Advantages of nonstationary Gabor transforms in beat tracking

Andre Holzapfel
INESC Porto
Porto, Portugal
aholza@inescporto.pt

Gino Angelo Velasco
Numerical Harmonic Analysis Group
Faculty of Mathematics
University of Vienna
Vienna, Austria
gino.velasco@univie.ac.at

Nicki Holighaus
Numerical Harmonic Analysis Group
Faculty of Mathematics
University of Vienna
Vienna, Austria
nicki.holighaus@univie.ac.at

Monika Dörfler
Numerical Harmonic Analysis Group
Faculty of Mathematics
University of Vienna
Vienna, Austria
monika.doerfler@univie.ac.at

Arthur Flexer
Austrian Research Institute for Artificial Intelligence (OFAI)
Vienna, Austria
arthur.flexer@ofai.at

ABSTRACT

In this paper the potential of using nonstationary Gabor transform for beat tracking in music is examined. Nonstationary Gabor transforms are a generalization of the short-time Fourier transform, which allow flexibility in choosing the number of bins per octave, while retaining a perfect inverse transform. In this paper, it is evaluated if these properties can lead to an improved beat tracking in music signals, thus presenting an approach that introduces recent findings in mathematics to music information retrieval. For this, both nonstationary Gabor transforms and short-time Fourier transform are integrated into a simple beat tracking framework. Statistically significant improvements are observed on a large dataset, which motivates to integrate the nonstationary Gabor transform into state of the art approaches for beat tracking and tempo estimation.

General Terms

Theory

Keywords

Beat tracking, nonstationary Gabor transform, music information retrieval

1. INTRODUCTION

The task of estimating the time-instants at which a person would tap his or her foot to a piece of music is known as beat tracking [6]. The related metrical level, which is also referred to as *tactus*, finds itself somewhere in the middle of the meter hierarchy of a piece of music, between the lowest (*i.e.* fastest) meter level, referred to as tatum [3], and the measure level.

Generally, beat tracking methods include specific processing steps. First, the input audio is used to *compute features*. There are several possibilities for the choice of the features, see Bello *et al.* [2] for a detailed overview. The obtained features have a lower sampling frequency than the initial audio file due to applied windowed analysis. They are aimed at stressing the phenomenal accent of note onsets in the signal, as these onsets appear at time instances that coincide with impulse positions at one or several levels of the meter hierarchy. For this reason, these features can be referred to as Onset Strength Signal (OSS). The OSS obtained from the audio file is used to *derive the tempo* of a piece of music (tempo in $bpm = 60/\text{beat period}$). As described in Grosche and Müller [8], this is usually performed by applying comb filters, autocorrelation methods or short-time Fourier Transforms (STFT). Because the tempo may vary throughout the piece, parameters of all those methods have to be adjusted in order to find a compromise between stable tempo estimations and the ability to react to tempo changes. Thus, tempo induction results in a series of beat period estimates at a certain sampling rate T_{hop}, and a pulse train can be generated with periods equal to these estimated beat periods. Then, a temporal alignment of this pulse train with the beat pulse a human would perceive in the piece of music has to be determined. This step will be referred to as *beat phase estimation*. For these estimations, various methods have been incorporated, such as the use of statistical models [9] or multiple agents [11].

In this paper, the goal is not to reach superior beat tracking results, but to investigate the possible advantages of using a novel transform for beat period and phase estimation, the nonstationary Gabor transform (NSGT) [1]. For this reason, more sophisticated ways to perform tempo induction are left for future work, in order to emphasize the charac-

teristics of the transform. As will be shown, NSGT has two properties that make it interesting for beat tracking. First, it is possible to adjust the frequency resolution in a flexible way, which can improve the resolution for low frequencies (*i.e.* slow tempi) over *e.g.* the STFT. Second, for this transform a perfect inverse transform exists. This makes it an interesting alternative to other approaches for increasing resolution, such as the combination of STFT and autocorrelation function [10]: A key feature of the NSGT is that once a beat period has been estimated, the beat phase information is contained in the phase information of the complex analysis coefficients. Thus, beat tracking can be performed by computing the inverse transform after appropriate filtering in the transform domain. Demonstrating the potential of such an approach for beat tracking using the NSGT is the main goal of this paper.

