Vibrato Behavior Analysis on Different Datasets and Automatic Vibrato Detection

Doga Buse Cavdir

Summer Internship Report UPF / 2016
Music Technology Group

Internship Supervisor

Xavier Serra
Department of Information and Communication Technologies
Universitat Pompeu Fabra, Barcelona, Spain

Table of Contents

1. Introduction

Musical vibrato is a fundamental feature based on the periodic pulsation of pitch, amplitude and/or timbre of a musical tone and it is widely used by many vocal, stringed instrument, and wind instrument performers [6,8,9,12,15]. Although it is such a common feature used by many great musicians and singers, its behavior varies from instrument to instrument, also shows differences among performers. Ferrante reported that even the age of the singer affects the vibrato rate and amplitude [4]. Furthermore he reveals the phenomenon about the decrease in vibrato rate over the last century to point out that both vibrato behavior and human perception of good vibrato changes over time.

Showing various and distinctive behavior, vibrato has been made a subject to many research. The first studies on vibrato dates back to the beginning of the 1930s [15]. Although Seashore analyzed vibrato in both singing voice and instruments, most of the researchers concentrated their efforts on the singing voice. Seashore reported the time pattern of pitch and intensity of vibrato claiming all instrumental vibrato is an imitation of the human vibrato [16] in 1937.
Even though vibrato measurement techniques have been improved broadly from 1930s to present, Seashore's conclusions have still been given place in current researches. One developed method over Seashore's studies was sonogram analysis. By this method, Prame examined the variation of vibrato parameter as rate, extend and intonation in ten singer's performance resulting in a vibrato rate of 5 Hz over the duration on 2 s [17]. In later year, Bretos and Sungberg investigated the same parameters in long sustained notes, in other words, for tones considerably longer than 2s sung by 10 soprano singers [18]. Having investigated longer tones, they reached a conclusion that the vibrato rate increased towards to end of the note.

Apart from singing, vibrato is significantly observable in bowed string instruments. In this dissertation, bowed instruments, mostly violin, is largely investigated. Moreover, in [16], Seashore brought an important clarification on that bowed string instrument is not the only instruments where vibrato occurs. Possibly in all orchestral and band instruments, even in fixed instruments such as piano and organ, vibrato can occur either naturally, by the help of mechanical devices or through sympathetic vibrations,respectively. What distinguishes them is the type of the vibrato due to variance in the degree of pulsation. According to [16,19], vibrato can be easily separated as pitch vibrato, amplitude vibrato and timbre vibrato physically, yet not perceptually. Different kinds of vibrato lead into one perception of vibrato movement. Hence, this ambiguity made the analysis and measurement of vibrato more complex.

Traditional approaches in vibrato analysis for automatic detection are based on manual inspection by musicians and engineers [20]. Moreover, various structural differences like shape and differences of vibrato type observed in different instruments, such as intensity vibrato in brass instruments usually perceived as tremolo [16], lead researchers either to put many limitations [20] and prevent them to achieve a detection methodology for all vibrato types or to give up from full automatization. For example, several recent analysis methods are restricted to sinusoidal vibrato only. Likewise, most of the vibrato studies are on Western operatic singing [4,18,22] or Western instrumental sounds in classical music [21]. Despite all the limitations, the labor intensive workload is a major bottleneck in the systematic study of the practice of vibrato [23]. Introducing the interactive toolbar automated detection, AVA in [23], filled the gap in the research area through fully automatic detection, analysis and visualization. Its importance is highly credited to being conducted in non-Western music.

Generally, vibrato detection methods are classified into two methods: note-wise and frame-wise. Note-wise methods needs note segmentation step in order to check vibrato existence in the note [23,24]. On the contrary, in [25] frame-wise methodology, where f0 information is extracted into a number of uniform frames, is applied with a novel approach based on Filter Diagonalization Method(FDM). In Yang and Chew's system using a Decision Tree method for vibrato detection based on FDM output exploits the FDM advantages, primarily the advantages of avoiding the limitations of Fourier Transform Uncertainty Principle [26] and avoiding an error-prone peak-picking process in the spectrum. Henceforth, it allows to extracts the sinusoids directly from the original signal without deriving them from the spectral information and to obtain the sinusoids from short frames [27]. Even though this approach avoids the sinusoid extraction deriving from spectral information, methods requiring direct analysis on signal's spectrogram representation have still been researched and developed. One of these novel vibrato analysis approaches is the template based vibrato analysis recently presented in [3]. Their work created a good alternative to methods based on fundamental frequency (F0) estimation. Local comparison of signal's spectrum with a set of predefined vibrato templates provides an advantage that the analysis does not rely on the estimation of a (possibly erroneous) F0-trajectory [3].Yet, the approach brings one disadvantage demanding high computational complexity due to necessity of use a quadruply nested loop.

