BASIC EVALUATION OF AUDITORY TEMPORAL STABILITY (BEATS): A NOVEL RATIONALE AND IMPLEMENTATION

Zhuohong Cai[1], Robert J. Ellis[1], Zhiyan Duan[1], Hong Lu[2], and Ye Wang[1]

[1] School of Computing
National University of Singapore
{a0109706,ellis,zhiyan,wangye}@comp.nus.edu

[2] School of Computer Science
Fudan University
honglu@fudan.edu.cn

ABSTRACT

The accurate detection of pulse-level temporal stability has important practical applications; for example, the creation of fixed-tempo playlists for recreational exercise (e.g., jogging), rehabilitation therapy (e.g., rhythmic gait training), or disc jockeying (e.g., dance mixes). Although there are numerous software algorithms which return simple point estimate statistics of "overall" tempo, none has operationalized the beat-to-beat *stability* of an inter-beat interval series. We propose such a method here, along with several novel summary statistics. We illustrate this approach using a public data set (the 10,000-item subset of the Million Song Dataset) and outline a series of future steps for this project.

1. INTRODUCTION

Motor synchronization with an auditory beat has been deemed a human cultural universal [20] and a "diagnostic trait of our species" [16]. Even infants show perceptual sensitivity to and motor coordination with musical rhythms [26,28]. A temporally stable beat facilitates rhythmic human movement during leisure activities such as exercise (for recent reviews, see [10,11]). It also serves as the basis for a class of gait rehabilitation therapies known as "Rhythmic Auditory Stimulation" or "Rhythmic Auditory Cueing" for Parkinson's disease (for reviews, see [12,22]), stroke [25], and others [27].

Numerous beat tracking algorithms have been developed which return a time series of detected beats for a given audio input (for reviews, see [5,18]), returning a simple "beats per minute" point estimate of tempo. None of these algorithms, however, has attempted to operationalize the beat-to-beat *stability* of that tempo over time, other than occasional efforts to note whether multiple excerpts taken from the same audio file have the same approximate tempo. Such a coarse estimate of tempo stability does not have the necessary precision for the type of clinical applications cited above applications, which not only need to know if a given audio file is stable, but the precise time indices at which it is stable (so as to preserve that information in the playlist).

To address these issues, we present a novel analysis tool: "Basic Evaluation of Auditory Temporal Stability" (BEATS) for Matlab (version ≥ 7.8). BEATS is *not* a beat tracking algorithm; instead, it uses the output of an existing beat tracking algorithm (i.e., beat and barline onset timestamps) to provide a full set of outcome statistics. Here, we focus on the "Million Song Subset" of 10,000 metadata files selected from the Million Song Dataset [1] (http://labrosa.ee.columbia.edu/millionsong/), with all audio files processed using the proprietary "Analyze" algorithm [9] developed by The Echo Nest (www.echonest.com). Compatibility with this data source has long-term advantages, as the full Echo Nest library contains over 34 million analyzed audio files.

2. METHODS

2.1 Data inputs

BEATS pulls four Echo Nest fields from each metadata file: `beats_start` and `bars_start` (the estimated onsets of successive beats and barlines, respectively); and `tempo` and `time_signature`. Next, the `beats_start` and `bars_start` vectors are transformed into an inter-beat interval (IBeI) series and an inter-bar interval (IBaI) series, respectively, by taking the first-order difference of each vector.

2.2 Initialization Thresholds

BEATS requires the user to specify three Initialization Thresholds:

(1) "Local Stability Threshold", θ_{Local}: a percentage value (default = 5.0%) used to define temporal stability at the level of individual and successive IBeIs (detailed below).

(2) "Run Duration Threshold", θ_{Run}: the minimum duration (default = 10 s) of a set of consecutive IBeIs (i.e., a "Run") that fall below θ_{Local}.

(3) "Gap Duration Threshold", θ_{Gap}: the maximum duration (default = 2 s) between the last timestamp of Run_j and the first timestamp of Run_{j+1}.

