SWING RATIO ESTIMATION

Ugo Marchand
STMS IRCAM-CNRS-UPMC
1 pl. Igor Stravinsky, 75004 Paris, France
ugo.marchand@ircam.fr

Geoffroy Peeters
STMS IRCAM-CNRS-UPMC
1 pl. Igor Stravinsky, 75004 Paris, France
geoffroy.peeters@ircam.fr

ABSTRACT

Swing is a typical long-short rhythmical pattern that is mostly present in jazz music. In this article, we propose an algorithm to automatically estimate how much a track, a frame of a track, is swinging. We denote this by swing ratio. The algorithm we propose is based on the analysis of the auto-correlation of the onset energy function of the audio signal and a simple set of rules. For the purpose of the evaluation of this algorithm, we propose and share the "GTZAN-rhythm" test-set, which is an extension of a well-known test-set by adding annotations of the whole rhythmical structure (downbeat, beat and eight-note positions). We test our algorithm for two tasks: detecting tracks with or without swing, and estimating the amount of swing. Our algorithm achieves 91% mean recall. Finally we use our annotations to study the relationship between the swing ratio and the tempo (study the common belief that swing ratio decreases linearly with the tempo) and the musicians. How much and how to swing is never written on scores, and is therefore something to be learned by the jazz-students mostly by listening. Our algorithm could be useful for jazz student who wants to learn what is swing.

1. INTRODUCTION

1.1. Swing

Music is not always played exactly as written on the score. These deviations from the score constitute the musician personal interpretation and are precisely chosen by the musician to make music more lively or to convey an emotion.

In this paper we focus on timing deviations. We distinguish different types of systematic timing deviation. The *rubato* refers to a rhythmic freedom, in which tempo is sped up then slowed down to make music more expressive. It was usual in Romantic period. The *notes inégales* is a performance practice of the Baroque and Classical period in which some notes with equal durations are performed with unequal durations. Finally the *swing* is often found in jazz music but also other music styles as we will see in part 3. Our work focuses on the swing estimation.

Swing is present at a specific time level. It describes the specific interaction between note durations at the eight-note level. When swing is present, two consecutive eight-notes are played following a long-short pattern: the first eight-note is lengthened, while the second is shortened proportionally.

The swing ratio is defined by the ratio between the duration of the long eight note and the short one. Common swing ratios found in jazz music are 1:1 (no swing at all), 2:1 (triple feel), 3:1 (hard swing). These three example are presented in Figure 1. It should be noted that the swing ratio value is not limited to the values 2 or 3, but it can take all floating values between 1 and 3.5.

Figure 1: *Different swing ratio. From left to right : 1:1 (no swing), 2:1 (triple feel), 3:1 (hard swing)*

Swing ratios are not written on the score, they are implicit. Thus, swing ratio can vary considerably between two musicians or even inside a single music piece.

1.2. Related works

On swing. In [1], Friberg et al. study the exact duration ratio between the long eight-note and the short one in recordings. The author shows that the swing ratio varies linearly with the tempo. At slow tempi, the swing ratio can reach 3.5:1. At fast tempi, it goes down to 1:1. The authors also show that the minimum absolute duration of the short eight-note in the long-short pattern is around 100ms. This suggests a physical limit to swing ratio, maybe due to perceptual factors.

In [2], Honing et al. show that professional jazz drummers have an extremely precise control over the swing ratio they want to achieve. The authors also found no evidence that the swing ratio scale linearly with tempo, as was suggested previously.

On swing ratio estimation. In [3], Gouyon et al. introduced a swing modification tool. The swing ratio is estimated by two methods. The first one is based on the estimation of the position of the second peak d_s in the inter-onset-histogram which is supposed to corresponds to the duration of the short eight-note. The swing ratio is then $\frac{d_e+d_s}{d_e-d_s}$ where d_e is the theoretical duration of a non-swinging eight-note and d_s is the estimated duration of the second peak of the histogram. The second method compares the inter-onset-histogram to several predefined histogram models (representing different swing ratio), and choose the model that best represent the inter-onset-histogram.

In [4], Laroche presents a joint estimation of the tempo, the downbeat and the swing ratio. For this, he first estimates the time-positions where the energy in a given frequency band grows quickly. Then he exhaustively tests all the triplet (tempo, downbeat time, swing ratio) and keeps the one that have the best likelihood.

