Sound Signature Detection by Probability Density Function of Normalized Amplitudes

Ion Bica[1], Zhichun Zhai[2], Rui Hu[3] and Mickey H. Melnyk[4]

Department of Mathematics and Statistics, MacEwan University, Canada

[1]bicai@macewan.ca [2]zhaiz2@macewan.ca [3]hur3@macewan.ca

[4]Piano, Edmonton, Canada; mickeyharold@yahoo.com

Abstract

In this paper, we propose to use the probability density function of normalized amplitudes (PDFNA) to detect distinctive sounds in classical music. Based on data sets generated by waveform audio files (WAV files), we use the kernel method to estimate the probability density function. The confidence interval of the kernel density estimator is also given. In order to illustrate our method, we used the audio data collected from recordings of three composers; Johann Sebastian Bach (1686-1750), Ludwig van Beethoven (1770-1827) and Franz Schubert (1797-1828).

Disclaimer: This paper addresses only to the genre of classical music, and it focuses on instrumental pieces played only on solo instruments. The results obtained are not intended to explain other musical genres.

1 Motivation

In the late 1970s, the Digital Revolution brought a new dimension to sound recording. The first commercial digital recorder was Sony PCM-1 in 1977 [7]. With this new technology, the sound could be stored in large audio files. In this paper we analyze WAV audio files to study the concept of distinctive sound for a classical music composer, i.e. we propose a method to decide how uniform or variable a composer's style is. In the present work, we are proposing a measure, applied to a few instrumental pieces, that give some preliminary and suggestive results, but not enough to draw definite conclusions; this work is the first attempt in this regard, and the goal for future work is recognizing patterns in machine learning.

When we listen to the sound created by a musical instrument our ears perceive the change in the air pressure around them; by higher variations, we understand louder sounds, and by faster variations (higher sound frequency) we understand higher pitch of the sound. The higher and faster variations (we will call it *sound variation*) create the sound experience that our ears absorb and send to the brain for processing. The brain is the central engine that is capable of analyzing the sound received, and based on its analysis it will classify the sound as loud, quiet, high pitch, low pitch, key signature, and so on. At the same time we believe that the brain, trained enough, will be able sometimes to single out the composer of an instrumental piece. If the sound variation falls within a certain bandwidth that the brain classifies as unique for a composer, we believe that the brain can recognize the composer. Then we can logically understand that the sound created by musical instruments are intrinsically correlated with the sound the composers meant to create in their compositions. In the present work, we are not interested in reoccurring motifs that some composers are instantly recognized by, such as the first two bars of Beethoven's Symphony No. 5 in C minor op. 67, or Bach's first bar from his Toccata and Fugue in D minor BWV 565.

Our present work focuses on compositions by three masters of classical music, with mathematical analysis performed with the help of MAPLE and MATLAB. The three composers are Johann Sebastian Bach, Ludwig van Beethoven, and Franz Schubert. We picked these composers because two of them, Bach

and Beethoven, are known to have distinct sound signatures (their composition style is uniform) [9], while Schubert is known not to have a distinct sound signature (his composition style is variable) [10]. Our question was why, and how could this be explained from a mathematical point of view? As well it is worth mentioning that Bach was a baroque composer, Beethoven was a classical composer whose music opened the way to romanticism, and Schubert was a romantic composer. At the same time, we have covered three different types of music eras.

Another reason what motivated us to choose these three composers was also based on the writings from [8], which describes their styles. For Bach's style: "His musical structures are built up into great units which posses an internal balance and overall sense of structure that may be described as architectural". As well, in [12], the reference to Bach's style is: "In instrumental pieces, Bach did not just write notes, but always was respectful of the distinctive sounds and effects of instruments.". Beethoven's style has classical and romantic elements, predominantly classical, and the sound variation of his music captures a distinctive form that is recognizable. Schubert's style is a "freestyle", where his scores are harmonically similar in the classical sense. However, there is a difference in texture and free modulations of key, which creates a larger sound variation; and therefore, correlated to variation in his compositional form.

Through the recordings that we sampled, our feeling was that Bach and Beethoven were more uniform in their compositions than Schubert, and we wondered whether the measures that we suggested would be consistent with that view.