2.3 Internal Calculations

The first statistic calculated by BEATS is the central tendency, λ (for "location") of the IBeI series: the single value which best characterizes the predominant IBeI. Obtaining an "optimal" value for λ can more challenging than simply taking the median or mode of a series. Consider the hypothetical IBeI series **S** shown in Figure 1, which exhibits two tempo changes (at the 21st and 41st IBeIs). In Matlab, `median`(**S**) = 0.869 and `mode`(**S**) = 0.477. (`mode` is known to be problematic for both non-discrete and non-quantized data.) Neither statistic effectively captures the central tendency of **S**.

To address this, we define an iterative loop in which the range of **S** is divided into an increasing odd number of bins *k* from 1 to 15. The loop stops at the largest value of *k* in which the most-frequent bin contains just over one-third of the data (or quits when *k* = 15). λ is then defined as the median value within the most-frequent bin. Under this definition, **S** has a λ = 0.993, which better captures its central tendency.

Having derived λ, the longest "Stable Segment" within an IBeI series can be identified. The first step in this process is to quantify local temporal stability in two ways: local *deviations* (from λ) and local *differences* (adjacent IBeIs). Local deviations are quantified by an absolute deviation from λ (ADL), calculated for each element *i* of IBeI series **S**:

$$S_{\mathrm{ADL},\boldsymbol{i}} = 100 \times \frac{|S_{\boldsymbol{i}} - \lambda|}{\lambda}. \qquad (1)$$

Local differences are quantified by a first-order absolute successive difference (ASD), calculated for each element *i* of **S**:

$$S_{\mathrm{ASD},\boldsymbol{i}} = 100 \times \frac{|S_{\boldsymbol{i}} - S_{i-1}|}{0.5 \times (S_i + S_{i-1})}, \qquad (2)$$

where $S_{\mathrm{ASD},1} = 0$ to preserve the series indexing. Both $\mathbf{S}_{\mathrm{ADL}}$ and $\mathbf{S}_{\mathrm{ASD}}$ are expressed relatively (i.e., as percentages) to facilitate comparisons across IBeI sequences in different tempo ranges.

Next, a binarized version of **S** ($\mathbf{S}_{\mathrm{Bin}}$) is created:

$$S_{\mathrm{Bin},i} = \begin{cases} 1, & \begin{cases} S_{\mathrm{ADL},i} \le \theta_{\mathrm{Local}} \\ S_{\mathrm{ASD},i} \le \theta_{\mathrm{Local}} \end{cases} \\ 0, & \text{otherwise} \end{cases}. \qquad (3)$$

$\mathbf{S}_{\mathrm{Bin}}$ identifies the locations of Runs (i.e., strings of 1s) and Gaps (strings of 0s) within the IBeI series itself. Finally, the Stable Segment is defined as the longest sequence of {Run_j, Gap_j, Run_{j+1}, … Gap_{n-1}, Run_n}, where each Run has a duration ≥ θ_{Run}, each intervening Gap has a duration ≤ θ_{Gap}, and the median IBeI value of each pair of neighboring Runs has a percent difference of ≤ θ_{Local}.

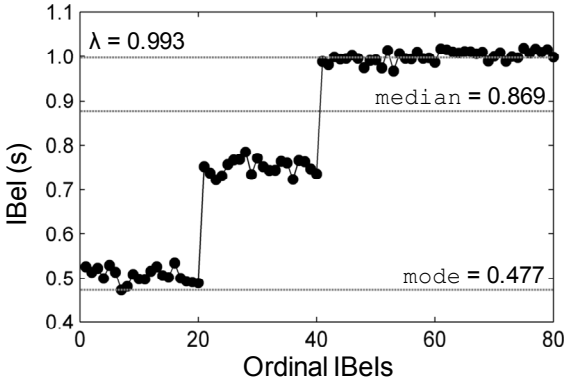

Figure 1. A hypothetical IBeI series, with IBeI values (*y*-axis) plotted ordinally (*x*-axis). The usual estimators of median and mode are both non-optimal. The newly-proposed λ statistic provides a better match.