1.3. Paper overview and organization

In this paper, we propose a novel algorithm for estimating the swing ratio of a track (part 2). We then evaluate its performance for a task of finding music tracks with/without swing. For this, we present the test-set we created for the purpose of swing detection

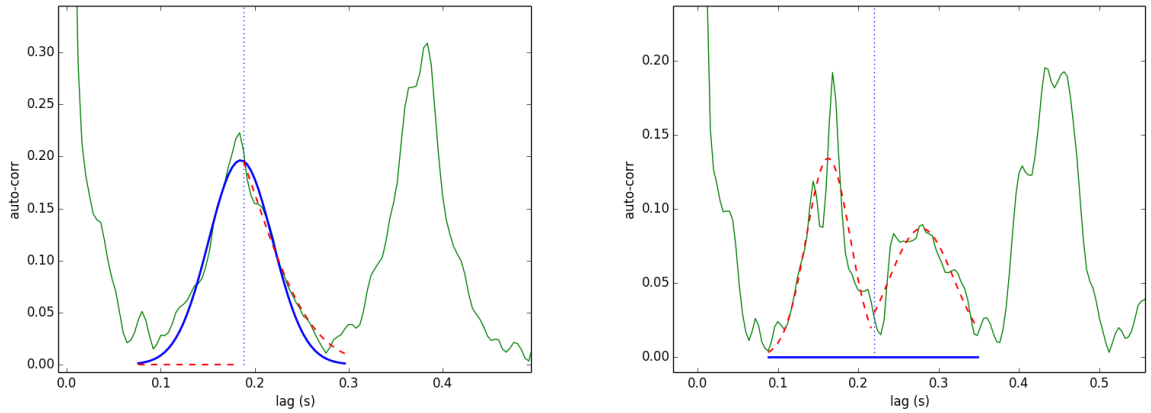

Figure 2: *Auto-correlation function for [Left] a signal without swing ('disco.00020.wav', first frame), [Right] a signal with swing ('jazz.00099.wav', first frame). The auto-correlation of the onset-energy-function (OEF) of the signal is the green thin line. The dotted vertical line (around 0.2 s.) corresponds to the theoretical duration of an eight-note. The thick blue line corresponds to the fitting of the peak of a non-swinging eight-note. The two red dotted lines correspond to the fitting of the peaks of the swing pattern (short and long eight-note).*

(part 3). We present our experiments and discuss their results in part 4.

2. SWING RATIO COMPUTATION

To estimate the swing ratio, we first compute an onset-energy-function[1] OEF (part 2.1). The auto-correlation function of the OEF allows highlighting the various metrical level of a rhythm pattern therefore the specific irregularities of the eight-note due to the swing (part 2.2). From the positions of the peaks in the auto-correlation function we then estimate the swing ratio (part 2.3).

2.1. Signal pre-processing

We calculate the onset-energy function $o(t)$ using the method proposed by Ellis [5][2].

$o(t)$ is then analyzed using a frame-analysis with a window length of 16s[3], and a hop-size of 1s[4]. For each frame, we then compute the normalized (at lag zero) auto-correlation of $o(t)$:

$$r(\tau) = \frac{1}{r(0)} \sum o(t)o(t-\tau) \quad (1)$$

2.2. Auto-correlation

The auto-correlation function of $o(t)$ allows highlighting the various metrical level of a rhythm pattern. Therefore it highlights the specific irregularities of the eight-note due to the swing. We illustrate this in Figure 2. The two panels illustrate the auto-correlation function for a signal without and with swing.

In a signal without swing (Figure 2 [Left]), there is one peak representing the tactus (the duration of a quarter-note) at 0.40s and one representing the duration of the eight-note at 0.20s. For a signal with swing (Figure 2 [Right]), the peak representing the quarter-note is still present at 0.45s, but the peak representing the duration of the eight-note is split into two peaks : one for the duration of the short eight-note at 0.15s and one for the duration of the long eight-note at 0.30s.

The goal of swing detection is to distinghish both cases.