In Section 2 the mathematical background will be introduced. Thereafter, Section 3 explains the method in which the NSGT will be applied to beat tracking. A system [8] that applies an STFT will serve for comparison with our approach, which will be summarized in Section 4, along with the evaluation criteria and the used evaluation data. Sections 5 and 6 describe our results and give the conclusions, respectively.

2. BACKGROUND

Several signal transforms providing more flexibility than the STFT have been proposed, cf. [7, 12] and references therein. The method proposed in this paper is based on an NSGT [1], which is a generalization of the widely used STFT. It allows for a time-frequency representation with adaptive frequency resolution.

To this end, first we compute the frequency representation (FFT) $\hat{x}$ of the whole input signal x. Afterwards, the inverse FFT (IFFT) of windowed sections of $\hat{x}$ (frequency slices) is computed. Analogous to the STFT, which computes, for each time slice, samples of a time-localized spectrum, the output of this procedure for each frequency slice is a sampled bandpass of the original input signal. In contrast to the STFT, we do not use the same window function at each position, but windows which vary with respect to frequency. Thus we achieve an adapted frequency resolution for each bin corresponding to a frequency slice.

Given a real-valued signal x of length L, and K *filters* (window functions) W_k with *support* (the interval where the vector is nonzero) of length $M_k \leq L$ centered at ω_k, the NSGT coefficients are given by

$$c_{k,m} = \frac{1}{M_k} \sum_{j=0}^{L-1} \hat{x}(j)\, W_k(j)\, e^{2\pi i m(j-\omega_k)/M_k}, \qquad (1)$$

where $k = 0, \ldots, K-1$, and $m = 0, \ldots, M_k - 1$. This corresponds to windowing $\hat{x}$ by W_k and applying an IFFT of length M_k.

Perfect reconstruction from the transform coefficients is possible because each IFFT has at least length M_k and the union of the filter supports covers the entire frequency axis [1]. Similar to the inverse STFT, $\hat{x}$ is reconstructed by applying an FFT to each frequency slice of its NSGT, followed by overlap-add. Finally, an IFFT is applied to the result to yield x.

Since we do not require the filters W_k to sum up to 1, another set of so-called *dual filters* is necessary to actually achieve perfect reconstruction in the overlap-add process. Explicitly, from [1], the dual filters are given by

$$\widetilde{W_k}(j) = \frac{W_k(j)}{\sum_{l=0}^{K-1} M_l \, |W_l(j)|^2}, \qquad (2)$$

$j = 0, \ldots, L-1$, and the reconstruction amounts to computing the following sum

$$\hat{x}(j) = \sum_{k=0}^{K-1} \sum_{m=0}^{M_k-1} c_{k,m} \widetilde{W_k}(j) e^{-2\pi i m(j-\omega_k)/M_k}, \qquad (3)$$

$j = 0, \ldots, L-1$.

In this contribution, a special type of NSGT, the varying-Q type NSGT (VQ-NSGT) is used. This construction is an extension of the constant-Q NSGT (CQ-NSGT) proposed in [1]. For each octave of the input signal's relevant frequency range, it allows the choice of a desired number of frequency bins and corresponding filters, hence an adapted frequency resolution. This is in contrast to the CQ-NSGT's fixed number of bins. Thus, a simple set of parameters can be used to tailor the frequency resolution in each octave to a particular application.

The filters W_k are constructed such that, for each octave, a fixed Q-factor and the desired number of bins are obtained. If H is a continuous Hann window, centered at 0 and nonzero on $[-1/2, 1/2]$, we take each W_k to be a sampled version of a translated and dilated H. More precisely, we construct the filters as follows:

Let $f_{\min}, f_{\max}$ and f_s denote the minimum and maximum frequencies, and the sampling rate (in Hz), respectively, such that $f_{\max}$ is less than the Nyquist frequency $f_s/2$. Furthermore, let b_n denote the number of bins in the nth octave, $n = 1, \ldots, N_{\text{oct}}$, where N_{oct} is the number of octaves necessary to cover the range from $f_{\min}$ to $f_{\max}$.