2.4 Outcome Statistics

BEATS calculates six statistics from each Stable Segment:

(1) "Stable Duration" (in seconds): the time between the first and last time stamps of the longest run.

(2) "Stable Percentage": Stable Duration relative to the duration of the entire IBeI series.

(3) "Estimated Tempo" (in beats per minute, BPM): the median IBeI value within the Stable Segment, multiplied by 60.

(4) "Estimated Meter": a more precise definition than the typical beats-per-bar value. Specifically, for a Stable Segment with a bar timestamp series {r_i, r_{i+1}, ...} and beat timestamp series {b_j, b_{j+1}, ...}, let B_i be the number of beat timestamps for which $r_i \le b_j < r_{i+1}$. Estimated Meter is then taken as the mean of all B_i. Only in the case when all B_i have the same value will an integer value result (e.g., 4.00), providing a simple way to identify the presence of a changing meter within the Stable Segment.

(5) "Percentile of Absolute Deviations from λ" (PADL_P): a statistically robust alternative to a percentage-based coefficient of variation (CV), used extensively in the gait literature (for a review, see [6]). For an IBeI series **S**, CV is defined as the standard deviation of *S* divided by the mean of *S*, and multiplied by 100. Because it makes use of the standard deviation, CV is susceptible to inflation by high-value outliers (e.g., a beat that is "dropped" by the beat tracking algorithm). By contrast, PADL_P offers a more robust formulation:

