

Keep the Beat

Beach Clark, Nathaniel Condit-Schultz

Introduction

In a BBC interview (2014), Jimmy Page discussed the recording session for Led Zeppelin’s iconic *Stairway to Heaven* (1971):

One of the cardinal rules when I was a studio musician was that you didn’t speed up. And I was keen to do something which had an acceleration to it, not only from the musical point of view but from the lyricist, so that the whole thing would start to gain a momentum as it went through, so it wasn’t just a monotone piece. And by that, I mean that it would subtly speed up, so you’re breaking the number one cardinal rule.

Page’s “cardinal rule” is certainly familiar to most musicians: a basic expectation of musical competence is that one can maintain a steady beat without “dragging” or “rushing” the tempo. The use of technology to facilitate and develop musicians’ ability to maintain stable tempos can be traced back to the 17th century, with the most important invention being the *metronome* introduced in the early 19th century (Fallows 2001). However, the unmusical sounds produced by traditional metronomes preclude their usage during actual music performance or production. In contrast, modern music production and performance technology allows musicians to produce perfectly tempo-stable music, whether by guiding musician’s performances with stable in-ear “clicks,” using inherently tempo-stable instruments (e.g., drum machines, turntables, looping equipment), or through post-performance digital editing.

Page’s discussion of *Stairway to Heaven* illustrates that tempo-stability is not necessarily aesthetically desirable. For as long as metronome technologies have existed they have met “approval and disapproval, with many fine teachers asserting that it is antimusical and promotes only wooden performance” (Fallows 2001). Indeed, the systematic variation of timing from metronomic stability—referred to as *rubato* in the Western tradition—is an essential part of musical performance (Sloboda 1983; Hudson 2001). In many musical traditions however—including modern American popu-

lar music—the use of tempo *rubato* is minimized in a “communicative tradeoff” with rhythmic complexity and syncopation (Temperley 2004), facilitating dance, trance, and synchronization. In these modern traditions (and in earlier Western performance traditions (Hudson 2001)), expressive timing is largely realized in timing displacements between different musical parts—playing “behind the beat” for instance—, even when the overall tempo remains steady (Prögler 1995). Thus, there is room for “expressive timing” even within the world of click tracks. Nonetheless, the notion that a perfectly stable tempo might be unaesthetic persists among many musicians: In another interview, Tim Comerford and Chris Cornell (of the band Audioslave) specifically point out that they “don’t play to clicks” [5], because they want the “organic feel.” Further, composers and performers are aware that changing tempo for aesthetic reasons can be an effective device [2]. Thus, whether or not to use “a click” to keep the tempo stable during music production or performance is a topic of artistic debate [2].

Just what range of tempo stability is considered aesthetic/professional is not well established. That tempo stability is generally *understood* by musicians to be desirable is clear. How much tempo-stability is *really* aesthetic? Consider the critically and commercially successful “Honky Tonk Woman” by the Rolling Stones, which increases its overall tempo from approximately 100 bpm to over 112 bpm in a period of about two minutes.

Tempo-stability has indirect production advantages: when all recordings/parts are synchronized to the same tempo, this makes it easier to manipulate and edit and them. Thus, tempo-stability may be artistically valuable because of *other* artistic license that it affords. Its possible that the aesthetic disadvantages of tempo-stability (if there truly are any) are simply outweighed by the technical/aesthetic advantages they afford.

The ability of performers to perform synchronized to a metronome without the “click” audible in the final performance/recording has been available at least since the 1950s. However, the nature of analog, tape-

based recording technology in this era would not be as conducive to this practice as modern technology—what’s more, the general view of “the studio as a creative instrument” had not yet widely developed, and musicians were more approach studio recording similarly to live performance, where using a click would have been inconceivable. It seems that the general practice of recording to a click, and the general expectation that “professional” music was recorded to a click, slowly emerged (in tandem) over time. The rise of digital recording in the 1980s, and more so the easy use of personal-computer-based Digital Audio Workstations from the 1990s on, make the feasibility and production advantages of recording to a click more acute. Inherently tempo-stable performance technology in the form of sequencers, drum machines, and turntables (as used by hip-hop DJs), appeared in the 1970s and proliferated rapidly through the 1980s–1990s.

This paper examines historical trends in “tempo-stability” through an analysis of 6,500 audio recordings from the era 1955–2019. We develop and compare several “tempo stability” metrics and measure their historical trends. We hypothesize that producing recordings “to click” increased in prevalence over the 1955–2019 time range.