In this current research, firstly another novel approach is used for a general behavior extraction. For computational analysis, mainly Essentia which is an open-source C++ library for audio analysis and audio-based music information retrieval is used [7]. From Essentia algorithms, mostly Spectral Peaks, Pitch Salience and Pitch Contours Melody, are used to analyze different datasets [10]. Having obtained pitch contours for each dataset, first derivative of contours are taken. The derivative trajectories are compared to each other, general behavior and exceptions to the common properties are determined. As a result of these comparison and analysis, five datasets are annotated, indicating the note, the presence of vibrato,parameters derived from derivative trajectory, manually extracted vibrato start and end times and finally vibrato durations. The data collection-annotation process and the first computational analysis part are explained in detail in Section 2 and 3,respectively. Secondly a spectral template based approach proposed in [3] is partially implemented.Finally, despite being unable to discuss the results of template based approach, a new idea of implementing another template based algorithm with pitch contour templates of different shapes, rates and extends is proposed to be studied. The details of this more Essentia based approach are given in Section 4. As a conclusion, a summary of overall work presented in this dissertation is provided while some future perspective for the continuation of the current work is discussed

2. Data Collection and Analysis

The first problem encountered during the current research was to find applicable datasets, which were going to be the basis of the analysis. Many datasets or many audio files within datasets could not be used for analysis mainly because of the reasons as different behavior of vibrato in lower pitch values as well as in noisy recordings. In order to observe the general behavior of vibrato by minimizing these constricts, four different datasets with recordings of violin and some mid-pitch woodwind instruments such as transverse flute and alto-saxophone are used. Two of the datasets are created by MTG and 2 are obtained from Freesound [13]. Although different instruments are tried to be involved and analyzed in the research, datasets are mainly for violin recordings with vibrato and no vibrato.

Recordings of MTG

For this vibrato analysis, one of the datasets previously created for Good-sounds[1] dataset and one dataset recorded with a violin to contribute the current research work are used for analysis and for some algorithm implementation trials. In both datasets, the recordings were made in the University of Pompeu Fabra.

i. Alto-saxophone Recordings

This dataset is one of the datasets created in the context of Good-sounds dataset where there are many different sound recordings for various instruments. The reason why alto-saxophone recordings are included in the research is because vibrato and tremolo are both observable in alto-saxophone notes which makes the analysis easier. Yet, it has been observed that it is still not compatible with vibrato behavior of violin resulting in a challenging study for automatic vibrato detection.

To be more precise about the incompatibility with violin, vibrato in alto-saxophone is observed as either square waves or triangular waves and their sinusoidal behavior is less significant then violin vibratos'. The other distinctive property which clarifies this difference is that in brass instruments vibrato is less observable than tremolo which could result in better sinusoidal forms in amplitude of signal rather than than pitch trajectory.

However, from the melody point of view, string instruments tend to show more irrelevant vibration at note changing points which affects the derivative peak analysis, explained in more detail in the following sections. Calculating the unwanted peaks in the peak percentage of the pitch contour's derivative derived the need to seek new parameters to test their periodicity as well as the frequency of the peaks. Moreover, periodicity parameter seemed to be useful in determining the shape of the vibrato on the pitch whereas the frequency could reveal some information about the rate.

Some examples of alto saxophone vibrato forms in melodies and notes are given with their corresponding violin vibratos in Figure 1,2,3 and 4.

---

[1] A real time system for measuring sound goodness in instrumental sounds.