$$\mathrm{PADM}_P = \mathtt{prc}(\mathbf{S}_{\mathrm{ADL}}, P), \qquad (4)$$

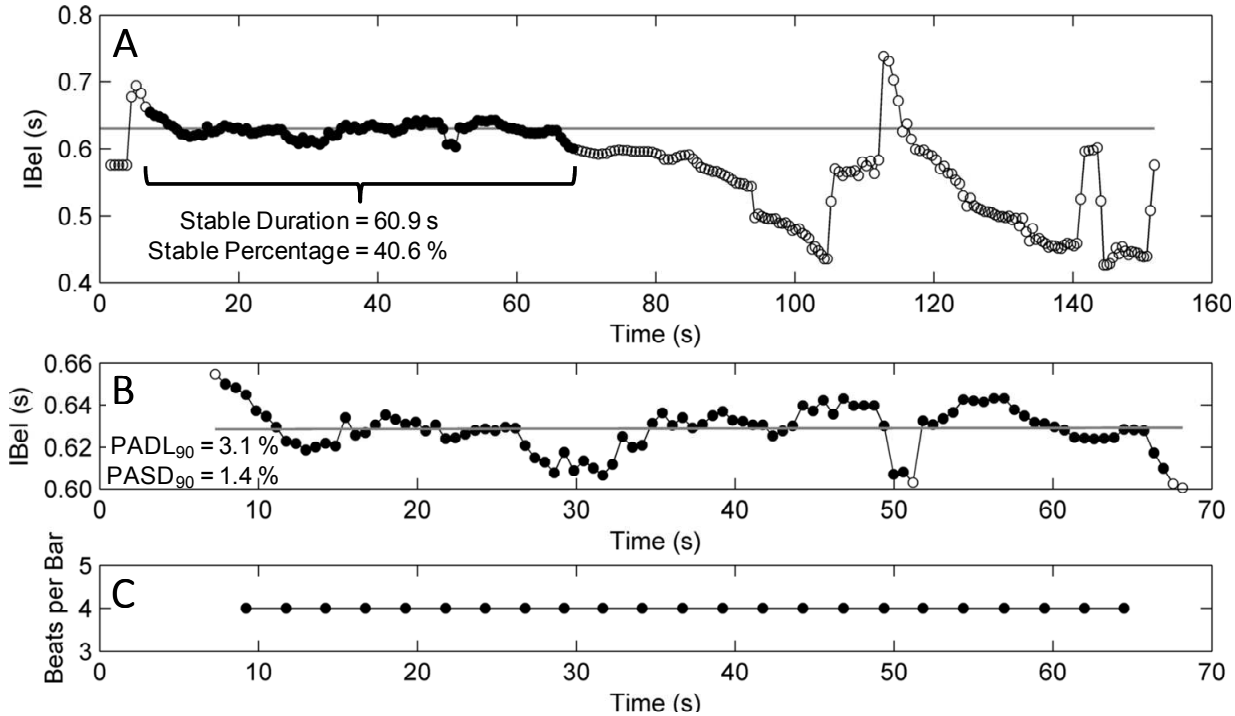

Figure 2. Visual illustration of the Outcome Statistics calculated by BEATS. Panel A shows the IBeI series (*y*-axis) as a function of real time (*x*-axis), with λ shown as a horizontal gray line (at *y* = .631) and the Stable Segment highlighted in filled circles. Panel B shows the Stable Segment in isolation; the best-fitting line (gray line) reveals a lack of temporal drift. Panel C shows all the number of beats per bar for all detected downbeats; Estimated Meter is consistent at four beats per bar throughout the Stable Segment.

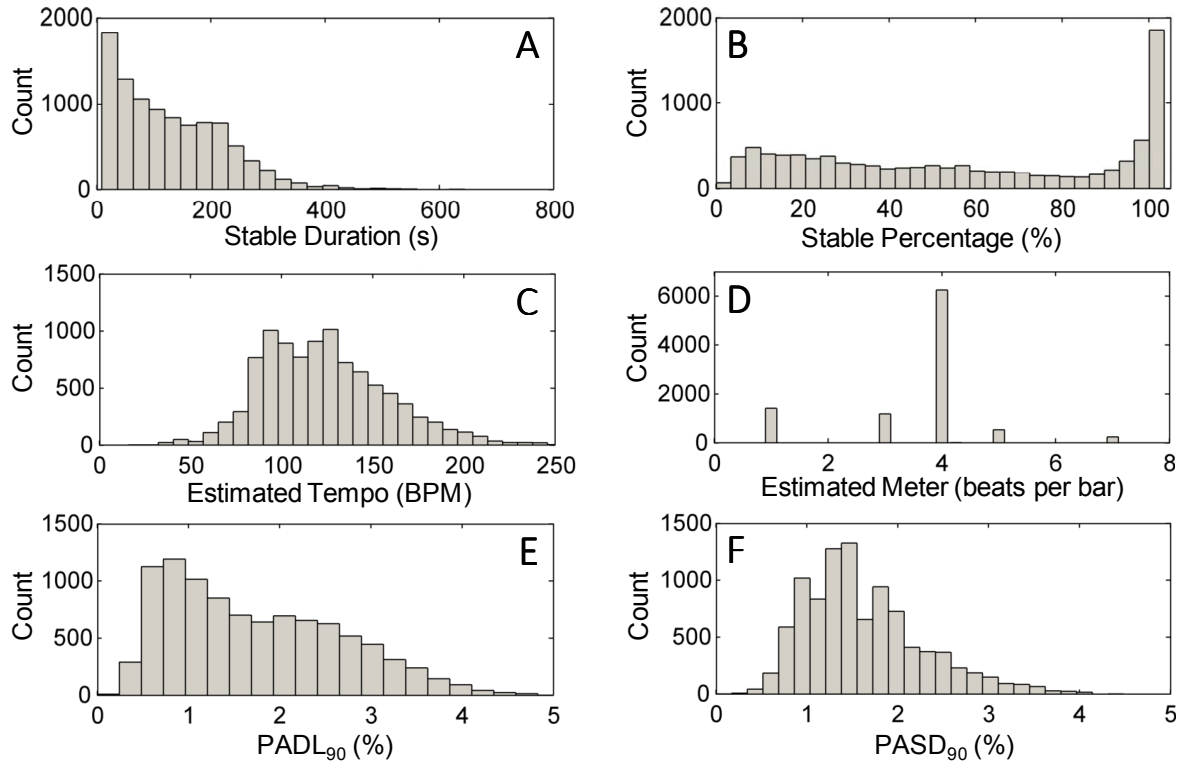

Figure 3. Histogram summaries of the six Outcome Statistics across the full 10,000-item dataset.