2.3. Peak finding

2.3.1. Eight-note duration estimation

In order to distinguish the two cases we need to know the theoretical position of the eight-note without swing. This position is derived from an estimation of the tempo (the tactus-level). For this we use the joint estimation of beat, downbeat and tempo algorithm of Peeters et al. [6]. The tempo value is hypothesised to be constant over the whole track duration. If we denote by T the estimated tempo in bpm (beat per minute), the theoretical duration (in second) of the eight-note d_e is then computed as $d_e = \frac{1}{2}\frac{60}{T}$ (half of the duration in second of a quarter-note : $\frac{60}{T}$).

[1] A onset-energy-function is a function taking high values when an onset is present and low values otherwise.

[2] The OEF is computed as follow: first, the audio is resampled to 8kHz, then the short-term Fourier transform (STFT) magnitude (spectrogram) using 32 ms windows and 4 ms step between frames is computed. Then it is converted to 40 Mel bands. The Mel spectrogram is converted to dB, and the first-order difference along time is calculated in each band. Negative values are set to zero (half-wave rec-tification), then the remaining, positive differences are summed across all frequency bands. This signal is passed through a high-pass filter with a cutoff around 0.4 Hz to make it locally zero-mean, and smoothed by convolving with a Gaussian enveloppe about 20 ms wide.

[3] A long window is required because there is not always a swinging eight-note around each beat. For some tracks, the swing information can be very sparse. Thus it is necessary to have a long window.

[4] A 1s hop-size was arbitrarily chosen. Taking a smaller hop-size wouldn't make sense considering that there are, at most, 4 beats per 1 second in our database.

Table 1: *Distribution into swing, noSwing and ternary of the proposed GTZAN-rhythm test-set.*

	blues	classical	country	disco	hip-hop	jazz	metal	pop	reggae	rock	**Total**
# Track with swing	44	1	15	0	0	45	0	0	25	6	**136**
# Track with noSwing	52	92	80	98	100	54	94	99	74	92	**835**
# Track with ternary	4	7	5	2	0	1	6	1	1	2	**29**

2.3.2. Eight-note detection in the auto-correlation

The values of the eight-note duration can be derived from the peak positions in the auto-correlation function of OEF. We also use the height of the peaks and the width of the corresponding lobes to have information on the reliability of these peaks. Because peak *finding* algorithms (such as finding local maxima) do not lead to good results (since the auto-correlation is often noisy), we propose to use a peak *fitting* algorithm [7, 8]. In this a predefined shape is fitted locally to a signal. This shape is defined by a Gaussian function (this leads to better results than the use of a 2nd order polynomial function).

In the absence of swing a peak must be present at d_e, in the presence of swing this peak is split in two peaks in the two intervals $\left[\frac{d_e}{2}, d_e\right]$ and $\left[d_e, \frac{3d_e}{2}\right]$ respectively. We therefore looks for peaks in these three intervals.

- The first interval $\left[\frac{d_e}{2}, d_e\right]$[5] corresponds to the short eight-note.
- The second interval $\left[d_e, \frac{3d_e}{2}\right]$ corresponds to the long eight-note.
- The third interval $d_e \pm \frac{d_e}{2}$ correspond to a non-swinging eight-note (this interval is here only for illustration purpose, it is not used by our algorithm).

In order to estimate the peak within each interval, we use a non-linear least squares algorithm to fit a Gaussian function. Thus we get 3 parameters for each peak: the amplitude A, the standard deviation σ and the mean μ (μ defines the estimated peak position). The parameters of the short and the long eight-notes are denoted A_s, σ_s, μ_s and A_l, σ_l, μ_l respectively.

2.3.3. Swing detection

Given the 6 parameters described above, we propose the following set of rules to decide if there is swing or not. The Gaussian functions representing the short and the long eight-notes must comply with the following rules:

- positive amplitudes $A_s > 0$, $A_l > 0$
- small widths $\sigma_s < \frac{d_e}{4}$, $\sigma_l < \frac{d_e}{4}$
- positions in the right interval $\frac{d_e}{2} < \mu_s < d_e$ and $d_e < \mu_l < \frac{3d_e}{2}$

If all the conditions are met, the frame is classified as swing.

