The FlexQ algorithm

Klaus Frieler

The FlexQ algorithm uses an optimization approach to find the optimal subdivision for all onsets between two beats. This is achieved by combining certain heuristic principles to construct a suitable loss function for which a subdivision with the least loss can be found by a simple grid search over all admissible subdivisions.

Formally, we consider M onsets $\{t_i\}_{1 \le i \le M}$ between two consecutive beats b_L and b_R, i. e., $b_L \le t_i \le b_R$ for all i. Without loss of generality, we can rescale all involved onsets by $\frac{t-b_L}{b_R-b_L}$ so that $b_L = 0$ and $b_R = 1$. A tatum K-grid G_K for a subdivision K is then the set of points

$$G_K = \left\{\frac{m}{K}\right\}_{0 \le m < K} = \left\{0, \frac{1}{K}, \frac{2}{K}, \ldots, \frac{K-1}{K}\right\}.$$

A quantization of the onsets t_i with respect to the grid G_K is defined by the following prescription, if $M \le K$, else the empty set. For each t_i, the closest grid point is the index

$$m_i = \min_m \arg\min_m \left| t_i - \frac{m}{K} \right|.$$

For the M onsets under consideration we thus obtain a set of closest grid indices. Requiring strict monophony entails that duplicated indices are not allowed. Hence, starting from the leftmost onset, we move all indices successively to the next free position if a duplicated index is found. This might result in new duplicated indices and the process is repeated till all indices occur only once, which might, however, not always be possible. The resulting quantization of the onsets is then either the set of causal grid points, denoted $\{m_i^*\}$, or the empty set.

This might be best illustrated by an example. Set $K = 2$, with grid $\{0, 1/2\}$, and onsets $\{3/8, 5/8\}$. The closest grids points for these onsets are then $m_1 = 1$ and $m_2 = 1$. Since the indices are the same, the second index should be moved, but there is no grid point left to the right, hence the quantization is the empty set. For the 4-grid $\{0, 1/4, 1/2, 3/4\}$, the closest grid points are $m_1 = 1$ and $m_2 = 2$, so no problem arises in this case. But for onsets $\{7/16, 9/16\}$, we have $m_1 = 2$ and $m_2 = 2$, which can be resolved by shifting m_2 one unit to the right with a resulting quantization of $m_1 = 2$ and $m_2 = 3$.

For a non-empty quantization, we can then define the quantization error as the sum of absolute differences between onsets and their modified closest grid points:

$$\Delta q = \sum_i \delta q_i = \sum_i \left| t_i - \frac{m_i^*}{K} \right|.$$

Now we have nearly all the elements for defining the FlexQ algorithm. The last component is the standard deviation of quantization errors:

$$s_q = \sqrt{\frac{1}{N-1} \sum_i \left(\delta q_i - \frac{1}{N} \Delta q \right)^2}$$

The loss function will be built from four preference rules.

1. Prefer smaller subdivisions.
2. Prefer binary and ternary subdivisions.
3. Prefer smaller quantization errors (deviations from the ideal grid points).
4. Prefer homogeneous deviations.

We can now state the loss function:

$$L(K) = \alpha_1 K + \alpha_2 \Omega(K) + \alpha_3 \Delta q + \alpha_4 s_q,$$

where

$$\Omega(K) = \begin{cases} 1, & \text{if K odd and } K > 3, \\ 0.5, & \text{if } K = 1 \text{ or } K = 3, \\ 0, & \text{otherwise.} \end{cases}$$

The function Ω embodies the second preference ("prefer binary and ternary subdivisions") only approximately, but in practice the differences are only marginal because very large grid sizes are not considered. The only practical differences might arise for the (actually) rare case of $K = 9$, which is a ternary subdivision but penalized by Ω. The $\alpha_{1,2,3,4}$ are free parameters.

An optimal grid K for a given set of M onsets $\{t_i\}$ between two beats b_L, b_R can then be found as

$$K_{\text{opt}} = \underset{M \le K \le K_{\max}}{\arg\min} \; L(K),$$

where $K_{\max}$ is defined via

$$\frac{b_R - b_L}{K_{\max}} < \alpha_5,$$

with free parameter α_5, which defines the smallest absolute distance between grid points. This parameter should be set in the order of 30 ms to 50 ms which corresponds to the discrimination threshold of two events and to the fastest observed human movement of about 20 Hz $\sim$ 50 ms. For $\alpha_5 = 40$ ms and IBI $b_R - b_L = 1$ s ($\sim$ 60 bpm), an upper bound of $K_{\max} = 25$ can be found.

From the optimal grid size, the optimal tatum positions are then given by the modified closest grid points m_i^*.

The default parameters for $\alpha_1, \ldots, \alpha_5$ were found using manual experimentation. In the current implementation, these are (MeloSpySuite/GUI parameter names in parentheses):

$$\begin{aligned}
\alpha_1 &= 1.0 \quad \text{(mismatchPenalty)} \\
\alpha_2 &= 1.0 \quad \text{(oddDivisionPenalty)} \\
\alpha_3 &= 8.0 \quad \text{(distPenalty)} \\
\alpha_4 &= 10.0 \quad \text{(spreadPenalty)} \\
\alpha_5 &= 0.02 \quad \text{(rhythmThreshold)}.
\end{aligned}$$

.