where $\texttt{prc}(\mathbf{S}_{ADL},P)$ is the Pth percentile of the $\mathbf{S}_{ADL}$ vector (Eq. 1).

(6) "Percentile of Absolute Successive Differences" ($PASD_P$): a robust alternative to the root-mean-square of successive differences (rMSSD), widely used in studies of heart rate variability to quantify beat-to-beat fluctuations (for a review, see [24]. However, just as squaring deviations from the mean make the standard deviation susceptible to outliers, squaring successive differences yields a potentially inflated RMSSD. By contrast, $PASD_P$ is defined as:

$$PASD_P = \texttt{prc}(|\mathbf{S}_{ASD}|, P)\ . \quad (5)$$

For both $PADL_P$ and $PASD_P$, BEATS uses $P = 90$ as its default. In practice, however, any value of P between 0 and 100 may be used.

2.5 Implementation

BEATS was run on the 10,000-item dataset using its default Initialization Thresholds (Section 2.2). (The rationale behind $\theta_{Local} = 5.0\%$ is explained in Section 4.1.)

3. RESULTS AND DISCUSSION

Figure 2 presents a visual illustration of the six Outcome Statistics calculated by BEATS, in a single file from the Million Song Subset (Track TRAHNHL128F14A4DDD: "In the Hall of the Mountain King" by Edvard Grieg, performed by the Staatskapelle Dresden; available at http://open.spotify.com/track/2cTXwtIFEeCNa0ZtbI97zh). This work is famous for its *accelerando*, which can be seen in the IBeI plot of Figure 2A (albeit with some confusion on the part of the Echo Nest "Analyze" algorithm [9], a point discussed further in Section 4.1). Such a recording would be of limited use for a constant-tempo exercise paradigm. A temporally stable segment, however (using $\theta_{Local} = 5.0\%$), can in fact be found between the 0′08″ and 1′09″, which can be more clearly appreciated in Figure 2B. The identified Stable Segment has a $PADL_{90} = 3.1\%$ and a $PASD_{90} = 1.4\%$, markedly different than if those statistics are calculated from the *entire* IBeI series ($PADL_{90} = 23.3\%$ and a $PASD_{90} = 3.1\%$). Finally, Figure 2C shows the number of beats per bar within the Stable Segment; this yields an Estimated Meter = 4.

Figure 3 presents a histogram for each of the six BEATS Output Statistics across the full 10,000-file data set. Of particular note is Figure 2B, which indicates that Stable Percentage varied widely across the data set. Indeed, only 18.6% of files were deemed temporally stable (i.e., as defined by $\theta_{Local} = 5.0\%$) over their entire duration (i.e., Stable Percent = 100). In other words, the probability that a song randomly selected from the MSD can be played it in its entirety as part of a rhythmic movement paradigm (i.e., has a moderately stable perceptual tempo with less than 5.0% local tempo variability) is < 20%.

Figure 4 presents a slightly different picture, plotting the percentage of files (y-axis) with a Stable Duration ≥ the x-axis value. Allowing BEATS to identify the Stable Segment within each audio file (if present) yields a higher percentage of files available for exercise playlists; for example, 55.7% of files are temporally stable over a duration of ≥ 90 s within the file—three times the number of files that are stable over their entire duration.