Illustrations: Swing detection is illustrated in Figure 2. On both figures, we show the auto-correlation of the OEF of the signal (green thin line). The theoretical duration of the eight-note d_e is given by the vertical dotted line. We also show the 3 peaks that were fitted as previously described: the thick blue line is the Gaussian fit for the non-swinging eight-note, and the two red dashed lines correspond to the fit of the short and the long eight-note of the swing pattern. Given the parameters of these fitted Gaussian functions and using the proposed rules described above, our algorithm decide that there is no swing on the Left panel of Figure 2 while there is swing on the Right panel.

2.3.4. Swing ratio estimation

For each frame classified as swing, its swing ratio s_r is finally computed as $s_r = \frac{\mu_l}{\mu_s}$.

Illustrations: For the previous example (Figure 2 [Right]), the swing ratio would be $s_r = \frac{0.28}{0.16} = 1.75$.

3. THE GTZAN-RHYTHM TEST-SET

There is currently no test-set on which the swing ratio is annotated. It is therefore difficult to compare numerically our results to previous works.

We therefore created our own test-set by extending the widely used GTZAN test-set for music genre[6]. While there exist controversies related to the use of the GTZAN test-set for music genre classification[9], its audio content is however represensative of real commercial music of various music genre. Also, this test-set has a good balancing between tracks with swing (blues and jazz music) and without swing.

Annotations: Each track of the test-set has been annotated into downbeat, beat/tactus (quarter-note) and eight-note positions[7]. In order to take into account tempo or rhythm pattern variations (the swing ratio is never constant over time) the annotation is performed over the whole duration of each track. A time region is said to contain swing if the eight-note positions are not exactly at the middle between the two adjacent quarter-notes.

The distribution into swing, non-swing and ternary is shown in Table 1. As can been seen, most of the swinging tracks are jazz (45 tracks) or blues (44 tracks). Swing is also present in country (15 tracks), reggae (25 tracks) and rock (6 tracks). The classical track classified as swing has a "march" rhythm. In total the test-set contains 136 swinging tracks, 835 non-swinging tracks and 29 ternary tracks.

4. EVALUATION

We evaluate our proposed algorithm for two tasks.

[5] The search ranges imply that the swing ratio is comprised between 1 and 3.

[6] The GTZAN test-set is made of 1000 audio excerpts of 30 s. duration evenly distributed in 10 music genres (blues, classical, country, disco, jazz, hip-hop, metal, pop, reggae, rock).

[7] We only annotated the eight-note position when swing exist or in case of ternary rhythm

Table 2: *Results of swing detection. Each column represents the results on a subset of the GTZAN-rhythm test-set. For this, each genre is subdivided into frames with swing and without. For each subset, we indicate its number of frames (# frame), the percentage of correct recognition of the automatic tempo estimation (Tempo-Estimation) and the two recalls for the classes swing and noSwing obtained using our algorithm (Exp-estimatedTempo and Exp-annotatedTempo).*

	blues		**classical**		**country**		**disco**		**hiphop**	
	swing	*noSwing*	*swing*	*noSwing*	*swing*	*noSwing*	*swing*	*noSwing*	*swing*	*noSwing*
# Frame	686	714	92	1308	294	1106	28	1372	0	1400
Tempo-Estimation (%)	63	74	41	63	61	72	100	96	0	97
Exp-estimatedTempo (recall %)	63.8	93.6	57.6	99.6	52.7	98.7	100	99.3	0	100
Exp-annotatedTempo (recall %)	95.3	94.1	92.4	99.6	84.4	97.6	100	99.3	0	100
	jazz		**metal**		**pop**		**reggae**		**rock**	
	swing	*noSwing*	*swing*	*noSwing*	*swing*	*noSwing*	*swing*	*noSwing*	*swing*	*noSwing*
# Frame	630	770	84	1316	14	1386	364	1036	112	1288
Tempo-Estimation (%)	37	56	83	64	100	87	57	75	62	84
Exp-estimatedTempo (recall %)	23.8	96.4	61.9	99.8	100.0	100	54.4	91.4	48.2	98.2
Exp-annotatedTempo (recall %)	60.2	97.4	76.2	99.1	100.0	99.6	93.4	95.3	83.0	99.1

In part 4.1, we test the ability of our algorithm (part 2.3.3) to correctly detect the tracks with swing and the ones without. Considering that the swing presence can vary over time, this is performed on a frame basis.

