intra-class correlation coefficients (ICC) as outlined by Koo and Li [43]. To determine the test–retest variability of sessions 1 and 2, a two-way mixed effects ICC with a single rater/measurement (ICC 3,1) analysis was computed with the BAT perception, BAT production asynchrony, and BAT production variability scores as the dependent variable [43]. Therefore 3 ICCs were computed each for the stroke and neurotypical groups. The ICC value was interpreted as

poor (<0.50), moderate (0.5–0.75), good (0.75–0.9), and excellent (>0.90) based on the recommendations of Koo and Li (2016) [43]. An ICC of 0.60 is the minimum required for clinical tests [37], and therefore was set a priori as the threshold for reliability of the online BAT.

To determine the precision and expected variability of each BAT score, the standard error of measurement (SEM) was calculated (SEM=SD* $\sqrt{(1-ICC)}$) [44, 45]. The SEM can be thought of as a measure of absolute reliability (i.e., consistency of the scores of individuals) [46], and was used to determine the minimal detectable change (MDC=SEM*1.96* $\sqrt{2}$) [46] each group for each measure of the BAT.

To determine learning effects when doing the online BAT, a 2×3 (group x time) repeated-measures analysis of variance (ANOVA), was used to compare all time points of BAT scores within both groups. The independent variable was group (i.e., stroke or neurotypical) and the dependent variable was BAT score (i.e., perception accuracy, production asynchrony, production variability). Significant within-subject main effects of time (p<0.05) were interpreted as evidence of learning. The generalized eta squared (η_G^2) was calculated for between and within-subject effects as a measure of effect size that is comparable across different research designs [47, 48].

Abbreviations

ANOVA	Analysis of variance
BAASTA	Battery for the Assessment of Auditory Sensorimotor and Timing Abilities
BAT	Beat alignment test
CA-BAT	Computerized adaptive beat alignment test
ICC	Intra-class correlation coefficient
GSI	Goldsmith musical sophistication index
MDC	Minimal detectable change
NT	Neurotypical
SEM	Standard error of measurement

Author contributions

SG conceived of the study, designed the study protocol, collected and analyzed the data, performed statistical analysis, interpreted the data and drafted the manuscript. AM, GM, and JC developed the study protocol, interpreted the data and revised the manuscript. KKP conceived of the study, designed the study protocol, is the grant holder, performed statistical analysis, interpreted the data and revised the manuscript. All authors approved the final manuscript.

Funding

The study was supported by the Canadian Partnership for Stroke Recovery with a Catalyst Grant. (no grant number available).

Data availability

No datasets were generated or analyzed during the current study.

Declarations

Ethics approval and consent to participate

This study was approved by the University of Toronto Research Ethics Board (Protocol #: 00039485). All participants provided informed consent.

Competing interests

The authors declare no competing interests.

Received: 13 August 2024 Accepted: 10 February 2025 Published online: 24 February 2025

References

1. Kyu HH, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, Abbastabar H, Abd-Allah F, Abdela J, Abdelalim A. Global, regional, and national disability-adjusted life-years (DALYs) for 359 diseases and injuries and healthy life

expectancy (HALE) for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2018;392(10159):1859–922.

