THE USE OF RECURSIVE MEDIAN FILTERS FOR ESTABLISHING THE TONAL CONTEXT IN MUSIC

Ilya Shmulevich and Edward J. Coyle

School of Electrical and Computer Engineering
Purdue University
West Lafayette, IN 47906

ABSTRACT

The recursive median filter is used to improve the structure of the output of the key finding algorithm for establishing tonal contexts of musical patterns in a musical composition. This is subsequently incorporated into a system for recognition of musical patterns. Krumhansl's key-finding algorithm is used as a basis. The sequence of maximum correlations that it outputs is smoothed with a cubic spline and is used to determine weights for perceptual and absolute pitch errors. Maximum correlations are used to create the assigned key sequence, which is processed by a recursive median filter. In most cases, the recursive median filter establishes the key more accurately than the standard median filter. Additionally, since the recursive median is idempotent, the key-finding output is guaranteed to be a root signal.

INTRODUCTION

Musical Pattern Recognition (MPR) is concerned with recognition or classification of musical patterns. Figure 2 shows a block diagram of the MPR system. It produces an error comprised of perceptual and absolute pitch and rhythm information between the *target* and *scanned* musical patterns. Thus, we view this process in a pattern recognition framework in the sense that we try to find the minimum distance from the set of scanned patterns to the target pattern. For a sequence $\left[q_1, q_2, \cdots, q_n\right]$ of *n* notes, we define a **difference of pitch vector** $\vec{p} = \left[p_1, p_2, \cdots, p_{n-1}\right]$, where $p_i = q_{i+1} - q_i$, as the encoding of this sequence. The example used here is J. S. Bach's Invention #8 in F Major (right hand only) and the target pattern is the one shown in Figure 1. For this pattern, $\vec{p} = \left[7,\text{-}8,8,\text{-}5,4,1\right]$. Because of perceptual invariance under transposition of pitch, the user is relieved of knowing the key of the actual musical pattern under this encoding.

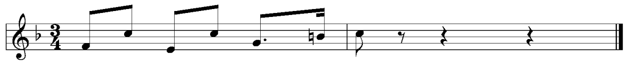

Figure 1 - Sample Target Pattern

The absolute pitch error is defined as $e_a = \left\|\vec{p} - \vec{p}_0\right\|_1$, where $\vec{p}$ and $\vec{p}_0$ represent the difference of pitch vectors of the scanned and target patterns respectively. The pitch error $e_q = \lambda \cdot e_p + (1-\lambda) \cdot e_a$ is a weighted combination of perceptual and absolute pitch errors, denoted by e_p and e_a in the block diagram. Below, we define the perceptual pitch error e_p and discuss the procedure for determining the weight λ.

PERCEPTUAL PITCH ERROR

All intervals of equal size are not perceived as being equal when the tones are heard in tonal contexts [2]. For example, the notes B C played in succession heard in the context of C Major (for instance, after hearing a strong key-defining sequence of notes) would be perceived as being more natural and stable than the same two notes heard in the context of D Major.

Since the ultimate goal is to recognize a target pattern memorized (possibly incorrectly) by a human being, it is important to consider certain principles of melody memorization and recall. For example, findings showed that "less stable elements tended to be poorly remembered and frequently confused with more stable elements." Also, when an unstable element was introduced into a tonal sequence, "... the unstable element was itself poorly remembered" [3]. So, the occurrence of an unstable interval within a given tonal context (e.g. a melody ending in the tones C C♯ in the C major context) should be penalized more than a stable interval (e.g. B C in the C major context) since the unstable interval is less likely to have been memorized by the human user. These perceptual phenomena must be quantified for them to be useful in the classification of musical patterns. Such a quantification is provided by the **relatedness ratings** found by Krumhansl [3]. Essentially, a relatedness rating between tone q_1 and tone q_2 ($q_1 \neq q_2$) is a measure of how well q_2 follows q_1 in a given tonal context. The relatedness rating is a real number between 1 and 7 and is determined by experiments with musically trained listeners. Results are provided for both major and minor contexts. So, a relatedness rating between two different tones in any of 24 possible tonal contexts can be found due to invariance under transposition.