We used the following piano recordings of Mickey H. Melnyk:

- Bach: Prelude, The Well Tempered Clavier, Book 1, No. 1 C major; Partita No. 3, Gavotte en rondeau, E major; Prelude, The Well Tempered Clavier, Book 2, No. 16 G minor. For the sound comparison with a different instrument (violin) we used the recording in [2].

- Beethoven: Moonlight Sonata, Op. 27, No. 2 C-sharp minor.

- Schubert: Impromptu, Op. 90, No. 3 G-flat major and Piano accompaniment; Gretchen am Spinnrade, D minor without voice; Sonata in A major, Op. posth. 120.

2   Other Works Done in Studying the Sound of a Musical Instrument

Many scientists study digital sound. The Fast Fourier Transform (FFT) is the most applied method in research about the study of the sound of an instrumental piece. As well, there is research involving dynamic spectra and wavelets on the sound of instrumental pieces. For the purpose of our study, only FFT is relevant to mention; the dynamic spectra analysis (spectrograms) isolate the spectrum of notes to capture the "essence" of the melodic structure in a musical passage, and wavelets help in computing scalograms to study the rhythmic structure in a musical passage via the Continuous Wavelet Transform [1].

FFT is an excellent optimized Discrete Fourier Transform (DFT) algorithm, [6] and [11], that extracts the main frequencies from a set of samples, which in our case are the main pitches from the raw data of an audio file. FFT converts the time-domain input signal into a frequency-domain picture, which is useful in studying the intensities of main pitches. In Figure 1 we can see FFT applied on recordings of Bach, Beethoven and Schubert. Even if the pictures look somewhat different, we cannot claim that they give any clear sign of a distinctive sound detection. We notice that Bach and Beethoven look somewhat similar, while Schubert looks rather different; this is how far we can go about a distinctive sound detection by using FFT, which is not enough to make an informed conclusion about the mandate of the present work.

The main use of Fourier analysis for audio files is in determining the chroma structure. The chroma structure of an audio file refers to Pitch Class Profiles (PCP). The profiles show the signal intensity distribution throughout a predefined set of pitch classes. An example of such analysis is found in [3].

An important conclusion for the present work: when detecting a distinctive sound in classical music we

need to focus on the entire time range of an audio file instead of overlapping small time periods from it, as FFT algorithm does. We had available many audio files that averaged to forty seconds per each file, and so the audio files contained large data.

The Probability Density Function of Normalized Amplitudes is our proposed method to decide how uniform or variable a composer's style is.

(a) *Bach, Prelude No. 1 (1,396,204 points)*

(b) *Beethoven, Moonlight Sonata (1,416,624 points)*

(c) *Schubert, Impromptu (908,212 points)*
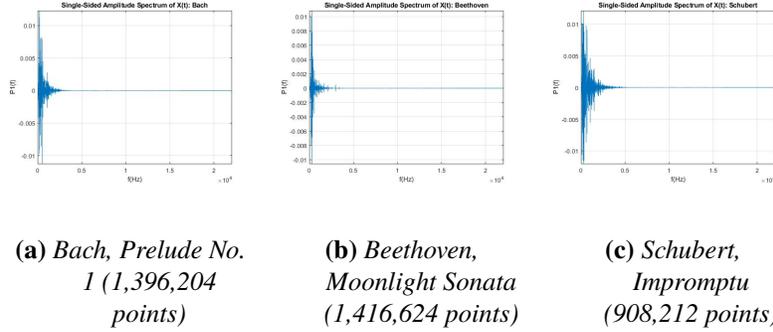
Figure 1: *FFT performed on Bach, Beethoven and Schubert respectively.*

3   E-ratio, EZ-ratio and Probability Density Function of Normalized Amplitudes

Sound is a continuous wave determined by two components: frequency and amplitude. Both, frequency and amplitude, vary in time. From digital audio files, such as Wav(.wav), FLAC (.flac), MP3 (.mp3), MPEG-4 AAC (.m4a, .mp4), OGG (.ogg), and certain compressed WAVE files, it is very hard or impossible to get the amplitude data in decibels (dB). But these audio files give us the normalized amplitudes, directly or indirectly, by applying a normalization process. We propose to use E-ratio, EZ-ratio and the probability density function of the normalized amplitude (PDFNA) to detect a distinctive sound in classical music.