In part 4.2, we use our swing ratio annotation (part 2.3.4) for two experiments. First we test the commonly accepted belief that the swing factor decrease linearly with the tempo. Then, we test if the swing factor can be used to recognize the musicians.

4.1. Swing detection

4.1.1. Exp-estimatedTempo

The algorithm we proposed in part 2.3.3 allows deciding if a frame of a given track has swing or not. We test our algorithm on each frame of each track of the GTZAN-rhythm test-set. Because, the method proposed in 2.3.3 uses a 16 s. window (for the computation of the auto-correlation), we compare our estimation to the corresponding local annotation. The corresponding local annotation is computed as follow: a 16-sec frame is said to be swing if it contains at least a single swing annotation.

In our experiment we assimilated ternary frames to swinging frames. This is because our method can not distinguish a ternary frame from a swinging jazz frame with the 'triple feel' (swing ratio of 2:1).

Results: The results are presented in Table 2. The first, second and third rows indicate the frame repartition in genre and in swing/noSwing classes. For example, we see that there are much more swinging frames in jazz (630) and blues (686) than in hiphop (0) or pop (12). The results of Exp-estimatedTempo are presented in terms of Recall[8] for each class (in %).

As one can see, the Recall for the noSwing classes are all very good (from 91.4% to 100%). In the opposite, the Recall for swing is not that good and strongly depends on the genre: in jazz, it is only 23.8%, in rock is is 48.2%, country 52.7% and reggae 54.4%. It is slightly better for blues (63.8%), classical (57.6%) and metal (61.9%). The Recall is perfect for disco and pop. However the number of swing frames to be detected is very small for these two classes.

The summary of the results over the genres is indicated in Table 3. The noSwing Recall is very high (98%) which means that almost all noSwing frames have been correctly detected. While the swing Recall is low (49.5%) the swing Precision is rather high (84%) which means that in most cases when a frame have been detected as swing it is indeed annotated as swing.

Table 3: *Confusion matrix for the estimation of the classes noSwing and swing for Exp-estimatedTempo.*

Annotated / Detected	noSwing	Swing	Recall
noSwing	**11479**	217	98.14%
Swing	1162	**1142**	49.57%
Precision	90.91%	84.03%	

4.1.2. Tempo-Estimation

Since our swing detection algorithm relies on a previous tempo estimation (the one provided by the algorithm of [6]), we test if the bad results obtained for the swing detection (low Recall) are due to wrong tempo estimations. For this, we compare the tempo estimation T_e to our manually annotated tempo T_a. An estimation is said to be correct if $\frac{|T_a - T_e|}{|T_e|} < 4\%$.

Results: The detailed results are presented in Table 2. The lowest results are obtained for genre for which tempo variation or timing can be large (*rubato* in classical music, expressive music timings in jazz). Highest results are obtained for genre with a steady tempo (hip-hop, pop and disco). Our tempo estimation is based on the assumption that the tempo is constant over a track, so tracks with a lot of tempo changes are poorly classified.

Overall, 75.1% of the tracks have their tempo correctly estimated. This value falls down to 56.4% if we consider only the tracks that have swing. This can explain the bad results obtained

[8]The recall is the number of items detected for a class divided by the total number of annotated item for that class.

for the swing detection (Recall=49.5%) but the good results for the noSwing detection (98%).

4.1.3. Exp-annotatedTempo

Because of this, we redid the swing detection evaluation using manually annotated tempo instead of the automatically estimated one. The manually annotated tempo is derived from the beat annotations of our GTZAN-rhythm.

Results The detailed results are presented in Table 2. Using annotated tempo always improves the results for the swing class. The Recalls of blues, classical, country jazz, reggae and rock are 32 to 40% higher. The Recall for the class metal/swing is 15% higher, and the other Recalls were already perfect in the previous experiment (disco, pop).

Concerning the noSwing classes, the mean-Recall are now all comprised between 94.1 and 100%, which is better than in Exp1.

The summary of the results over the genres is indicated in Table 4. Using the annotated tempo improved a lot the results: the global mean-Recall increases from 73.9% to 90.6%.

These experiments show that we can estimate the swing knowing the 8th-note tactus. However deciding which level is the 8th-note remains a far more complicated problem.