The relatedness ratings described above are not defined when $q_1 = q_2$. However, it is very common to see two successive equal tones in music and we need to be able to assign a rating for them within the given tonal context. A study by Krumhansl and Kessler (1982) provides the answer. In it, the authors provide

probe tone ratings for each of the 12 tones after a strong key-defining context is established. These ratings are also numbers from 1 to 7 and signify how well each probe tone “fits into” the context in a musical sense [4]. The experimentally measured probe tone ratings correlate quite strongly with the distribution of tones in tonal-harmonic music [3, p. 77]. The probe tone ratings expose an ordering of stability on the set of tones and this ordering is referred to as a **tonal hierarchy**. Tonal contexts with similar tonal hierarchies are said to be close to one another. The probe tone ratings will be used as a relatedness rating when two consecutive tones are identical. For example, in the key of C major, the relatedness rating of C-C is equal to 6.35, which is just the probe tone rating of C in that context. Thus, we create new modified matrices of relatedness ratings whose diagonals are the probe tone ratings. We define vectors

$$\vec{\alpha} = \left[\alpha_1, \alpha_2, \cdots, \alpha_{n-1}\right] \text{ and } \vec{\beta} = \left[\beta_1, \beta_2, \cdots, \beta_{n-1}\right]$$

to be the vectors of relatedness ratings for the scanned and target patterns respectively, in the same tonal context. Finally, the perceptual pitch error is defined as $e_p = \left\|\vec{\alpha} - \vec{\beta}\right\|_1$.

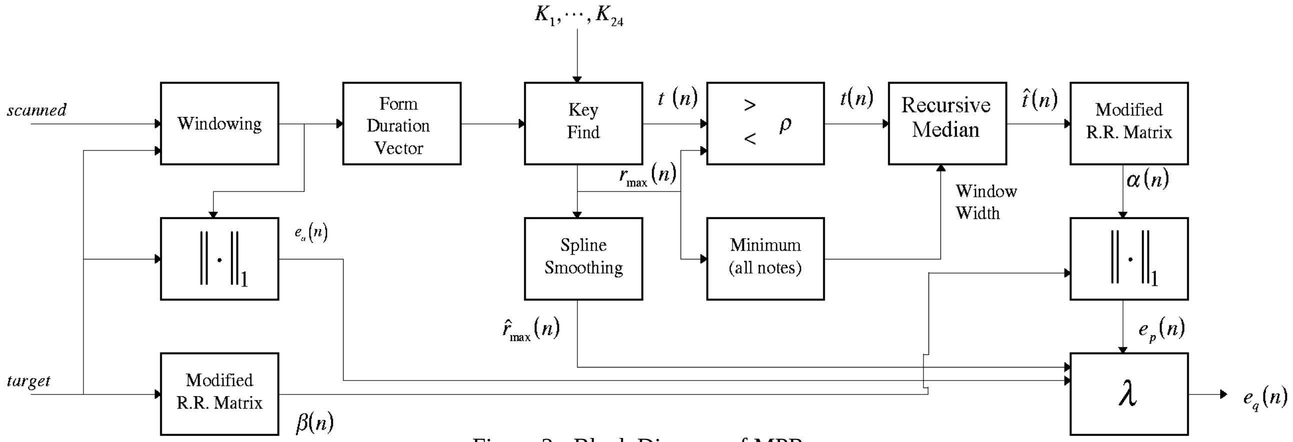

Figure 2 - Block Diagram of MPR

ESTABLISHING THE TONAL CONTEXT

What exactly is the meaning of the tonal context of a pattern? If the pattern is of short length (1 note, for example), then speaking about its tonal context is meaningless. Similarly, if the pattern is very long, it may consist of several tonal contexts and the transitions between them are called modulations. Finally, quite often, a tonal context is a matter of degree in that for a given pattern, there are several possible candidates for tonal context. So, just because the key-signature of a given composition happens to be F major, for example, it does not imply that the relatedness rating vectors $\vec{\alpha}$ and $\vec{\beta}$ must be chosen for that particular tonal context, since modulations and shifting tonal centers are likely to occur.

We thus see the need for a key-finding algorithm which will present us with a most likely tonal context for a given musical pattern and this tonal context will be subsequently used for the relatedness rating vectors. Such an algorithm was developed by Krumhansl [3] and is based on the fact that “most stable pitch classes should occur most often” [5]. That is, tones that are sounded most frequently are the ones with high probe tone ratings, in a given tonal context (e.g. in a C major context, C, G, and E occur most often). We now make certain modifications to the algorithm in [3] and present a method for determining the parameter λ.