3.1   E-ratio and EZ-ratio

For audio data analysts, the zero-crossing rate ZCR is very familiar, and it defines the ratio of the number of zeros to the sample size. ZCR measures the times that the amplitudes pass the value zero, and it provides us a way to measure the frequency of an audio wave. We propose a similar rate, E-ratio, to measure the ratio of extreme (significant) amplitudes of an audio wave.

We define the E-ratio as follows: $ER(\beta) = E(\beta)/n$, where $E(\beta)$ is the total number of extremes (local maxima and local minima) whose absolute values are bigger than a given positive constant $\beta \in (0, 1)$ and n is the sample size. $ER(\beta)$ provides us with information about significant amplitudes of an audio wave.

We define the EZ-ratio as follows: $EZR(\beta) = ER(\beta)/ZCR$. $EZR(\beta)$ measures the ratio between the number of significant amplitudes and the number of zero-crossing.

In the analysis of our audio files we used the proposed method, PDFNA, and the above mentioned ratios.

3.2   The Probability Density Function of Normalized Amplitudes

As we mentioned in Section 2, digital data files give us many sample points of a continuous wave. The left of Figure 2 is the plot of a sample from Bach Prelude No. 1 with $1, 396, 204$ many data points (the same that we used in Section 2). The right of Figure 2 is the plot of the first 2000 points, which correspond to about half second of music. As explained in Section 2, when detecting a distinctive sound in classical music, we need to focus on the entire time range instead of overlapping small time periods of an instrumental piece, i.e. we need to study the amplitude and frequency spectrums over the entire time range. For amplitude, it follows from the top plot of Figure 2 that it has more chance to have low to medium amplitudes, and not large
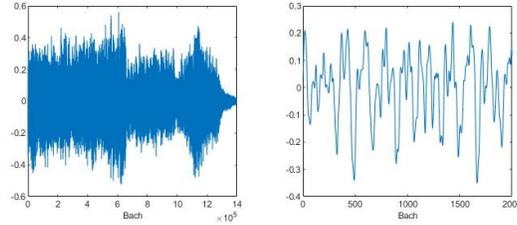
Figure 2: *Left: plot of whole data points of Bach; Right: plot of first 2000 data points of Bach*

amplitudes. For frequency, we can measure how many times the wave crosses the baseline (i.e. zero). Thus, the bigger the frequency means the more chance that the wave passes the baseline. Therefore we propose to use the probability density function to measure these two important components, amplitude and frequency, at the same time in the entire time range recorded. For Bach, the PDFNA is given by the kernel method (left picture in Figure 3a). The confidence interval of the kernel density estimator is given as well (right picture in Figure 3b). Now we will give a brief introduction of the kernel density estimation method. More details are found in [5] and [13].
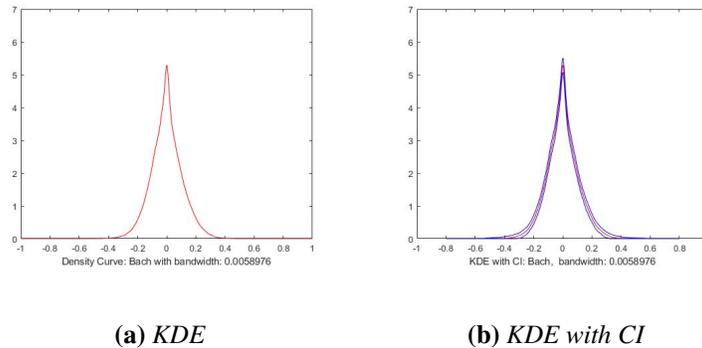
(a) *KDE*   **(b)** *KDE with CI*

Figure 3: *KDE and KDE with CI for Bach.*

3.3 Kernel Density Estimation

Consider a continuous random variable X with the probability density function $f(x)$. The probability density function is an integrable non-negative function $f(x)$ over a domain D.