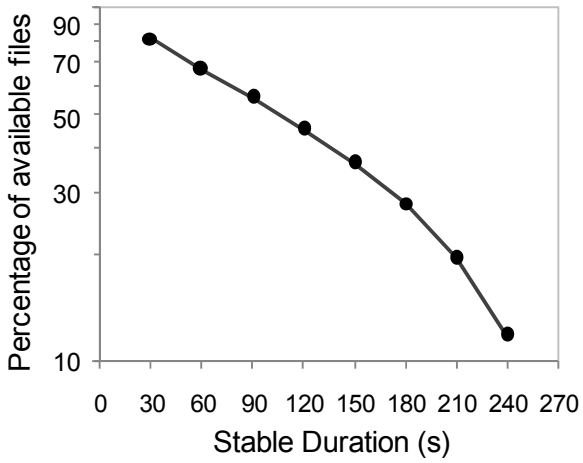

Figure 4. The percentage of files in the 10,000-item dataset which have a Stable Duration ≥ the corresponding x-axis value.

The power of BEATS lies in the flexible way its Outcome Statistics may be combined to deliver a stimulus set optimized to a user's specific needs. For example, a gait training paradigm requiring highly stable music might use the following set of inclusion criteria: Stable Duration ≥ 120 seconds, Tempo between 50 and 150 BPM, Estimated Meter = 4.0, $PADL_{90} \leq 2.5$, and $PASD_{90} \leq 2.5$. This combination of inclusion criteria retains 24.2% of the 10,000-file dataset (again, using $\theta_{Local} = 5.0\%$). Although this percentage may seem low, it is *scalable*. That is, assuming that the remainder of the Million Song Dataset yields similar distributions for the six Outcome Statistics, nearly 250,000 candidate songs could be made available for rhythmic synchronization paradigms (using these same inclusion criteria), and more still if the entire 34-million-item Echo Nest library were leveraged.

4. CONCLUSIONS, CAVEATS, AND FUTURE DIRECTIONS

We present a novel tool to evaluate auditory temporal stability (BEATS). An important departure that BEATS makes from previous methods is that it seeks to identify the most *temporally stable* segment within an inter-beat interval (IBeI) series for an entire audio file, rather than

derive a point estimate of tempo for the entire IBeI series. This increased flexibility enables BEATS to identify a greater number of candidate pieces of music that satisfy the requirements of rhythmic exercise applications.

4.1 Caveats

For ease of illustration in the present report, a single Local Stability Threshold (θ_{Local} = 5.0%) was used to instantiate BEATS and generate the associated figures. This value was chosen based on prior studies which have explored just-noticeable differences (JNDs) for changes in tempo (e.g., [3,8,19,23]), with reported values ranging from 10% (for single pairs of intervals) to 2% (for longer sequences). The stimuli in each of these cited studies, however, were all (1) isochronous (i.e., all intervals equally spaced in time), and (2) contained no more than 10 temporal intervals per sequence. Both factors limit the generalizability of these studies to actual extended excerpts of music, which is frequently (1) non-isochronous and (2) of a longer duration (enabling a stronger reference tempo to be formed, and thus a more finely tuned ability to detect change). θ_{Local} = 5.0% was chosen as a compromise, but warrants further experimental validation. That is, determining a threshold for "perceptually stable" in a non-isochronous IBeI series with varying degrees of local and global variability across trials (and across different tempo ranges) would greatly increase the utility of BEATS.

Another issue, highlighted by Figure 2, concerns the accuracy of the beat tracking algorithm itself. That is, BEATS is ignorant of the fidelity of the algorithm used to derive an inter-beat and inter-bar interval series. In the case of Figure 2, the derived IBeI series (as derived by the Echo Nest "Analyze" algorithm [9]) does not match the steady acceleration of tempo present within the audio file. Furthermore, preliminary exploration of the 10,000-item dataset suggests that highly complex or multi-layered rhythm loops that have an underlying perceptual pulse may nevertheless flummox a beat tracking algorithm.

Although this may mean that BEATS is conservative (in that it will classify some pieces of music as "temporally unstable" when they in fact may not be), such conservativeness may be beneficial in practice, as it will rule out pieces of music that may in fact be too challenging for listeners to synchronize with.