We let f_k be the kth center frequency, k running from 1 to the total number of frequency bins. The number of frequency bins per octave varies, so the center frequencies on the nth octave are given by $f_k = f_{B_n+k_n} = 2^{n-1} f_{\min} 2^{\frac{k_n-1}{b_n}}$, for $k_n = 1, \ldots, b_n$, where B_n signifies the total number of bins below the nth octave: $B_1 = 0$, $B_n = \sum_{l=1}^{n-1} b_l$. For each f_k, we set the filter length Ω_k to be $f_{k+1} - f_{k-1}$, resulting in the following Q-factors: $Q_n = f_k/\Omega_k$, with $Q_n = (2^{1/b_n} - 2^{-1/b_n})^{-1}$ for $k = B_n + 2, \ldots, B_{n+1}$. Note that $k = B_n + 1$ serves as a smoothing transition bin between consecutive octaves and the Q-factor is between Q_{n-1} and Q_n. With these parameters, the filters W_k are given by

$$W_k(j) = H((j\tfrac{f_s}{L} - f_k)/\Omega_k),\ j = 0, \ldots, L-1.$$

We note that each filter W_k is centered at $\omega_k = f_k L/f_s$.

Since the union of filter supports has to cover the entire frequency axis, additional filters are considered. These include the zero frequency, two frequencies for each octave between $f_{\max}$ and $f_s/2$, and $f_s/2$ itself. Moreover, due to the symmetry of $\hat{x}$, center frequencies beyond $f_s/2$ are positioned in a symmetric manner. Letting K_{add} be the number of additional center frequencies, the total number of frequency positions is $K = K_{\text{add}} + 2\sum_{n=1}^{N_{\text{oct}}} b_n$.

3. NSGT FOR BEAT TRACKING

The method in which NSGT is applied to beat tracking is summarized in Figure 1. As a first step, an Onset Strength Signal (OSS) is computed from a monaural audio signal. As the way of computing such a function is not the focus of this paper, the OSS proposed by Ellis [6] is used, which is derived from the time derivative of a magnitude spectrum (Spectral Flux). While the input waveform is sampled at a sampling frequency of 44100 Hz, the obtained OSS has a sampling frequency of 160Hz. The second step is taking the

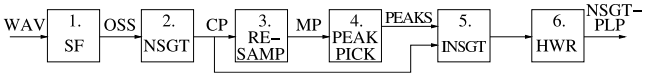


Figure 1: Diagram of the NSGT-PLP computation

NSGT of the OSS in order to get a description of the time-varying periodicities in the signal. This description will be referred to as a (complex) periodogram, CP, throughout the text. Periods in the range from $f_{min} = 0.5Hz$ ($30bpm$) to $f_{max} = 16Hz$ ($960bpm$) are taken into consideration. This range can be divided into $N_{oct} = 5$ periodicity octaves. The NSGT gives us a time-period representation with a varying period resolution, as depicted in an example periodogram magnitude in Figure 2.

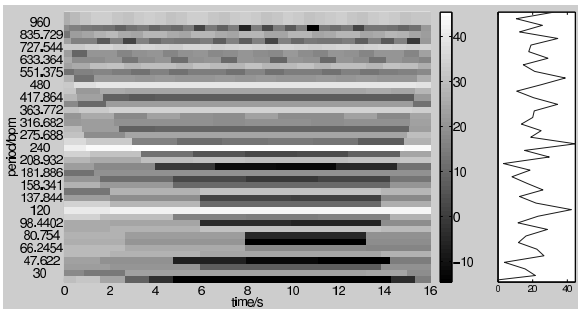


Figure 2: Periodogram magnitude of drum beat at 120bpm (left), and its mean values over time (right). The maximum value of the mean is located at 240 bpm. Throughout the duration of the sample there is a peak at this bin (white line in periodogram). Periodicity bins are non-linearly spaced, as can be seen from the ordinate labels.