Table 4: *Confusion matrix for the estimation of the classes noSwing and swing for Exp-annotatedTempo.*

Annotated / Detected	noSwing	Swing	Recall
noSwing	**11514**	182	98.44%
Swing	399	**1905**	82.68%
Precision	96.65%	91.27%	

4.2. Swing ratio as a function of tempo and artist

In this part, we use the annotation of the GTZAN-rhythm into swing ratio (part 2.3.4) for two experiments. First we test the commonly accepted belief that the swing factor decrease linearly with the tempo. Then, we test if the swing factor can be used to recognize the musicians.

4.2.1. Swing ratio as a function of tempo

In Figure 3, we illustrate the annotated swing ratio as a function of the tempo for the subset of our test-set corresponding to blues and jazz music (where swing is often present).

We test the commonly accepted belief that swing is a linearly decreasing function of tempo. This was proposed by the experiment of Friberg [1]. We superimposed to Figure 3, the linear regression (dotted line) used in Figure 1 of Friberg [1]. We can see that in our experiment the swing ratio does not scale linearly with the tempo. Our observed swing ratios are all far from the linear assumption proposed by Friberg (dotted line) at tempi < 130, and they does not show the same trend. Our results are therefore in agreement with the results of Honing et al. [2] saying that there may be no correlation at all between tempo and swing ratio. We also see that there may be a preference for the triple feel swing ratio (2:1) in our test-set, as about a third of our examples have a swing ratio around 2 (comprised between 1.9 and 2.1).

4.2.2. Swing ratio as a function of an artist

In Figure 3, we also indicates the artist[9] [10] corersponding to each swing ratio.

We test if there is a relationship between the artist and the swing ratio. We see that some artists may tend to play with low swing ratio (Robert Johnson) or high swing ratio (Magic Slim). Tracks from Stevie Ray Vaughan share the same swing factor but also the same tempo. (it is therefore difficult to say if the swing is characteristic of the artist or of the tempo). Except these examples predicting the musicians from its swing ratio and tempo seems difficult.

4.3. Applications

Our swing ratio estimation could be useful to jazz students who want to learn swing and to musicologists who would benefit from a swing estimation tool to study jazz music/ musicians.

How much to swing is often learned by students by listening to a lot of jazz recordings. It is not something that can be learned from books, and it requires a lot of practice to master. One application of this swing estimation could be a system helping jazz students to learn how much to swing :

- A jazz teacher plays a jazz piece and its swing ratio is estimated.
- The student plays the same piece and the computer says if the student swings enough or too much (compared to the teacher's reference), allowing the student to learn the right swing ratio.

It should be noted that we do not indicate what is the ideal swing ratio for a music piece.

5. CONCLUSION AND FUTURE WORKS

In this article, we presented an algorithm to automatically detect the swing presence in an audio music track and estimate its swing ratio. For the purpose of the evaluation of this algorithm, we proposed and shared the "GTZAN-rhythm" test-set, which is an extension of a well-known test-set by adding annotations of the whole rhythmical structure (downbeat, beat and eight-note positions). We showed that the algorithm allows a mean-Recall of 74% for the recognition of swing/ noSwing. By analysing the errors, we discovered that most of the errors are due to wrong tempo estimation which is used as input parameter of our algorithm. Using annotated tempo as input of our algorithm, allowed to reach 91% mean-Recall. We finally used our annotations to study the relationship between the swing ratio and the tempo and the musicians. Future works will concentrate on the development of a swing ratio estimation algorithm which does not necessitate this tempo estimation. Finally, given that swing ratio is a rhythmic descriptor, it could be used as input, among other audio descriptors, to machine learning algorithms to perform other Music Information Retrieval tasks.

[9] We are grateful to Bob Sturm for sharing the artist information of the GTZAN tracks.

[10] Given that the various tracks of an artist are coming from the same album we suppose that the musicians are the same for a given artist.