Alternatively, research from another sub-domain of audio content analysis, *score–performance matching* (e.g., [7,21]), may provide techniques to more robustly quantify changes in tempo over time, enhancing the ability of BEATS to detect excerpts of tempo stability.

4.2 Future Directions

By summarizing temporal stability using simple summary statistics, the output of BEATS can become the input to search engines for which tempo is a key feature (e.g., [4,14,15]). In its current state, however, BEATS is a work in progress. Our own future goals for this project include (1) implementing BEATS on much larger datasets (such as the entire Million Song Dataset, or even larger Echo Nest datasets), and (2) developing a high-quality web-based user interface ("iBEATS") that will offer visualizations (box plots, scatter plots) and flexible parameter settings (buttons and sliders) to efficiently sort and sift through large amounts of metadata (including artist, release date, and genre tags) to create customized playlists for clinical (e.g., gait rehabilitation) or commercial (e.g., rhythmic exercise) applications.

5. ACKNOWLEDGMENT

We thank three anonymous reviewers for helpful comments. This research is supported by the Singapore National Research Foundation under its *International Research Centre @ Singapore* funding initiative, and administered by the Interactive Digital Media Programme Office.

6. REFERENCES

[1] T. Bertin-Mahieux, D. P. Ellis, B. Whitman, and P. Lamere, "The million song dataset," *in ISMIR 2011: Proceedings of the 12th International Society for Music Information Retrieval Conference*, October 24-28, 2011, Miami, Florida, 2011, pp. 591–596.

[2] S. Dixon, "An interactive beat tracking and visualisation system," in *Proceedings of the International Computer Music Conference*, Havana, Cuba, 2001, pp. 215–218.

[3] C. Drake and M. C. Botte, "Tempo sensitivity in auditory sequences: evidence for a multiple-look model," *Percept. Psychophys.*, vol. 54, no. 3, pp. 277–286, Sep. 1993.

[4] F. Gouyon, "Dance music classification: A tempo-based approach," in *Proceedings of the International Conference on Music Information Retrieval*, Barcelona, 2004.

[5] F. Gouyon, A. Klapuri, S. Dixon, M. Alonso, G. Tzanetakis, C. Uhle, and P. Cano, "An experimental comparison of audio tempo induction algorithms," *Audio Speech Lang. Process. Ieee Trans.*, vol. 14, no. 5, pp. 1832–1844, 2006.

[6] J. M. Hausdorff, “Gait dynamics in Parkinson’s disease: common and distinct behavior among stride length, gait variability, and fractal-like scaling,” *Chaos Woodbury N*, vol. 19, no. 2, p. 026113, Jun. 2009.

[7] H. Heijink, P. Desain, H. Honing, and L. Windsor, “Make me a match: An evaluation of different approaches to score—performance matching,” *Comput. Music J.*, vol. 24, no. 1, pp. 43–56, 2000.

[8] R. B. Ivry and R. E. Hazeltine, “Perception and production of temporal intervals across a range of durations: Evidence for a common timing mechanism,” *J. Exp. Psychol. Hum. Percept. Perform.*, vol. 21, no. 1, pp. 3–18, 1995.

[9] T. Jehan, “Analyzer Documentation.” 2011. http://developer.echonest.com/docs/v4/_static/AnalyzeDocumentation.pdf

[10] C. I. Karageorghis and D.-L. Priest, “Music in the exercise domain: a review and synthesis (Part I),” *Int. Rev. Sport Exerc. Psychol.*, vol. 5, no. 1, pp. 44–66, Mar. 2012.

[11] C. I. Karageorghis and D.-L. Priest, “Music in the exercise domain: a review and synthesis (Part II),” *Int. Rev. Sport Exerc. Psychol.*, vol. 5, no. 1, pp. 67–84, Mar. 2012.