When applying the NSGT, we have to take into account the time resolution T_{hop} that is desired in the next computational steps. This resolution determines how often the tempo estimation is updated (*e.g.* every 4s). For a given resolution T_{hop}, we would like to guarantee that the lowest time resolution in the NSGT is still about as large as T_{hop}. For each octave n, T_{hop} and the number of bins in an octave b_n are related by $T_{hop} \leq \frac{Q_n}{f_{min} * 2^{n-1}}$, with $Q_n = (2^{1/b_n} - 2^{-1/b_n})^{-1}$, as detailed in Section 2. For example, given a T_{hop} of four seconds and $f_{min} = 0.5Hz$ (*i.e.* $30bpm$) the number of bins in the five periodicity octaves are $b = \{3, 6, 12, 12, 12\}$, where $b_{max} = 12$ was chosen as the maximum periodicity resolution that we want to obtain. Further increase to larger values of b_{max} was not found to improve experimental results. As can be seen for the varying values in b and as explained in Section 2, typically our transform is a varying-Q transform,

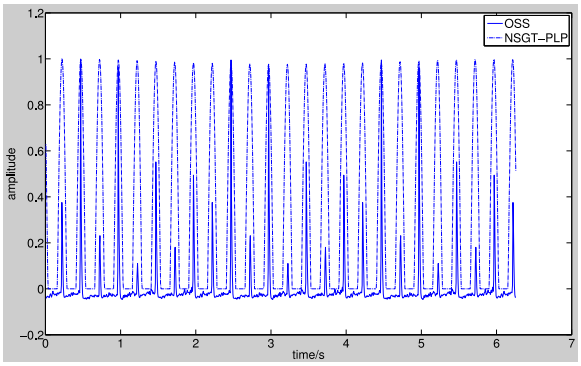


Figure 3: Excerpt of OSS and resulting NSGT-PLP for the drum beat used to create the periodogram in Figure 2. Amplitudes have been normalized to have the same range.

and the number of bins in every octave is adapted to the desired resolution. After fixing the numbers of bins, the set of periodicity domain windows W_k is obtained as described in Section 2, and the transform is computed for the input OSS. It has to be pointed out again that the NSGT is currently applied to the whole OSS, and windowing is performed in the periodicity domain, enabling flexible resolution adjustment.

The output of the NSGT is the complex periodogram CP, with a varying resolution over periodicity. The remaining steps to obtain a simplistic beat tracking are to resample CP to obtain a uniform magnitude periodogram, MP, perform a peak picking in MP, and, using CP, apply the inverse transform on the coefficients related to the detected peaks.

In order to perform peak picking, the non-uniform sampled CP has to be re-sampled (Block 3 in Figure 1) to the uniform sampled magnitude periodogram, MP, with the sampling period T_{hop}. For this, a moving average filter is applied to every periodicity bin of the magnitude of CP, with the length of the filter being T_{hop}. Then the values for MP are obtained by linear interpolation.

In this paper our focus lies on examining the advantages of applying the NSGT instead of an STFT. For that reason the following peak picking was chosen to be as simple as possible. More sophisticated schemes that take account of *e.g.* temporal development in order to get a smoother tempo curve were considered to obscure the comparison of the two transforms. Instead, for each time slice of MP, only the highest peak between 50bpm and 600bpm will be picked.

Obtaining a simple sinusoid that can be used for beat tracking is a straight forward procedure in the case of the NSGT. In the complex periodogram CP all values which are not related to a peak in MP are simply set to zero. Then the inverse transform of this filtered CP is computed.

In order to investigate the improvements that result from the usage of the NSGT, the proposed system will be compared to a similar approach that applies STFT [8]. Therein, the obtained signals are referred to as Predominant Local Pulse functions (PLP), and that approach will be summarized in Section 4. In order to obtain similar waveforms also for the NSGT, a half wave rectification is applied to the obtained inverse transform, which is denoted by HWR in Figure 1. The resulting waveform will be referred to as

NSGT-PLP throughout the following sections. In Figure 3 we zoomed into the drum beat that was used to create the periodogram in Figure 2, and depict the OSS and the resulting NSGT-PLP.