Previous Work

The challenge of estimating the (possibly variable) tempo of commercial recordings is a prominent area of research in music information retrieval. However, research into how tempo fluctuates in a non-classical setting is relatively sparse. Dannenberg et al. [3] developed an algorithm that aids in characterizing tempo data collected from humans tapping the tempo while listening to a recording. Their research focused on developing an algorithm that was able to identify local tempo changes and describe overall tempo variation in a song. They used the algorithm to show that the variations in tempo for professional musician performances were much smaller than those of amateur musicians. Our research is similar in that we compare tempo variation, but differs in that we rely on data computed by Spotify using their tempo tracking algorithms. In a blog article, Lamere [4] graphed beat data collected from audio to allow visualization of tempo changes, demonstrating that proper scaling of the data on both the time and variation axes supports the ability to differentiate tracks that used a click track from those that didn’t. This study used the Spotify data to identify the songs with the smallest and largest standard deviation in tempo. Fig-

ures 1–5 show examples of songs with a very small standard deviation and those with larger standard deviations. While this study still did not identify a specific threshold in tempo variation that would characterize whether an artist used a click track or not, the anecdotal and visual evidence suggests that use of click tracks (and other production techniques such as sequenced reference tracks) in the recording process produces recordings with less significant variations.

Music scholars have also investigated ways that music technology can improve live performances. Dannenberg et al. [6] developed a framework to be used in developing new interfaces between technology and live performances[6]. Their work focused on ways that computer generated music can be synchronized with human players in live performance situations. The existence of hardware and software systems that allow for coordinated performances between humans and recorded and/or synthesized music suggests that popular musicians today have technology at their disposal that extends beyond a simple click track for synchronizing human and computer-generated musical parts. It also suggests that tempo stability may not be the only criterion for determining the extent to which recorded music (or even live music) is augmented by technology.

Micro-timing in popular music has been the subject of much discussion and research; however, attention has nearly uniformly focused on tempo-stable micro-timing. For example, Fruhauf et al. (2013) examined whether students perceive rhythm tracks with small variations in tempo as better or poorer quality. They found that rhythm tracks that were accurate to the quantized beat were perceived as being higher quality than those that hits that were either slightly ahead or slightly behind the beat. This finding is contrary to the sentiment expressed by Cornell and other popular musicians who feel that slight timing variations improve the quality of the track.

Project Overview

The goal of this project was to investigate whether the use of stable-tempo music technology by popular musicians has resulted in observable decreases of tempo variation in commercial music over the period 1955–2019. We sample 6,500 recordings from the Spotify collection, one hundred from each year in the target time span. We hypothesize that computational tempo estimates will show decreasing within-recording variation in the period after roughly 1985.

Measuring Tempo

The Spotify application programming interface (API) offers several precomputed tempo estimates for each song in the collection: estimates by measure, by section, and global estimates for entire recordings. For our purposes, we elect to use the by section values, as they represent large enough spans for reasonably stable tempo estimates, but with enough differentiation to identify changes of tempo over the course of a recording. Theoretically, changes of tempo are more likely to occur across sections than between, which also suggests the suitability of a section-by-section approach. The “sections” which divide the tempo estimates are also precomputed, automatically extracted features. The length of sections in the data vary from 3–xxx measures, xx.xxx–xxx.— seconds. The number of sections per recording quantiles. . . xxx

The exact nature and “true” accuracy of the Spotify API tempo estimates is not known.

xxx More about the Spotify algorithm? Ask Alexander?

To define tempo stability in a measurable way, we used the range of tempi of the sections of a song as returned by the Spotify API, and the overall change in tempo over the timespan of the song.

An important distinction can be made between “unstable” tempo—where variations are perceived as “variations” or shifts with categorically “the same” tempo—and true *tempo changes*, where a conscious, clear alteration of tempo is heard. How would such a distinction be made in practice? By doing so, we are able to conduct a wider study that looks at changes over time in a large corpus of popular music. In this study we look at historical trends in tempo stability to attempt to understand why some controversy surrounds this issue.