- Sihvonen AJ, Särkämö T, Leo V, Tervaniemi M, Altenmüller E, Soinila S. Music-based interventions in neurological rehabilitation. Lancet Neurol. 2017;16(8):648–60.
- Thaut MH, McIntosh GC. Neurologic music therapy in stroke rehabilitation. Curr Phys Med Rehabil Rep. 2014;2:106–13.
- Thaut MH, McIntosh GC, Rice RR. Rhythmic facilitation of gait training in hemiparetic stroke rehabilitation. J Neurol Sci. 1997;151(2):207–12.
- Wang L, Peng JL, Ou-Yang J-b, Gan L, Zeng S, Wang H-Y, Zuo G-C, Qiu L. Effects of rhythmic auditory stimulation on gait and motor function in Parkinson's Disease: a systematic review and meta-analysis of clinical randomized controlled studies. Front Neurol. 2022;13:818559.
- Thaut MH, Abiru M. Rhythmic auditory stimulation in rehabilitation of movement disorders: a review of current research. Music Percept. 2010;27(4):263–9.
- Patterson KK, Wong JS, Knorr S, Grahn JA. Rhythm perception and production abilities and their relationship to gait after stroke. Arch Phys Med Rehabil. 2018;99(5):945–51.
- Crosby LD, Wong JS, Chen JL, Grahn JA, Patterson KK. An initial investigation of the responsiveness of temporal gait asymmetry to rhythmic auditory stimulation and the relationship to rhythm ability following stroke. Front Neurol. 2020;11:1214.
- 9. Grahn JA, Brett M. Rhythm and beat perception in motor areas of the brain. J Cogn Neurosci. 2007;19(5):893–906.
- Leow LA, Parrott T, Grahn JA. Individual differences in beat perception affect gait responses to low-and high-groove music. Front Hum Neurosci. 2014;8:811.
- 11. Cameron DJ, Pickett KA, Earhart GM, Grahn JA. The effect of dopaminergic medication on beat-based auditory timing in Parkinson's disease. Front Neurol. 2016;7:19.
- 12. Inversen JR, Patel AD. The Beat Alignment Test (BAT): Surveying bat processing abilities in the general population. In: 10th International Conference on Music Perception and Cognition; 2008.
- 13. Iversen JR, Patel AD. The beat alignment test (BAT): Surveying beat processing abilities in the general population. In: International conference for music perception and cognition; Sapporo, Japan. 2010.
- 14. Ready EA, McGarry LM, Rinchon C, Holmes JD, Grahn JA. Beat perception ability and instructions to synchronize influence gait when walking to music-based auditory cues. Gait Posture. 2019;68:555–61.
- Dalla Bella S, Farrugia N, Benoit C-E, Begel V, Verga L, Harding E, Kotz SA. BAASTA: Battery for the assessment of auditory sensorimotor and timing abilities. Behav Res Methods. 2017;49:1128–45.
- Fujii S, Schlaug G. The Harvard Beat Assessment Test (H-BAT): a battery for assessing beat perception and production and their dissociation. Front Hum Neurosci. 2013;7:771.
- Harrison PM, Müllensiefen D. Development and validation of the computerised adaptive beat alignment test (CA-BAT). Sci Rep. 2018;8(1):12395.
- Iodice F, Romoli M, Giometto B, Clerico M, Tedeschi G, Bonavita S, Leocani L, Lavorgna L, Digital Technologies, Web and Social Media Study Group of the Italian Society of Neurology. Stroke and digital technology: a wake-up call from COVID-19 pandemic. Neurol Sci. 2021;42:805–9.
- 19. Smith EE, Mountain A, Hill MD, Wein TH, Blacquiere D, Casaubon LK, Linkewich E, Foley N, Gubitz G, Simard A. Canadian stroke best practice guidance during the COVID-19 pandemic. Can J Neurol Sci. 2020;47(4):474–8.
- Laver KE, Adey-Wakeling Z, Crotty M, Lannin NA, George S, Sherrington C. Telerehabilitation services for stroke. Cochrane Database Syst Rev. 2020. https://doi.org/10.1002/14651858.CD010255.pub3.
- 21. Begel V, Verga L, Benoit C-E, Kotz SA, Dalla Bella S. Test-retest reliability of the battery for the assessment of auditory sensorimotor and timing abilities (BAASTA). Ann Phys Rehabil Med. 2018;61(6):395–400.
- 22. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. J Clin Epidemiol. 2003;56(8):730–5.
- Chun YJ, Patterson PE. A usability gap between older adults and younger adults on interface design of an Internetbased telemedicine system. Work. 2012;41(Supplement 1):349–52.
- 24. Greenwald P, Stern ME, Clark S, Sharma R. Older adults and technology: in telehealth, they may not be who you think they are. Int J Emerg Med. 2018;11:1–4.
- Caughlin S, Mehta S, Corriveau H, Eng JJ, Eskes G, Kairy D, Meltzer J, Sakakibara BM, Teasell R. Implementing telerehabilitation after stroke: lessons learned from Canadian trials. Telemed e-Health. 2020;26(6):710–9.
- Sachdev P, Brodaty H, Valenzuela M, Lorentz L, Looi J, Berman K, Ross A, Wen W, Zagami A. Clinical determinants of dementia and mild cognitive impairment following ischaemic stroke: the Sydney Stroke Study. Dement Geriatr Cogn Disord. 2006;21(5–6):275–83.
- Sensenbrenner B, Rouaud O, Graule-Petot A, Guillemin S, Piver A, Giroud M, Béjot Y, Jacquin-Piques A. High prevalence of social cognition disorders and mild cognitive impairment long term after stroke. Alzheimer Dis Assoc Disord. 2020;34(1):72–8.
- 28. Rowe FJ, Hepworth LR, Howard C, Hanna KL, Cheyne CP, Currie J. High incidence and prevalence of visual problems after acute stroke: an epidemiology study with implications for service delivery. PLoS ONE. 2019;14(3): e0213035.
- Koohi N, Vickers DA, Lakshmanan R, Chandrashekar H, Werring DJ, Warren JD, Bamiou D-E. Hearing characteristics of stroke patients: prevalence and characteristics of hearing impairment and auditory processing disorders in stroke patients. J Am Acad Audiol. 2017;28(06):491–505.
- 30. Lynn SK, Barrett LF. "Utilizing" signal detection theory. Psychol Sci. 2014;25(9):1663-73.
- 31. Zwarun L, Hall A. What's going on? Age, distraction, and multitasking during online survey taking. Comput Hum Behav. 2014;41:236–44.
- 32. Bernhardt J, Hayward KS, Dancause N, Lannin NA, Ward NS, Nudo RJ, Farrin A, Churilov L, Boyd LA, Jones TA. A stroke recovery trial development framework: consensus-based core recommendations from the Second Stroke Recovery and Rehabilitation Roundtable. Int J Stroke. 2019;14(8):792–802.
- Woods KJ, Siegel MH, Traer J, McDermott JH. Headphone screening to facilitate web-based auditory experiments. Atten Percept Psychophys. 2017;79:2064–72.