For a continuous random variable it is usually impossible to collect all the information to find the density function. We need to estimate the probability density function based on a sample $\{x_1, x_2, \cdots, x_n\}$ with the sample size n. There are many ways to do so, such as empirical cumulative distribution functions, histogram estimates, kernel density estimates or penalized likelihood approaches. We will introduce the kernel density estimate (KDE) method. We will follow the tutorial of Chen [5].

In statistics the kernel function $K(x)$ is a nonnegative smooth and symmetric function. A very popular kernel is the Gaussian kernel $f(x) = e^{-x^2/2}/\sqrt{2\pi}$. If $K(x)$ is a kernel function, by the change of variable, $K_h(x) := K(\frac{x}{h})/h$ is also a kernel function. For an independent and identically distributed sample $\{x_1, x_2, \cdots, x_n\}$, the KDE gives us an estimate $\hat{f}_n(x)$ of $f(x)$ as follows

$$\hat{f}_n(x) = \sum_{i=1}^{n} K_h(x - x_i)/n. \tag{1}$$

Here h is the bandwidth, which plays an extremely important role for obtaining a quality KDE. Minimizing the mean integrated square error (MISE) of $\hat{f}_n(x)$ is one way to select the bandwidth: $\mathbf{MISE}(\hat{f}_n(x)) = \int E(f(x) - \hat{f}_n(x))^2 dx$. When the kernel is Gaussian, the optimal bandwidth is about $\hat{\sigma} \cdot (3n/4)^{-1/5}$, where $\hat{\sigma}$ denotes the sample standard deviation. Theoretically the choice of the kernel function, $K(x)$, has very little effect on the estimator. But in practice, the choice of the kernel function makes a difference. In this paper we will apply the kernel method to different data files using the Gaussian kernel and the optimal bandwidth.

3.4 Confidence Interval of KDE

In (1), $\hat{f}_n(x)$ is the average of $Y_i = K_n(x - x_i)$. Then, we can derive a $(1 - \alpha)100\%$ confidence interval as

$$\hat{f}_n(x) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\frac{\hat{f}_n(x)\sigma_K^2}{nk}}. \tag{2}$$

Here $\sigma_K^2 = \int K^2(x)dx$. For more details, see [13]. The method to get (2) is known as "plug-in approach" in [5], since the estimator of variance has been simply plugged in. But, only when $h \to 0$ and $nh \to \infty$, this confidence interval is asymptotically valid, see [5]. Thus, the value of h is very small when n is very large. In this paper n is very large, and so (2) is not suitable.

Alternatively, we estimate the variance by applying bootstrap method. We randomly draw B (in this paper, we take $B = 100$) many samples (with size n) with replacement from the original n sample points. Based on these B samples, we can get B many KDEs and $\widehat{\text{Var}}(\hat{f}_n(x))$ the sample variance as an estimate of the variance of $\hat{f}_n(x)$. Then, a $(1 - \alpha)100\%$ confidence interval can be defined as

$$\hat{f}_n(x) \pm Z_{1-\frac{\alpha}{2}} \sqrt{\widehat{\text{Var}}(\hat{f}_n(x))}. \tag{3}$$

This confidence interval is asymptotically valid regardless of h. Rigorously speaking, this interval is not the confidence interval for $f(x)$ but for $E(\hat{f}_n(x))$ for individual x. In this paper, we will only use (3).

4 Analysis

We had a large number of audio files recorded for the studied composers, and we analyzed them using our MAPLE and MATLAB codes. The audio files captured different time passages for each instrumental piece played by our piano performer, Mickey H. Melnyk. Figures 4, 5, 6 and 7 illustrate samples of our findings. For each audio file, we found the optimal bandwidth, and we assured the best fit of the density estimate to the true density. A bandwidth that is too small will have the tendency to under-smooth the density estimate, while a bandwidth that is too large will have the tendency to over-smooth the density estimate. In both cases, the density estimate of the data will not give the closest fit to the true density. That is why we determined the optimum bandwidth for each audio file. The choice of bandwidth is essential for kernel density estimation, [4]. In the present work (we used the Gaussian kernel) the optimal bandwidth is $\hat{\sigma}(3n/4)^{-1/5}$. Consider two different samples with sample size n_1 and n_2 correspondingly. When n_1 is very close to n_2, $(3n_1/4)^{-1/5}$ will be very close to $(3n_2/4)^{-1/5}$. The optimal bandwidth for the two samples will be the product of the same constant and the sample standard deviation of the corresponding sample. When n_1 is very close to n_2, we can compare the variation of different samples by comparing the optimal bandwidths. Then it is intuitive to interpret the optimal bandwidth as follows: smaller bandwidth means less variation in the sample, while larger bandwidth means higher variation in the sample. In our analysis, a composer's uniform compositional structure and style are given by the Kernel Density Estimation plot (KDE plot) and the optimal bandwidth, respectively. Consistency in the shape of the KDE plots means a consistent compositional structure. Consistency in the optimal bandwidth means uniformity in the compositional style. A variable optimal bandwidth means a non-consistent sound variation in a composer's style that can make the composer hard to identify. In conclusion we look for the consistency of a composer's compositional style by looking for consistency in bandwidth.