In our experiments for both NSGT-PLP and STFT-PLP the range of considered periodicities was set equally to $f_{min} = 0.5Hz$ ($30bpm$) to $f_{max} = 16Hz$ ($960bpm$), to guarantee for a valid comparison. In our experiments, analysis hop size parameter T_{hop} was changed as well, but no influence on the differences between the two transforms were observed by changing this parameter.

4. EVALUATION METHOD

In order to determine if the application of NSGT leads to an improvement over more classical transform methods, the proposed method is compared to the method introduced by Grosche and Müller [8]. The most significant difference is the application of an STFT instead of the NSGT for obtaining a beat periodogram. System parameters, such as the tempo search range in the peak picking, the simple peak picking method, and the application of half-wave rectification, are the same between the proposed method and the method by Grosche and Müller [8]. As explained before, this will allow us to trace back performance differences to the differences in the transforms, and optimization towards better beat tracking performance, *e.g.* by more sophisticated tempo induction in the periodogram, is left for future work. Differing from [8], we used the same OSS [6] for the STFT based method to guarantee comparability.

Evaluation of beat tracking algorithms has been addressed recently by Davies *et al.* [4]. Several measures have been proposed in the literature, each having a different way to penalize deviations from the beat annotations. As described in the previous sections, the local tempo induction does not consider factors such as tempo continuity over time. Thus, from all the available evaluation measures described by Davies *et al.* [4], we chose two criteria:

1. AML_t: a measure that ranges from 0% to the best value of 100%. It allows for local errors in the beat estimation, and considers a beat sequence with tempo half, equal or double the annotated tempo correct, as long as it is aligned in phase or exactly on the off-beat. Tempo doubling or halving is likely to happen as we are simply picking the highest local peaks between 50bpm and 600bpm, which will include both double and half tempo for most pieces of music.

2. F-measure: a measure that ranges from 0% to the best value of 100% . It was also applied by Grosche and Müller [8], and it penalizes phase errors (at correct tempo) with a value of zero, while it decreases from 100% to 66% in case of tempo halving or doubling. Just like AML_t it does not require continuity of the beat sequence.

Motivated by the evaluation presented by Grosche and Müller [8], these two measures will be computed in two different setups:

1. Peak picking (PP): the PLP functions derived using STFT and NSGT will be interpreted as the output of a beat tracker, and only a peak picking will be performed to obtain the time instances of the beats. To quantify the improvement compared to the OSS a peak picking as described by Bello *et al.* [2] is performed on the OSS. The derived time instances are rated against the ground truth using AML_t and F-measure.

2. Beat tracking (BT): the PLP computation is interpreted as a filtering of the initial OSS, with the goal to obtain a signal concentrated at the beat tempo and phase. This filtering, either performed using STFT or NSGT, is plugged into the original beat tracking code by Ellis [6], and again AML_t and F-measure is determined.

The evaluation data in this paper was first used by Dixon [5]. For our experiments only the data with relatively stable tempo will be used, i.e. no sudden discontinuities are found in the annotations. The obtained data comprise 1220 audio files of many different styles of music. Length of the excerpts varies from $11s$ up to almost 2 minutes. As detailed by Dixon [5], this data set covers a wide variety of musical styles and is currently the largest available dataset for beat tracking evaluations.
The system parameters are set to a kernel length of $4s$ and hop size of $0.2s$ in the STFT-PLP, for the NSGT-PLP T_{hop} was set to $4s$ and the maximum number of bins per octave, b_{max}, is set to 12.

5. RESULTS

Table 1: Mean of the AML_t and F-measure over all 1220 files, values are shown in %.