Methodology

Data

The Spotify database contains features and metadata for approximately forty million songs and is continuously updated with new material based on listener tastes. For this project, we sampled 6,500 recordings from the Spotify collection, one hundred tracks sampled randomly from each year in the range 1955–2019. We accessed the recording duration, section start and end timestamps, release year and artist, and section

level-tempo estimates. We also collected the beat start time for each beat and the bar start time and duration for each bar. For data collection, we used the following process:

Data Exclusion

A significant number of tracks (653xxx) had sections with no tempo computed; sections which the Spotify beat tracking algorithm (evidently) failed to identify a tempo. These null-tempo sections may reflect moments of silence, non-musical noise, spare sections, or heavily rubato, “free” sections, with no clear tempo—other unknown algorithmic artifacts. These null sections were overwhelmingly the first (xxx) or last (xxx) section in the recording, consistent with the idea that they are attributed to silence/noise/sparseness. We elected to discard null sections which were either the first or last section, and to discard completely any recordings (xxx) with null sections that were not the first or last section. It is possible that some sections are null, in part, because they have an unstable tempo; if this is the case, our data exclusion may reflect our results.

Harmonic errors

A known issue in tempo estimation (whether automated or human) is so-called “octave” error: estimates that are harmonic multiples of the “true” tempo, most often powers of two (hence octaves). Rhythmic music inevitably has multiple, interlocking metrical pulses, at integer ratios (forming musical “meter”), and human listeners often disagree which beat-level is the “true” tempo (Martens 2011). Not surprisingly, tempo extraction algorithms frequently evince the same problem, and to a higher degree. Fortunately, for our purposes the “true” tempo level is irrelevant so long as we can compare the *changes* in tempo across sections. The only issue that can arise is if the Spotify algorithm’s tempo estimates makes difference beat-level choices in different sections: for instance, estimating 80 beats per minute (bpm) in one section, then 160 beats per minute in the next.

Of course, “true” tempo changes do occur in music, and are usually harmonic: for instance, switching to “double time” mid piece. In our opinion, such harmonic changes are not “true” tempo changes, and once again can be ignored. We defined “harmonic” tempo changes as being changes by a ratio of 2, 3, 4, or their respective reciprocals.

To explore the issue of “harmonic errors” in our dataset, we investigated how often sequential sections within a song had tempos that were roughly ($\pm 10\%$) doubles or triples. Very few (xxx) doubling/halving changes were evident in the data, suggesting that the Spotify algorithm is successful at avoiding these (typically most common) errors. However, xxx cases of clear tripling errors were evident.

1. We ran a query against the Spotify API specifying the selection of any album with a vowel in the name and collected the names of 255,000 albums, along with the release year for each album. The albums contained a total of 1.9 million tracks.
2. For each year in the range 1955 – 2019, we used the api to collect the track level data for a random selection of 100 tracks from the list of albums
3. For each of the selected tracks, we collected the section and bar-level statistics from Spotify using the Spotify API. The statistics we collected were song tempo, song length, section start and duration, section tempo and section key and mode. Our Preliminary analysis showed that using the section-level statistics allowed us to summarize tempo stability data in a concise, easy to interpret manner.
4. For each track we computed additional statistics - minimum and maximum section tempo, the range of tempi from minimum to maximum and the difference in tempo between the sections. We also computed the normalized tempo range as the difference between the minimum and maximum tempo divided by max tempo, and the normalized difference in tempo as the difference between section tempi divided by the max tempo. Finally, we computed the standard deviation of the section tempi for each song.
5. A significant number of tracks (653) had sections where no tempo was computed. These sections could be characterized as sections that did not contain enough sounds that Spotify’s beat tracker was able to identify as onsets. The tempo for these sections was set to the median of the tempi for the remaining sections. A small number of tracks had sections whose tempo differed by a factor of approximately three from neighboring sections. We determined these to be artifacts of the algorithms used to calculate the tempo by listening to several of the tracks. We did not adjust the tempo of these sections.
6. For the normalized tempo range and the normalized tempo difference, we calculated a histogram of the values. We also calculated a regression analysis by year of the normalized tempo ranges

and the normalized tempo differences.

7. We sorted the tracks by the standard deviation of the section tempos to identify songs that had very stable tempi and those that had highly variable tempi. We listened to songs at various points in the range to try to summarize the tempo attributes.

Calculations

We propose the following “robust” tempo variability measure, which is unaffected by doubling or tripling harmonic errors. Given two tempos, t_1 and t_2 ,

$$\min(2^{(\log_2 t_1 - \log_2 t_2) \bmod 1}, 3^{(\log_3 t_1 - \log_3 t_2) \bmod 1})$$

Results

Based on the goodness-of-fit statistics for both the normalized tempo range and the normalized tempo difference, the data in this sample does not support the hypothesis that there was a reduction in the variability in tempo over the period (tempo difference regression $r=.1328$, tempo range regression $r=.1301$). The histogram for the normalized tempo difference showed that 87% of the songs had overall tempo differences between the sections of between negative 5.3% and 7.2%. The histogram for the normalized tempo range showed that 86% of the songs had tempo ranges of between 1% and 10.5% (about 1–10 bpm).