For Bach (Figure 4) the KDE plots are consistent on three different pieces performed on two different instruments. Even if melodically the differences between the three instrumental pieces are clear, in terms of structure they suggest that they follow a uniform compositional structure, which agrees with our mention in Section 1 regarding his compositional structure, [8]. All the optimal bandwidths that we obtained for all the audio files that we tested for Bach were consistent and very narrow, about 0.005. Our study on Bach is a step forward in showing that his compositional structure and style are uniform.

(a) *Prelude No.1 (piano)*   **(b)** *Partita No. 3 (violin)*   **(c)** *Prelude No. 16 (piano)*
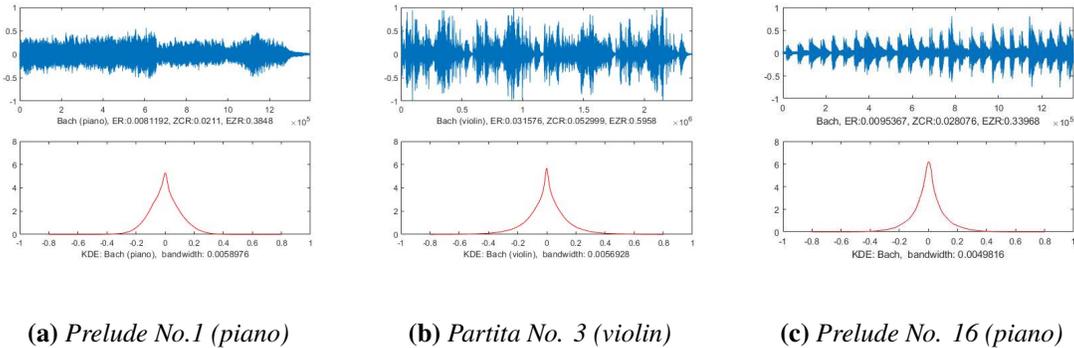
Figure 4: *Bach*

For Beethoven (Figure 5) our findings were very similar to the findings for Bach. Movement 1 and Movement 3 of the Moonlight Sonata sound very different, but in our data analysis we found that the compositional structure of each movement is almost identical (see the KDE plots). We tested audio files for the two movements that captured the entire time passage for each movement, and we also saw a consistency in the optimal bandwidths. All the optimal bandwidths that we obtained for all the audio files that we tested for Beethoven were very narrow, about 0.004. This is a sign that he used a uniform composition structure and style for this sonata. These findings are an encouraging step forward in showing that Beethoven's compositional structure and style are uniform.

(a) *Moonlight Sonata, Movement 1* **(b)** *Moonlight Sonata, Movement 1* **(c)** *Moonlight Sonata, Movement 3*
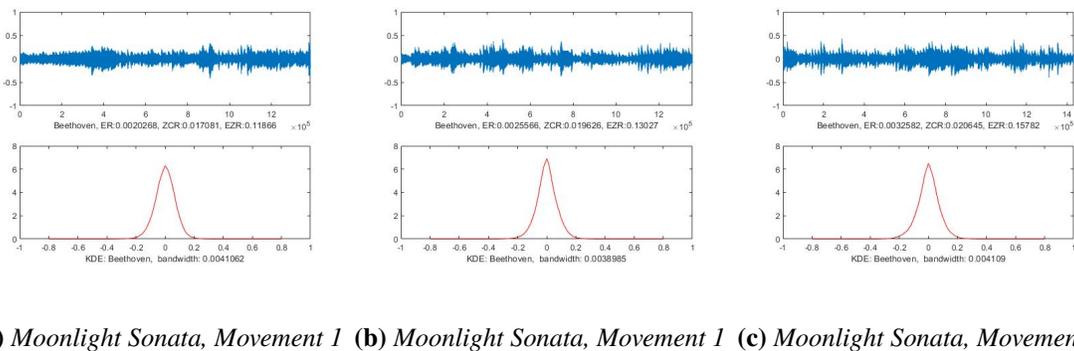
Figure 5: *Beethoven*