The set of the 12 probe tone ratings for a given key is referred to as the **probe tone profile** for that key. There are 24 such profiles (12 major and 12 minor). The $i^{\underline{\text{th}}}$ profile vector is denoted by $\mathbf{k}_i$. The input to the algorithm is a 12-element vector **i** whose elements are total durations of the 12 tones in the musical pattern being scanned. Proceeding on the relationship between the number of tone occurrences and probe tone ratings, we correlate vector **i** with each of the 24 probe tone profile vectors and produce a 24-element vector of correlations, $\mathbf{r} = \left[r_1, \cdots, r_{24}\right]$. The highest correlation, $r_{\max}$, is the one that corresponds to the most likely tonal context of the musical pattern being scanned.

Suppose a musical composition that we wish to scan for the purpose of recognizing the target pattern consists of *m* notes and the target pattern itself consists of *n* notes (typically, $m >> n$). We slide a window of length *n* across the sequence of *m* notes and for each window position, the key-finding algorithm outputs a key assignment. Thus, we generate a sequence $\mathbf{t} = \left[t_1, t_2, \cdots, t_{m-n+1}\right]$ of key assignments .

We can see that each t_i is a number between 1 and 24 (on a computer, it is convenient to use 0 to 23). See Figure 3 for an assigned key sequence.

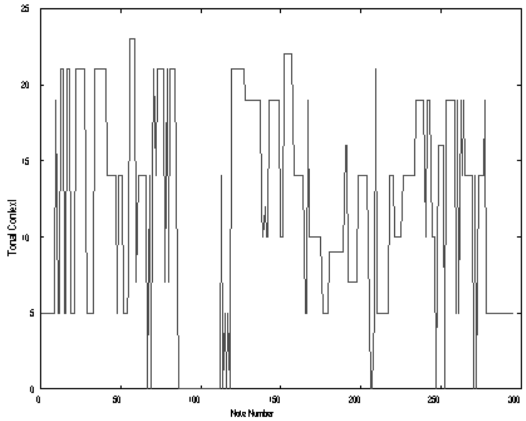

Figure 3 - Assigned Key Sequence

As can be seen from that figure, there is quite a bit of variation in certain regions of the sequence. Moreover, we can see some impulses lasting only one note, which would seem to indicate that the tonal context changes for one note and then changes back - a very unlikely circumstance. This exposes a weakness of the key-finding algorithm [3] in that it may be sensitive to window length as well as the distribution of pitches within the window. Whatever the tonal context may be, it makes little sense to talk of two modulations occurring one note apart. Besides this, there are small areas of oscillations, especially those close to edges between two flat regions. These edges signify modulations and as the window slides across them, the key-finding algorithm is unable to determine a prevalent tonal context due to the presence of pitches that have high probe tone ratings in two different profiles. As a result, the assigned key values oscillate until a prevalent tonal context is established. Such small oscillations and impulses are undesirable, not only because they do not reflect our notions of modulations, but primarily because they affect the relatedness rating vectors, which inherently depend on the tonal context produced by the key-finding algorithm. Since the values of the assigned key sequence often appears arbitrary in the regions of oscillation, the perceptual pitch error is distorted in these regions.

As a solution to the above problem, we employ the recursive median filter [6] with a large enough window to remove not only the impulses, but also the small regions of oscillations. The output of the recursive median filter is defined as