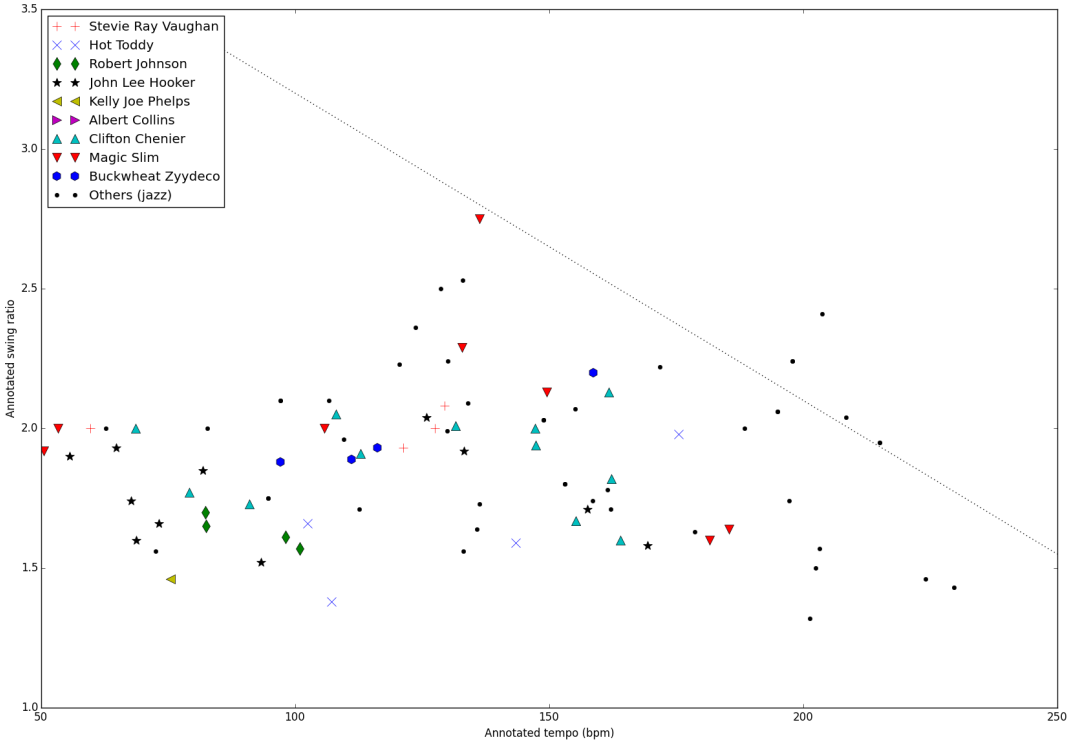

Figure 3: *Swing ratio as a function of tempo. Each type of point marker correspond to a different artist. The dotted line correspond to the swing as a function of tempo from the experiment of Friberg [1].*

Downloading the GTZAN-rhythm test-set.

We deeply thank Quentin Fresnel for helping us with the annotations of the "GTZAN-rhythm" test-set. The "GTZAN-rhythm" test set can be downloaded at `http://anasynth.ircam.fr/home/media/GTZAN-rhythm/`.

Acknowledgements

This work was founded by the French government Programme Investissements d'Avenir (PIA) through the Bee Music Project.

6. REFERENCES

[1] Anders Friberg and Andreas Sundström, "Swing ratios and ensemble timing in jazz performance: Evidence for a common rhythmic pattern," *Music Perception*, vol. 19, no. 3, pp. 333–349, 2002.

[2] Henkjan Honing and W Bas De Haas, "Swing once more: Relating timing and tempo in expert jazz drumming," 2008.

[3] Fabien Gouyon, Lars Fabig, and Jordi Bonada, "Rhythmic expressiveness transformations of audio recordings: swing modifications," in *Proc. Digital Audio Effects Workshop (DAFx)*, 2003.

[4] Jean Laroche, "Efficient tempo and beat tracking in audio recordings," *Journal of the Audio Engineering Society*, vol. 51, no. 4, pp. 226–233, 2003.

[5] Daniel PW Ellis, "Beat tracking by dynamic programming," *Journal of New Music Research*, vol. 36, no. 1, pp. 51–60, 2007.

[6] Geoffroy Peeters and Helene Papadopoulos, "Simultaneous beat and downbeat-tracking using a probabilistic framework: theory and large-scale evaluation," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1754–1769, 2011.

[7] Kenneth Levenberg, "A method for the solution of certain non–linear problems in least squares," 1944.

[8] Donald W Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial & Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[9] Bob L Sturm, "The gtzan dataset: Its contents, faults, and their effects on music genre recognition evaluation," *IEEE Transactions on Audio, Speech and Language Processing*, 2013.