In our analysis on Schubert (Figures 6 and 7), we noticed that his compositional structure and style are either non-consistent (Figure 6) or consistent (Figure 7). In the Impromptu, the optimal bandwidth is non-consistent, which suggests that Schubert's style is variable, and there is a non-consistency as well in his compositional structure (non-consistent KDE plots). In the Sonata though, Schubert struck a very structured compositional form (almost identical KDE plots on all our audio files) and a very uniform style (consistent optimal bandwidth throughout the entire instrumental piece, about 0.0035). In Gretchen am Spinnrade, the analysis showed an inconstancy in the compositional style, i.e. a mix of homophonic technique with ostinato.

Our findings show that Schubert displays variation in his compositional style and compositional structure.

As well, our guess is that a composer's uniform or variable compositional style is given by the use of extreme amplitudes versus zero-crossings. The more extreme amplitudes than zero-crossings, the more variation in a composer's style. Among the three composers studied, Schubert showed non-constant use of extreme amplitudes versus zero-crossings in comparison to Bach and Beethoven (see the EZR table below).

EZ-Ratio	Bach	Beethoven	Schubert
EZR (%)	40	14	20-65

(a)   **(b)**   **(c)**   **(d)**
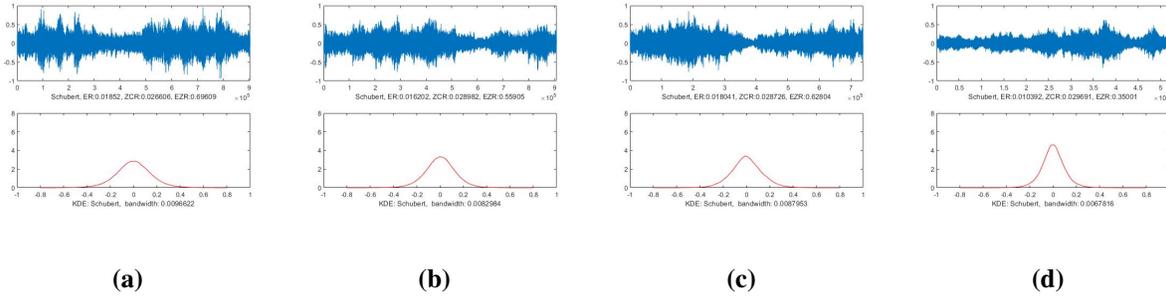
Figure 6: *Schubert, Impromptu No. 3*

(a)   **(b)**   **(c)**
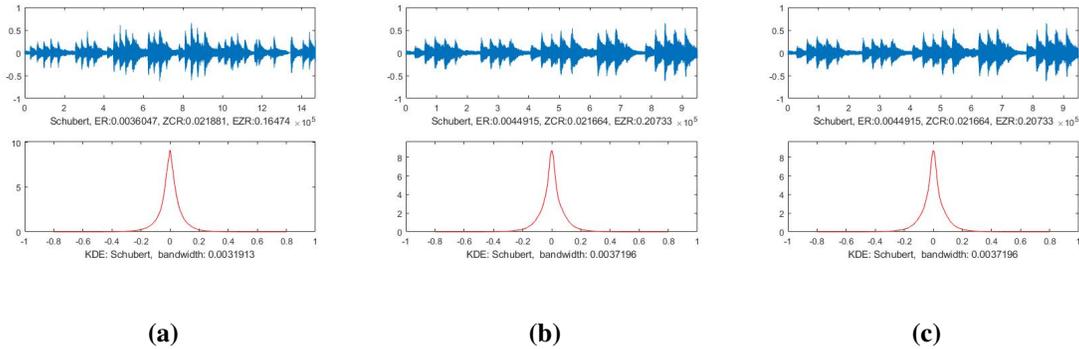
Figure 7: *Schubert, Sonata*