$$y_i = med\left(y_{i-v,\cdots,}y_{i-1}, x_i, \cdots, x_{i+v}\right)$$

where the samples $y_{i-v}, \cdots, y_{i-1}$ have already been computed during previous positions of the sliding window. It has been shown that the recursive median filter has a higher immunity to impulsive noise than the standard median filter. This makes it a better choice for our purpose than the standard median filter [7]. Moreover, the output of the recursive median is more correlated than the output of the standard median. This is due to the fact that y_i is dependent on previous output values. This correlation in the output is advantageous since the tonal context at a particular position is more strongly dependent on previous values of the assigned key sequence than on future values. Finally, it is well known that the recursive median filter is idempotent. This property implies that any signal is reduced to a root signal after one pass; i.e., it is invariant to further passes of the same filter. This assures us that the assigned key sequence cannot be improved by more filter passes. The window width of the recursive median filter needs to be chosen. If we are to employ the recursive median filter in order to remove oscillations in the regions of modulation, we must establish a high measure of **tonal structure** prior to and after the region of modulation. The amount of notes necessary to establish this, of course, depends on key membership of the notes as well as their relationship to the tonal center (i.e. stability). However, it has been shown that the maximum correlation, $r_{\max}$, is strongly correlated with the degree of tonal structure [5]. Therefore, if $r_{\max}$ is small, indicating a low degree of tonal structure, we should expect to use more notes to establish the latter. This implies that the window width of the recursive median filter should be inversely related to $r_{\max}$. Recall that for every window position of the key-finding algorithm, we have a maximum correlation, thus giving rise to the sequence $r_{\max}(i)$ of maximum correlations. We would like the window width, W, of the recursive median filter to be a function of the lowest of the maximum correlations. That is,

$$W = f\left(\min\left[r_{\max}(i)\right]\right)$$

and one possible function is

$$f(r) = \left\lceil \frac{k}{r^{1/v}} \right\rceil$$

where $\lceil \cdot \rceil$ is the next **odd** integer. Experiments show that values of k = 17 and ν = 8 give good results for the recursive median filter. The parameter ν simply controls the rate of growth of the window width with respect to the lowest maximum correlation.

Figure 4 shows an assigned key sequence processed by a window width 19 recursive median filter, where $\min\left[r_{\max}(i)\right] = 0.46$. As can be seen, the impulses and oscillations are completely removed and yet the key assignments (and modulations) reflect what would be expected upon a visual inspection of the composition.

WEIGHTING OF PERCEPTUAL ERRORS

Now that we can successfully generate the assigned key sequence **t**, all that remains is the determination of parameter λ. In the key-finding algorithm, the higher the maximum correlation, the more reliable is the key assignment. Consequently, the relatedness rating vectors $\vec{\alpha}$ and $\vec{\beta}$ become more suitable for use in the perceptual pitch error and hence, the perceptual error should get more weight. Figure 5 shows $r_{\max}(i)$, the sequence of maximum correlations generated by the key-finding algorithm.

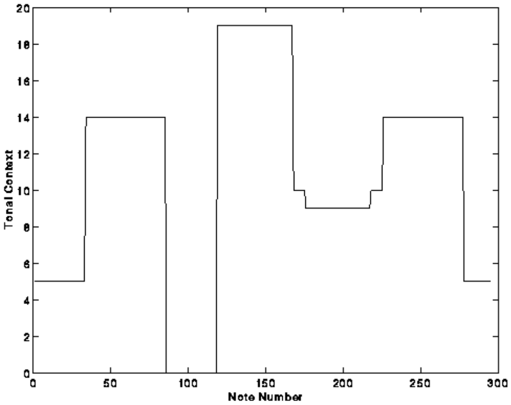

Figure 4 - Recursive Median Filtered Assigned Key Sequence (WW = 19)