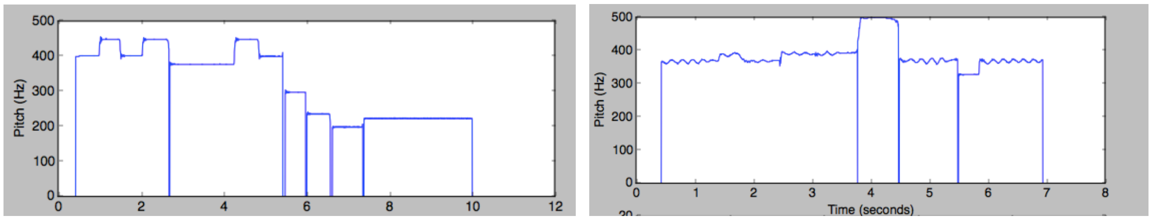

Figure 1: Pitch trajectories of two different melodies where frame size equal to 512 is used. Alto saxophone on the left, violin on the right.

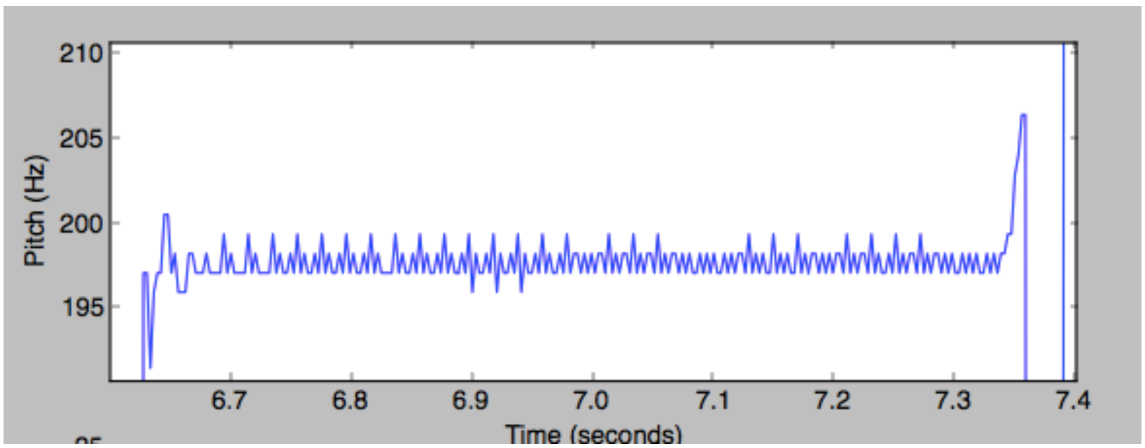

Figure 2: Vibrato of the alto saxophone between time instance 6.6 : 7.3 seconds

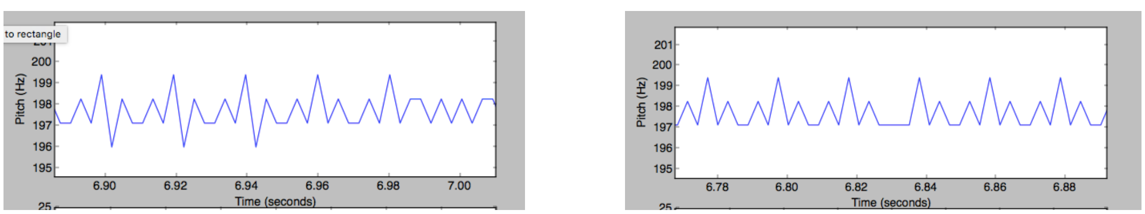

Figure 3: Detailed shapes of alto saxophone vibrato between 6.90:7.00

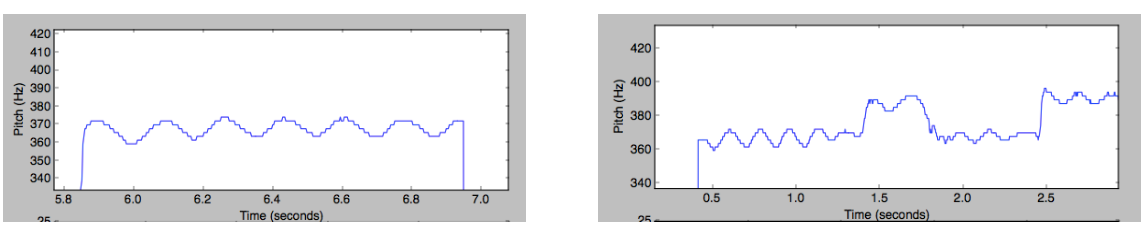

Figure 4: Detailed shapes of violin vibrato between 5.58:7.00 seconds on the left hand side and 0.5:2.5 seconds on the right hand side.