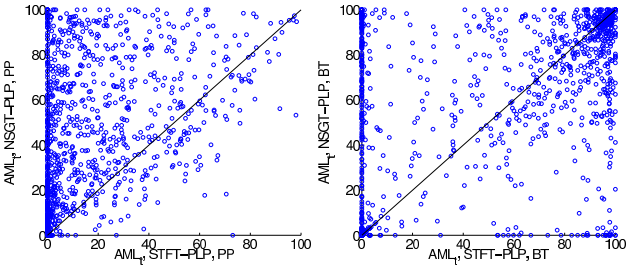
	Peak picking (PP)			Beat tracking (BT)		
	STFT	NSGT	OSS	STFT	NSGT	OSS
AML_t	14.9	**38.8**	0.7	51.8	**60.1**	68.0
F-m	42.9	**49.5**	28.6	46.9	**50.2**	56.0

In Table 1 the mean values over all 1220 files of the two evaluation criteria are given. All differences between those values are statistically significant (ANOVA followed by a series of t-tests with level of significance of $\alpha = .05$, Tukey's HSD adjustment was used to account for the effect of multiple comparisons). From the left three columns it is apparent that F-measure and in particular AML_t penalize the peaks in the original OSS that do not coincide with the beat annotation. The more interesting conclusion from the peak picking experiment is that the NSGT can improve significantly over the STFT, with the difference being larger for AML_t. In the right half of Table 1 the mean values for the original beat tracking system by Ellis [6] are depicted in the right column (OSS). Obviously, the simple sinusoidal waveforms of the STFT- and NSGT-PLP cannot reach these values, as they only contain one (possibly wrong) periodicity, which was chosen from the periodogram peaks. However, comparing the two PLP functions it is apparent also for the BT setup that the performance can be improved by applying the non-stationary Gabor transform, which is emphasized using bold numbers in Table 1.

In order to interpret the reasons for the better performance of the NSGT-PLP we can have a look at the scatter plots shown in Figures 4 and 7. These plots show the performance measures of the NSGT-PLP on the ordinate, and the

measures of the STFT-PLP on the abscissa. Thus, a distribution of points above the diagonal indicates advantages for the NSGT-PLP, and the way points are grouped into clusters can give conclusions about the differences between the methods.

AML_t scatter plots



(a) (b)

Figure 4: Scatter plots of AML_t in the two different setups.

For the AML_t criterion, Figure 4.(a) depicts the scatter plot for the PP setup and Figure 4.(b) for the BT setup. For the PP setup, there is a majority of the data clearly above the diagonal, with many songs that score zero for the STFT-PLP, which can be recognized from the clustering of points on the ordinate. It is revealed that the large difference in performance is related to two reasons:

1. A tendency towards estimation of very high tempi in the STFT-PLP, which is depicted in Figure 5. There, the quotient of the estimated tempo median and the ground truth tempo median is depicted for those songs that improved by 10% or more in the AML_t measure in the case of NSGT-PLP, *i.e.* those songs which are clearly above the diagonal in Figure 4.(a). Figure 5 proves that the high periodicity peaks are stronger emphasized for the STFT, while for the NSGT the double tempo is strongest in most cases. In Figure 4.(a) those samples which score zero using the STFT-PLP (those on the ordinate) are usually characterized by a very stable and high tempo estimate, and for that reason are penalized most using the AML_t, which allows only for tempo halving and doubling.

2. Large variances in the estimated local tempi obtained from the STFT-PLP, an effect that appears only for the samples above the diagonal and not on the ordinate in Figure 4.(a). The variances of those samples are depicted in Figure 6, where it is apparent that for the NSGT-PLP the tempo estimation is very stable, which complies with the stable tempo of the audio data that was used in this experiment. The variances in the STFT-PLP are much higher, which is caused by two facts: the tempo estimation contains spurious large jumps, and it is characterized by small random deviations throughout the files. Increasing the window hop size of the STFT from $0.2s$ up to $2s$ (half overlap) did also not lead to improvement in the error measures depicted in Table 1, and did not solve the described problems.