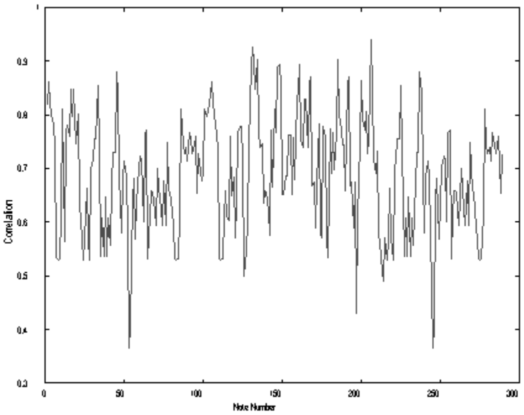

Figure 5 - Sequence of Maximum Correlations

Because of the sensitivity of the key-finding algorithm to pitch distribution within the window, it is common to see small-scale oscillations in the sequence $r_{\max}(i)$. See, for example, the region around note 37 in Figure 5. Such oscillations are undesirable since they tend to imply that the confidence of the key-finding algorithm can change from high to low or vice versa in a matter of one note. Also, the perceptual pitch error becomes overly sporadic. The downward and upward trends, however, such as the one around note 52, should be preserved, since they indicate a genuine decrease of confidence of the key-finding algorithm in a given region. Because of these considerations, median filtering the sequence $r_{\max}(i)$ is not appropriate as it would remove such trends, considering them to be impulses. So, in favor of reducing the jitter of the sequence of perceptual pitch errors $e_p(i)$, preserving local trends, and removing small-scale oscillations, we apply a cubic smoothing spline to the sequence $r_{\max}(i)$, thus creating a new sequence $\hat{r}_{\max}(i)$. It is precisely the values of $\hat{r}_{\max}(i)$ that determine our parameter λ at every position of the window, giving rise to the sequence λ(i) of the parameter values. Figure 6 shows the smoothed maximum correlation sequence. We choose to restrict the range of the sequence λ(i) to

$$a \le \lambda(i) \le b, \quad 0 < a < b < 1$$

for the following reasons. λ(i) should never be allowed to reach a value of 1, since that would effectively ignore the absolute pitch error and put all the weight on the perceptual pitch error. Similarly, λ(i) should not reach a value of 0 because the assigned key sequence **t** has been median filtered, thus making the relatedness rating vectors $\vec{\alpha}$ and $\vec{\beta}$ suitable for use in the perceptual pitch error. Values of a = 0.25 and b = 0.5 have shown to be successful. Our choice of b in effect does not allow the perceptual error to "outweigh" the absolute error. To this end, we restrict the range of the parameter λ by

$$\lambda(i) = m \cdot \left(\hat{r}_{\max}(i) - \max\left(\hat{r}_{\max}(i)\right)\right) + b$$

where

$$m = \frac{b-a}{\max\left(\hat{r}_{\max}(i)\right) - \min\left(\hat{r}_{\max}(i)\right)}$$

making λ(i) just a scaled version $\hat{r}_{\max}(i)$.

CONCLUSIONS

We develop a method for establishing the tonal context of a musical pattern. This is a crucial element in the framework of musical pattern recognition. The recursive median filter is a good choice for improving the structure of the assigned key sequence. However, it would be worth exploring the possibility of using a variable-width filter, where the window width depends on the maximum correlation output by the key-finding algorithm. Moreover, a weighted median filter with weights determined by maximum correlations may be another solution.

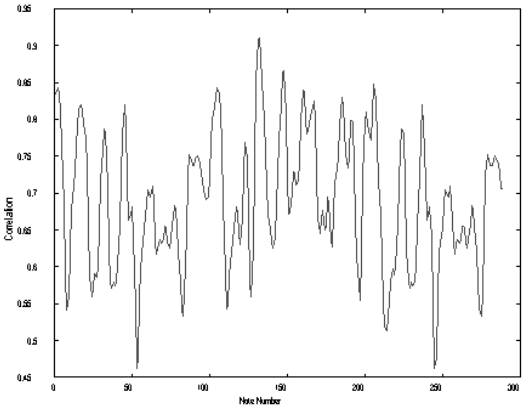

Figure 6 - Spline Smoothed Maximum Correlation Sequence

REFERENCES

[1] I. Shmulevich, E. J. Coyle, "Musical Pattern Recognition," *in preparation*.

[2] C. L. Krumhansl, R. N. Shepard, "Quantification of the hierarchy of tonal functions within a diatonic context," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 5, pp. 579-594, 1979.

[3] C. L. Krumhansl, *Cognitive Foundations of Musical Pitch*, New York: Oxford University Press, 1990.

[4] C. L. Krumhansl, E. J. Kessler, "Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys," *Psychological Review*, vol. 89, pp. 334-368, 1982.

[5] A. H. Takeuchi, "Maximum key-profile correlation (MKC) as a measure of tonal structure in music," *Perception & Psychophysics*, vol. 56, pp. 335-346, 1994.

[6] T. A. Nodes, N.C. Gallagher, "Median filters: some modifications and their properties,'' *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-30, no. 5, pp. 739-746, 1982.

[7] G. R. Arce, N. C. Gallagher, T. A. Nodes, "Median filters: theory for one- and two-dimensional filters," in *Advances in Computer Vision and Image Processing*, T. S. Huang editor, JAI Press, 1986.