- 34. Patterson KK, Knorr S, Grahn JA. Rhythm perception and production abilities relate to motor impairment and temporal gait variability after stroke. In: International society for posture and gait research world congress; Jun 28-Jul 2; Seville, Spain. 2015.
- 35. Jordan PW, Thomas B, McClelland IL, Weerdmeester B. Usability evaluation in industry. Boca Raton: CRC Press; 1996.
- 36. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. Arch Orofac Sci. 2017;12(1):1.
- Shoukri MM, Asyali M, Donner A. Sample size requirements for the design of reliability study: review and new results. Stat Methods Med Res. 2004;13(4):251–71.
- Mullensiefen D, Gingras B, Musil J, Stewart L. The musicality of non-musicians: an index for assessing musical sophistication in the general population. PLoS ONE. 2014;9(2): e89642.
- Mullensiefen D, Gingras B, Stewart L, Musil J. The Goldsmiths Musical Sophistication Index (Gold-MSI): technical report and documentation v0. 9. London: Goldsmiths. University of London; 2011.
- 40. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, Bakken S, Kaplan CP, Squiers L, Fabrizio C. How we design feasibility studies. Am J Prev Med. 2009;36(5):452–7.
- 41. Belfiore LC; Michaelchuk WW, Brooks D, Patterson KK Feasibility of a technology-assisted, live-streamed home dance program for people with chronic stroke. In International Neurorehabilitation Symposium—RehabWeek2019; June 24-28, 2019; Toronto, Canada. 2019.
- 42. United Nations DoEaSA, population division: World population ageing 2019 highlights (ST/ESA/SER.A/430). In. New York: United Nations; 2019.
- Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63.
- 44. Darter BJ, Rodriguez KM, Wilken JM. Test–retest reliability and minimum detectable change using the k4b2: oxygen consumption, gait efficiency, and heart rate for healthy adults during submaximal walking. Res Q Exerc Sport. 2013;84(2):223–31.
- 45. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40.
- 46. Dontje ML, Dall PM, Skelton DA, Gill JM, Chastin SF, Team SU. Reliability, minimal detectable change and responsiveness to change: Indicators to select the best method to measure sedentary behaviour in older adults in different study designs. PLoS ONE. 2018;13(4): e0195424.
- 47. de Gil PR, Pham T, Rasmussen P, Kellermann A, Romano J, Chen Y-H, Kromrey JD. GEN_ETA2: A SAS[®] macro for computing the generalized eta-squared effect size associated with analysis of variance models. In: Sas Global Forum, San Francisco, Ca. Copyright; 2020.
- Olejnik S, Algina J. Generalized eta and omega squared statistics: measures of effect size for some common research designs. Psychol Methods. 2003;8(4):434.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.