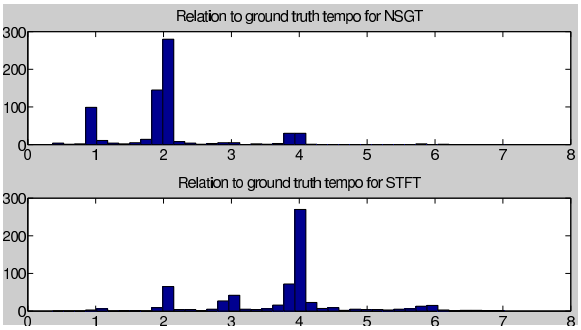


Figure 5: Histograms of relations between estimated tempo and ground truth for AML_t in the PP setup. The subset of songs was used for that NSGT improved performance by 10% or more.

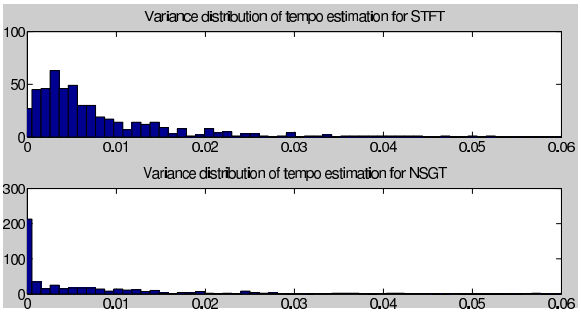
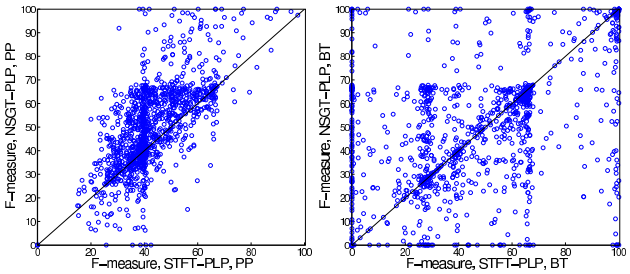


Figure 6: Histograms of variances in estimated tempo for AML_t in the PP setup for the subset of songs above the diagonal but not on the ordinate in Figure 4.(a).

Going from Figure 4.(a) to 4.(b), it can be seen that the improvement for NSGT-PLP gets smaller. This is caused by the fact that the state-of-the-art beat tracker compensates at least to some extent for the variations in the tempo estimates, however the samples with widely overestimated tempi remain on the ordinate of Figure 4.(b).

F-measure scatter plots



(a) (b)

Figure 7: Scatter plots of F-measure in the two different setups.

Looking at the scatter plots for the F-measure in Figures 7.(a) and 7.(b) reveals some more interesting insights into the causes of the improved performance when using the NSGT-PLP. First, it can be seen that in Figure 7.(b) a lot of samples appear on the ordinate, indicating a zero value

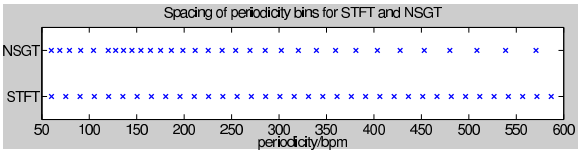


Figure 8: Alignment of periodicity bins in STFT and NSGT.

for the STFT-PLP, while this does not happen for the PP setup. These errors are related to a phase misalignment of the beat sequence estimated by the Ellis beat tracker using the STFT-PLP, because the most likely way to get to a F-measure of exactly zero is using a correct tempo but tapping close to the off-beat. Thus, using the NSGT improves the phase estimation of the beat sequence, which is caused by the straight forward method of including the original phase in the inverse transform.