[12] I. Lim, E. Van Wegen, C. De Goede, M. Deutekom, A. Nieuwboer, A. Willems, D. Jones, L. Rochester, and G. Kwakkel, “Effects of external rhythmical cueing on gait in patients with Parkinson’s disease: a systematic review,” *Clin. Rehabil.*, vol. 19, no. 7, pp. 695–713, 2005.

[13] J. Langner and W. Goebl, “Visualizing expressive performance in tempo-loudness space,” *Comput. Music J.*, vol. 27, no. 4, pp. 69–83, 2003.

[14] Z. Li, Q. Xiang, J. Hockman, J. Yang, Y. Yi, I. Fujinaga, and Y. Wang, “A music search engine for therapeutic gait training,” in *Proceedings of the international conference on Multimedia*, 2010, pp. 627–630.

[15] Z. Li and Y. Wang, “A domain-specific music search engine for gait training,” in *Proceedings of the 20th ACM international conference on Multimedia*, New York, NY, USA, 2012, pp. 1311–1312.

[16] B. H. Merker, G. S. Madison, and P. Eckerdal, “On the role and origin of isochrony in human rhythmic entrainment,” *Cortex*, vol. 45, no. 1, pp. 4–17, 2009.

[17] M. F. McKinney and D. Moelants, “Ambiguity in tempo perception: What draws listeners to different metrical levels?,” *Music Percept.*, vol. 24, no. 2, pp. 155–166, 2006.

[18] M. F. McKinney, D. Moelants, M. E. P. Davies, and A. Klapuri, “Evaluation of audio beat tracking and music tempo extraction algorithms,” *J. New Music Res.*, vol. 36, no. 1, pp. 1–16, 2007.

[19] N. S. Miller and J. D. McAuley, “Tempo sensitivity in isochronous tone sequences: the multiple-look model revisited,” *Percept. Psychophys.*, vol. 67, no. 7, pp. 1150–1160, 2005.

[20] B. Nettl, “An ethnomusicologist contemplates universals in musical sound and musical culture,” in *The origins of music*, B. Wallin, B. Merker, and S. Brown, Eds. Cambridge, MA: MIT Press, 2000, pp. 463–472.

[21] A. Robertson, “Decoding Tempo and Timing Variations in Music Recordings from Beat Annotations.,” in *ISMIR*, 2012, pp. 475–480.

[22] T. C. Rubinstein, N. Giladi, and J. M. Hausdorff, “The power of cueing to circumvent dopamine deficits: a review of physical therapy treatment of gait disturbances in Parkinson’s disease,” *Mov. Disord.*, vol. 17, no. 6, pp. 1148–1160, 2002.

[23] H. H. Schulze, “The perception of temporal deviations in isochronic patterns,” *Percept. Psychophys.*, vol. 45, no. 4, pp. 291–296, 1989.

[24] Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, “Heart Rate Variability: Standards of Measurement, Physiological Interpretation, and Clinical Use,” *Circulation*, vol. 93, no. 5, pp. 1043–1065, Mar. 1996.

[25] M. H. Thaut, G. C. McIntosh, and R. R. Rice, “Rhythmic facilitation of gait training in hemiparetic stroke rehabilitation.,” *J. Neurol. Sci.*, vol. 151, no. 2, pp. 207–212, Oct. 1997.

[26] I. Winkler, G. P. Háden, O. Ladinig, I. Sziller, and H. Honing, “Newborn infants detect the beat in music,” *Proc. Natl. Acad. Sci.*, vol. 106, no. 7, pp. 2468–2471, 2009.

[27] J. E. Wittwer, K. E. Webster, and K. Hill, “Rhythmic auditory cueing to improve walking in patients with neurological conditions other than Parkinson’s disease-what is the evidence?,” *Disabil. Rehabil.*, vol. 35, no. 2, pp. 164–176, 2013.

[28] M. Zentner and T. Eerola, “Rhythmic engagement with music in infancy,” *Proc. Natl. Acad. Sci.* U.S.A., Mar. 2010.