The fluctuations on the pitch contours of alto saxophone is observed better when the frame size is reduced to 512 from 2048. When the frame size is set to 2048, there is no attenuations as if it does not embody any vibrato. This contradiction led to analyze the spectrogram of the recordings which are known to be played with vibrato. Spectral analysis again did not reveal vibrato existence. Lastly, referring to the vibrato/tremolo analysis for numerous instruments explained the reason why it did not visualize the fluctuations as in violin or flute even though it is possible to hear [2]. It is because tremolo is more dominant than vibrato in most of the brass instruments even though it sounds almost identical to vibrato. The spectral comparison of vibrato and tremolo from DiCristofano's and Manaster's analysis is illustrated below [2]. Figure 5 represents the spectral appearances of tremolo and vibrato.

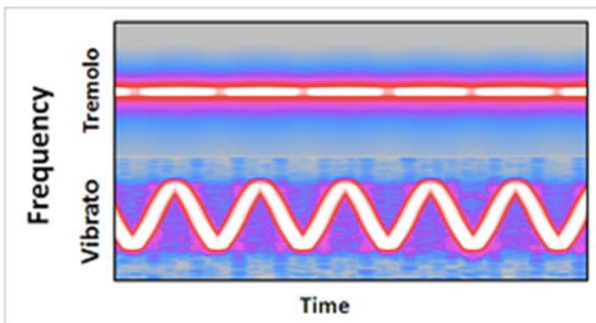

Figure 5: Vibrato and Tremolo
(Frequency vs Time)

As can be seen from Figure 6, a similar shape is observable in the spectrogram of a single alto saxophone note with vibrato. The sinusoidal attenuations are more significant in the higher harmonics. The following spectrogram is visualized by Sonic Visualizer, a software from Queen Mary University of London.

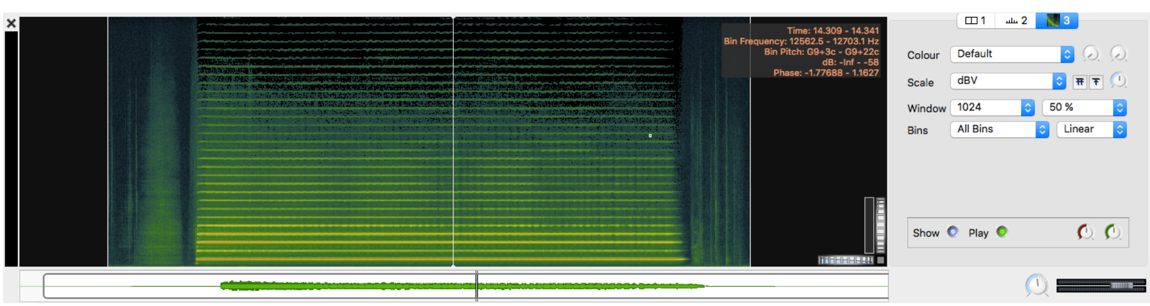

Figure 6: Spectrogram of a single alto saxophone note

Furthermore, the dataset annotations reveals not only whether the vibrato in recordings are good or bad vibratos, but also gives more information of some features like timbre, timing and some techniques. Apart from the extensiveness, variety in features obscured the attempts to a novel algorithm creation. From all reasons above, a new dataset for violin was created, audio files were recorded and annotated.

ii. Violin Recordings

In addition to the good-sounds recordings of alto sax, violin audio files with different rates of vibrato are recorded. This dataset includes two types of recordings. One is single note recordings with three different vibrato rates between G3-A#5 while the other is recordings of melodies with different starting notes. In both cases, when vibrato exists within the note or melody, there is significant sinusoidal variation.