Probably the most relevant finding that can be obtained from the F-measure scatter plots is related to the rectangle located in the middle of these scatter plots, with corner points at about $(28, 28)$ and $(66, 66)$. An F-measure of 66% is usually related to a tempo halving or doubling, which is not perceived as problematic and is not considered as an error in other measures, such as AML_t. This error happens mostly for the NSGT-PLP, which has been observed previously when analyzing the AML_t scatter plots. A more severe error which appears often for the STFT-PLP results in the clustering of data close to the F-measure of 28%. This error is related to a tempo over-estimation by a factor of 1.33, which results in perceptually unacceptable beat sequences. Analyzing the affected pieces it was observed that this type of error happens almost only for pieces with slow tempo annotation, *i.e.* with tempi slower than $100bpm$. To understand the reason for this phenomenon it is helpful to look at the periodicity domain bin spacing of both STFT and NSGT. In Figure 8, it can be seen that in general the bins are more densely situated for the NSGT in low periodicities, while for the higher ranges the linear spacing of the STFT bins leads to a higher resolution. As explained in Section 3, the distribution of bins in the NSGT is equal to constant-Q in the higher octaves, while the number of bins in lower octaves is limited by the demanded time resolution T_{hop}. Especially in the low periodicity region, periodicity estimations on the STFT are likely to present errors in the order of factor 1.3, while this is avoided in the NSGT by the spacing of the bins. Increasing the analysis window size for the STFT to 8s, which increases the periodicity resolution, showed no improvement in the overall accuracy measures. This indicates that the crucial point is not only the high resolution in the low periodicities, but also the way coefficients are spaced throughout the whole periodicity range. Thus, the combination of avoiding 1.33- and 4-times tempo overestimation by using the NSGT is considered to be an important finding for the improvement of beat tracking systems.

6. CONCLUSIONS

In this paper, the nonstationary Gabor transform was applied in a simple framework to a beat tracking experiment, and the obtained results show significant improvement over the application of a STFT. The reasons of this improvements are related to the more flexible resolution of the transform in the time-frequency plane. This flexibility is not even limited to constant-Q, and comes always with an inverse transform, which can simplify the phase alignment of a beat estimation significantly. Next steps in the work using this transform will be an integration of a more sophisticated period estimation, and experiments with data having changing tempi. On such data, the NSGT could be applied in a setting with varying T_{hop}, by taking shorter analysis steps in presence of unstable periodogram regions, similar to the adaption of window sizes in the vicinity of transient regions.

7. ACKNOWLEDGMENTS

This research was supported by the Vienna Science and Technology Fund (WWTF): MA09-024 (Audiominer), the Austrian Science Fund (FWF):[T384-N13] and [S10602-N13], and the FCT/MCTES (SFRH/BPD/51348/2011).

8. REFERENCES

[1] P. Balazs, M. Dörfler, F. Jaillet, N. Holighaus, and G. A. Velasco. Theory, implementation and applications of nonstationary Gabor frames. *preprint*, 2011.

[2] J. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. Sandler. A tutorial on onset detection in music signals. *IEEE Trans. on Speech and Audio Processing*, 13(5):1035–1047, Sept. 2005.

[3] J. A. Bilmes. Timing is of the essence. Master's thesis, Massachusetts Institute Of Technology, 1993.

[4] M. Davies, N. Degara, and M. Plumbley. Evaluation methods for musical audio beat tracking algorithms. Technical report, QMUL, C4DM, 2009.

[5] S. Dixon. Evaluation of the audio beat tracking system BeatRoot. *Journal of New Music Research*, 36(1):39–50, 2007.

[6] D. P. W. Ellis. Beat tracking by dynamic programming. *Journal of New Music Research*, 36(1):51–60, 2007.

[7] G. Evangelista. Flexible Wavelets for Music Signal Processing. *Journal of New Music Research*, 30(1):13–22, Mar. 2001. special issue on Music and Mathematics.

[8] P. Grosche and M. Müller. Extracting predominant local pulse information from music recordings. *IEEE Trans. on Audio, Speech, and Language Processing (in press)*, 2011.

[9] A. P. Klapuri, A. J. Eronen, and J. T. Astola. Analysis of the meter of acoustic musical signals. *IEEE Trans. on Audio, Speech, and Language Processing*, 14(1):342–355, 2006.

[10] G. Peeters. Beat-tracking using a probabilistic framework and linear discriminant analysis. In *Proc. DAFx-09*, 2009.

[11] N. Scaringella and G. Zoia. A real-time beat tracker for unrestricted audio signals. In *Proc. of SMC'04*, 2004.

[12] C. Schorkhuber and A. Klapuri. Constant-q transform toolbox for music processing. In *Proc. of SMC'2010*, 2010.