Discussion and Future Work

Based on the goodness of fit for the regression tests, the data does not support the hypothesis that musicians used technology to significantly reduce the variability in tempo in their recordings. Previous research has shown that it is possible to make recordings that have near 0 variability, but musicians have not taken advantage of this capability. This data shows that throughout the period studied tempo stability was extremely high (variability was low). Based on anecdotal evidence as in [5], at least some musicians believed that trying to improve the “tightness” of a recording by using click tracks reduces the authenticity of the track. However other research such as (Fruhauf, Kopiez, and Platz 2013) suggests that audiences do perceive performances to be higher quality when the tempo of a song stays relatively constant. Assuming the professional musicians are aware (to

some extent) of both these points of view, it may be that for most bands, the level of tempo stability that can be achieved without going to extraordinary lengths is “good enough”.

One avenue for further research is to determine whether different sampling methods would produce different results. For example, in certain genres such as EDM and hip-hop, the heavy use of sequenced midi tracks, samples and loops would seem likely to produce tracks with very stable tempi. One potential future study might select tracks from the spotify collection based on the standard deviation of the section tempi to see whether certain genres are more likely to have stable tempi than others.

References

- [2] <https://music.stackexchange.com/questions/30536/is-it-acceptable-to-change-tempo-in-the-middle-of-a-song-or-is-this-a-bad-idea> (Accessed October 27, 2019)
- [3] R. B. Dannenberg and Mohan, Sukrit, “Characterizing Tempo Changes in Musical Performance.” 05-Aug-2011.
- [4] Paul Lamere: In Search of the Click Track, <https://musicmachinery.com/2009/03/02/in-search-of-the-click-track/>
- [5] Chris Cornell & Tim Commerford - Cuba <https://www.youtube.com/watch?v=LUjcrfa7u38>
- [6] R. B. Dannenberg, “New interfaces for popular music performance,” in *Proceedings of the 7th international conference on New interfaces for musical expression - NIME '07*, New York, New York, 2007, p. 130.
- [7] E. Räsänen, O. Pulkkinen, T. Virtanen, M. Zollner, and H. Hennig, “Fluctuations of Hi-Hat Timing and Dynamics in a Virtuoso Drum Track of a Popular Music Recording,” *PLoS ONE*, vol. 10, no. 6, p. e0127902, Jun. 2015.
- [9] S. Dixon, “An Empirical Comparison of Tempo Trackers,” p. 9.
- [10] H. Honing, “When a Good Fit is not Good Enough: A Case Study on the Final Ritard,” p. 4, 2004.
- Fallows, David. 2001. “Metronome.” In *The New Grove Dictionary of Music and Musicians*. Oxford University Press. <https://doi.org/10.1093/gmo/9781561592630.article.18521>.
- Fruhauf, J., R. Kopiez, and F. Platz. 2013. “Music on the Timing Grid: The Influence of Microtiming on the Perceived Groove Quality of a Simple Drum Pattern Performance.” *Musicae Scientiae* 17 (2): 246–60. <https://doi.org/10.1177/1029864913486793>.
- Hudson, Richard. 2001. “Rubato.” In *The New Grove Dictionary of Music and Musicians*. Oxford University Press. <https://doi.org/10.1093/gmo/9781561592630.article.24039>.
- Martens, Peter A. 2011. “The Ambiguous Tac-tus: Tempo, Subdivision Benefit, and Three Listener Strategies.” *Music Perception: An Interdisciplinary Journal* 28 (5): 433–48.
- Page, Jimmy. 2014. “Jimmy Page: How Stairway to Heaven Was Written.” Edited by Amanda Bruckshaw. *Www.youtube.com/Bbcnews*. BBC News. <https://www.youtube.com/watch?v=DDo4CA13LbY>.
- Prögler, J. A. 1995. “Searching for Swing: Participatory Discrepancies in the Jazz Rhythm Section.” *Ethnomusicology* 39 (1): 21. <https://doi.org/10.2307/852199>.
- Sloboda, John A. 1983. “The Communication of Musical Metre in Piano Performance.” *The Quarterly Journal of Experimental Psychology Section A* 35 (2): 377–96. <https://doi.org/10.1080/14640748308402140>.
- Temperley, David. 2004. “Communicative Pressure and the Evolution of Musical Styles.” *Music Perception: An Interdisciplinary Journal* 21 (3): 313–37.