In conclusion, it is important to notice that all the optimum bandwidths that we obtained were small (regardless whether they were consistent or non-consistent), for all the composers that we studied. In our case, it means that the variation window of a composer's style is not large at all per se. In the world of classical music, people debate many times a composer's style, and whether a composer can be identified just by listening to an instrumental piece. The analysis done here makes us strongly believe that our brains are finely tuned to identify very small differences in the variations of a composer's style. If the optimal bandwidth is "consistent enough" (i.e. more uniformity in the composer's style), then it is possible for the brain to identify the composer.

5   Commentary

Our study on Bach showed that he followed a very structured and refined compositional form, [9]. Our data processing on his music suggests that his compositional form leads to minimal variation within his polyphonic technique, which leads us to believe that he created a distinct sound signature for his music. His genius is

known as: "Suffice it to say that for many composers and for countless listeners, Bach's music is supreme−to quote Wagner: 'the most stupendous miracle in all music'.", [9].

Beethoven is known as a very creative composer, [10]. He was the bridge between classical and romantic who radically transformed every musical form in which he worked. Our analysis on his music showed very small sound variations in his compositional form (i.e. uniform compositional style), which leads us to believe that he created a unique and recognizable form for his music.

Schubert's composition style is known as: "He had many styles, and his music sounds like that of other composers", [10]. Our analysis on his music showed a variation in his compositional style and in his compositional structure, which leads us to believe that he followed a "freestyle" in composing his music.

References

[1] Alm, J.,F., Walker, J. S., (2002). Time-Frequency Analysis of Musical Instruments. *SIAM REVIEW, Vol. 44, No. 3*, 457-476.

[2] Bach, J. S., Partita E major, Gavotte en rondeau - Sirkka Vaisanen, violin.
File a2002011001 − e02 − 16kHz.wav,
http : //www.music.helsinki.fi/tmt/opetus/uusmedia/esim/index − e.html

[3] Bello, J. P., Chroma and tonality. MPATE-GE 2623 Music Information Retrieval. New York University.

[4] Chen, S., (2015) Optimal Bandwidth Selection for Kernel Density Functionals Estimation. *Journal of Probability and Statistics, Volume 2015, Article ID 242683, 21 pages*, http : //dx.doi.org/10.1155/2015/242683.

[5] Chen, Yen-Chi (2017) A tutorial on kernel density estimation and recent advances, *Biostatistics & Epidemiology*, 1:1, 161-187.

[6] Cooley, J. W., Tukey, J. W., (1965). An Algorithm for the Machine Calculation of Complex Fourier Series. *Mathematics of Computation, Vol. 19, No. 90*, 297-301.

[7] Digital Recording - History - The Virtual Gramophone - Library and Archives Canada. https : //www.collectionscanada.gc.ca/gramophone/028011 − 3021 − e.html

[8] Hindley, G., (1971). The Larousse Encyclopedia of Music. First Edition. *The Hamlyn Publishing Group Limited*.

[9] Kennedy, M., (1980). The Concise Oxford Dictionary of Music. Third Edition. *Oxford University Press, Walton Street, Oxford OX2 6DP*.

[10] Latham, A., (2002). The OXFORD Companion to MUSIC. First Edition. *Oxford University Press, Great Clarendon Street, Oxford OX2 6DP*.

[11] Lenssen, N., Needell, D., (2014). An Introduction to Fourier Analysis with Applications to Music. *Journal of Humanistic Mathematics Volume 4 | Issue 1*, 72-89.

[12] Marschall, R., (2011). Johann Sebastian Bach. Thomas Nelson, Inc. Published in association with the literary agency of WordServe Literary Group, Ltd., 10152 S. Knoll Circle, Highlands Ranch, Colorado 80130.

[13] Zucchini, W., Applied smoothing techniques: Part 1: Kernel density estimation, Lecture Notes, University of Science and Technology of China, Hefei, China, Oct. 2003.