AUTOMATIC BEAT-SYNCHRONOUS GENERATION OF MUSIC LEAD SHEETS

DURRIEU Jean-Louis, durrieu@enst.fr

TELECOM ParisTech - TSI / LTCI
46 rue Barrault, F-75634 Paris Cedex 13, France

WEIL Jan, weil@nue.tu-berlin.de

TUB - Communication Systems Group
Einsteinufer 17, 10587 Berlin, Germany

ABSTRACT

Most of the popular music scores are written in a specific format, the lead sheet format. It sums up a song by representing the notes of the main melody, along with the chord sequence together with other cues such as style, tempo and time signature. This sort of representation is very common in jazz and pop music, where the accompaniment playing the chord sequence usually is improvised. The aim of our study is to bring together two techniques, a chord detection system and a lead melody transcriber, in order to produce a lead sheet. In addition to the respective issues inherent to each problem, we also need to address tempo estimation, time signature estimation, and, based on these estimations, time quantification of both the chord sequence and the melody line. We propose a tempo tracker that aligns the beats to the audio, and adapt the chord detection and melody extraction systems so as to take into account this new piece of information. Future works include cover song detection based on lead sheet representation, query-by-similarity applications and so on.

1. INTRODUCTION

The lead-sheet format in music is well-known among jazz and rock players. It consists of the main melody, along with the chords of the accompaniment. It can also include more information such as the style of the song, the lyrics, the structure, etc.

Here we are interested in combining two existing systems: a chord detection algorithm and a melody extractor, in order to obtain such a representation. We however missed the temporal information such as tempi and time signature. Tempo estimation is a well studied problem and we base our system on some previous works [1]. We also designed a method that aligns the beat to the data. The time signature estimation still is an open problem, for which we propose some general directions.

Some improvements can also come from the fusion of the results of the different algorithms. We propose some of such improvements, but we expect that further studies could unravel even more of those correlated results.

This document is organized as follows: first, we present the proposed beat tracking and pulse alignment algorithms. Then we explain the chord and melody estimation modules. Thereafter, a short evaluation of these tasks is proposed. A short insight in what can be done for time signature is also presented before the conclusion. At last, we conclude with some futur works and perspectives.

2. BEAT TRACKING MODULE

In order to produce musically relevant lead-sheets, we need to determine the temporal structure of the song, i.e. the tempo, dealt in this section, and the time signature, dealt in section 5. In order to do so, the tempo is first estimated on 10s-long frames, with a .5s hopsize. This estimation is based on a detection function proposed in [1]. From this function, at each window an auto-correlation function (ACF) is computed, which gives us a "ACF-map". A viterbi algorithm allows us to find the optimal tempo-path, with a trade-off between the tempo variation smoothness and the maxima of the ACF-map. We also output an estimate of the "tatum", which supposedly is the smallest time-unit of the song, in use for the melody quantification part.

We tackle the beat/tatum location problem using a dynamic programming approach. We use the same hopsize for the windows as previously, but their length is at least 10 beats/tatums per window - i.e. 10 times the maximum time lag between two beats. For each window, an impulse comb is generated with a period corresponding to the estimated tempo. The cross-correlation between the comb and the data in the window is stored in a matrix, the maxima of which give the time lags or "phases" necessary to align the combs to the data. In order to avoid off-beat problems, which are common in rock and jazz music, we designed a Viterbi algorithm that smoothes the variability of location of the pulses. Instead of smoothing the path "horizontally" in this phase-matrix, it takes into account the tempo changes and favors phase locations where they are expected to occur according to the previous window.

At last, for each window, we place the pulses according to the estimated phase, taking care of possible "double" pulses by choosing a location between them where the onset detection is at a local maximum.

This work was partly supported by the European Commission under contract *FP6-027026-K-SPACE*.

3. DETECTION MODULES

3.1. Chord sequence detection

The chord detection method we developped is close to the system introduced in [3]: the chosen features are the tonal centroids, derived from the chroma vectors. A Hidden Markov Model (HMM) is assumed for the chord sequence. As in [3], we assume the transition probabilities to depend only on the interval between the chords. Further studies should aim at using key-specific HMMs, in order to estimate the main key at the same time.

In order to integrate beat information, either we compute the features within segments obtained thanks to the beat location given in section 2 or we constrain the Viterbi decoding of the chord sequence: within each segment, the state (i.e. the chord) is assumed constant. However, neither of these two solutions gave better results for now.

3.2. Main melody extraction

The main melody transcription module is based on the leading melody estimation of [2]. A source-filter model catches F_0 candidates for each frame, and the main melody is computed thanks to a Viterbi smoothing algorithm, accomplishing a trade-off between the energy and the frequency proximity of consecutive candidates. This system only outputs a frame-wise sequence of frequencies in Hz. In order to obtain the desired sequence of temporally quantified notes (i.e. on the Western music scale), we use the tatum estimation of section 2. It provides segments on which we can decide which note was intended. Most of pop music singers do not have strong vibrato, which makes this task rather straight-forward in those cases: a simple decision like taking the mean or the median of the output frequency sequence within each segment will give satisfying results. More studies on vibrato estimation may be useful in order to deal with classical music.

As pointed out in section 3.1, the algorithm can also separate the singer voice from the background music. This output can then be used as a pre-processing step to other tasks such as chord detection or multi-F_0 estimation.

4. EVALUATION

The evalution of such a transcription system as a whole is not clear yet. However, we can evaluate the different modules separately.

The Chord detection was tested on a database of beatles songs along with MIDI synthetized ones. The recognition recall vary from 65% to 70%.

As for the main melody extractor, as was stated in [2], performs amongst the state-of-the-art systems, with 78% frame-wise recall on the pitched frames. One of the main drawback of this module for now is the lack of silence detection in vocal activity. As such, we observe a significant drop in the results when taking into account non-vocal frames with 65% of global recall. It also leads to spurious notes in the transcription. In order to avoid these, some heuristics can be applied, e.g. penalizing segments in which the melody is varying to deeply.

5. ABOUT TIME SIGNATURE

As we discussed in the previous sections, we also need an estimate of the time signature of the song. This signature is a fraction: the denominator gives the musical unit related to the beat, while the numerator tells how many of these units there are in 1 measure.

We propose the following direction for future works on the topic: there usually are two trends for choosing the denominator. The song either has binary rhythmic patterns or ternary ones. In the first case, one usually can assume the unit to be the eighth note, with symbol 4, in the other case, it is often chosen as the sixteenth note, symbol 8. As a first approximation, one can assign either of these two denominators. The tatum to beat ratio may give some insight as to which of them it should be.

Assuming that the chord changes mainly occur on the beats, and more specifically on the up-beats, at the beginning of measures, the numerator could be infered from the harmonic structure. More evidence is needed for this last assumption, but this should give a rather straight-forward way of estimating the time signature of the analyzed song.

6. CONCLUSIONS

In this study, we have found that each system can take advantage of the beat/tatum estimation, especially on the quantification step. This seems to produce musically relevant material. The result is not yet completely ready and we still need to estimate the time-signature. This feature is closely related to the beat and tatum ratio, but also, we believe, to the melodic and harmonic structure. Further studies aim at designing a robust way of estimating the time-signature as well as the overall structure of the musical piece, which would for example help avoiding repetitions in the output lead-sheet.

7. REFERENCES

[1] M. Alonso, G. Richard, and B. David. Extracting Note Onsets from Musical Recordings. *ICME*, 2005.

[2] J.-L. Durrieu, G. Richard, and B. David. Singer melody extraction in polyphonic signals using source separation methods. *ICASSP*, 2008.

[3] K. Lee and M. Slaney. Acoustic Chord Transcription and Key Extraction From Audio Using Key-Dependent HMMs Trained on Synthesized Audio. *IEEE Trans. on ASLP*, 2008.