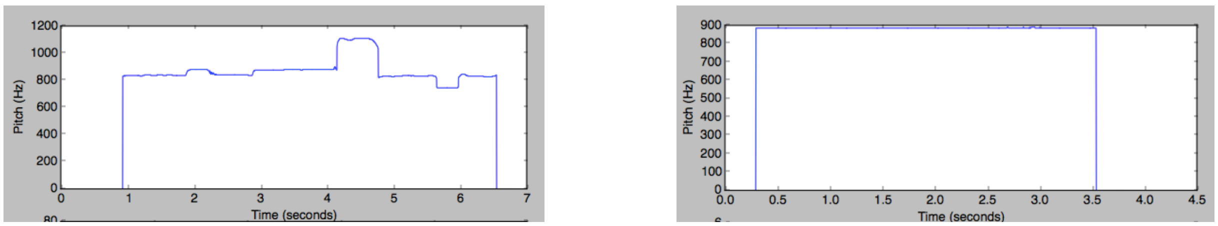

Figure 7 : Pitch trajectories of no vibrato cases of a melody on the right, of a single note G#5 on the left.

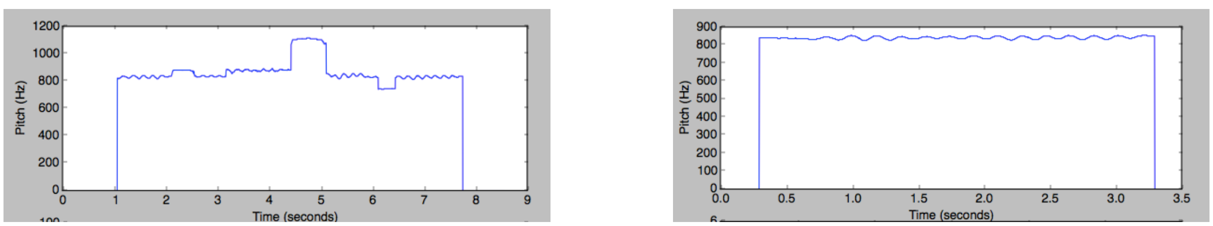

Figure 8: Pitch trajectories of vibrato cases of a melody on the right, of a single note G#5 on the left.

In the dataset, not only some melodies and notes are included but also it consists of different rates of vibrato such as slow rate, standard rate and fast rate.

Here some illustrations of the pulsation in the pitch with different rates are plotted in Figure 9 and Figure 10.

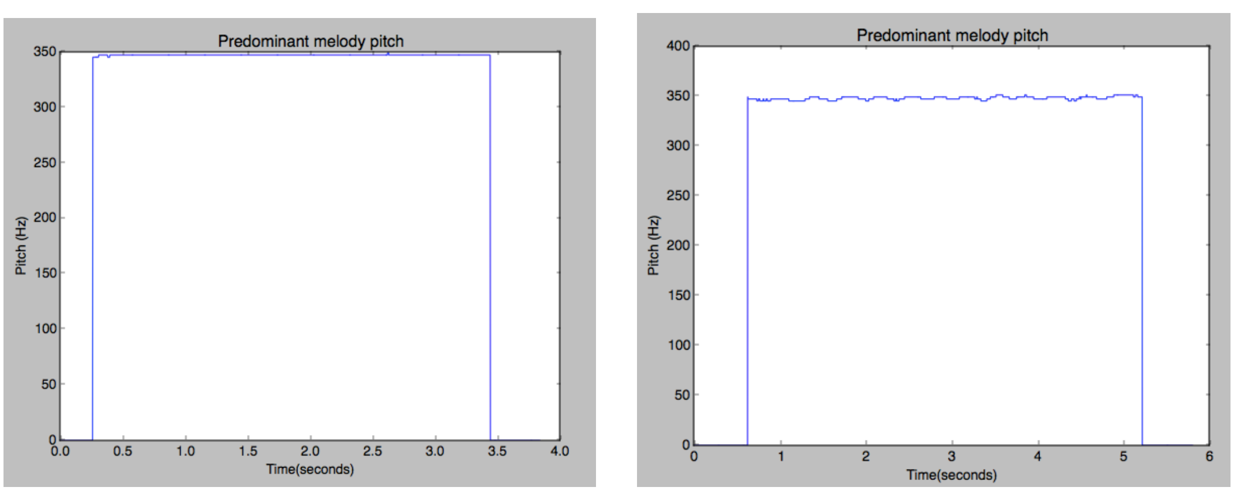

Figure 9: Pitch trajectory with no vibrato and slow vibrato.On the left hand side, pitch trajectory of F4 note with no vibrato is present. On the right hand side, same pitched note has vibrato at a slow rate.

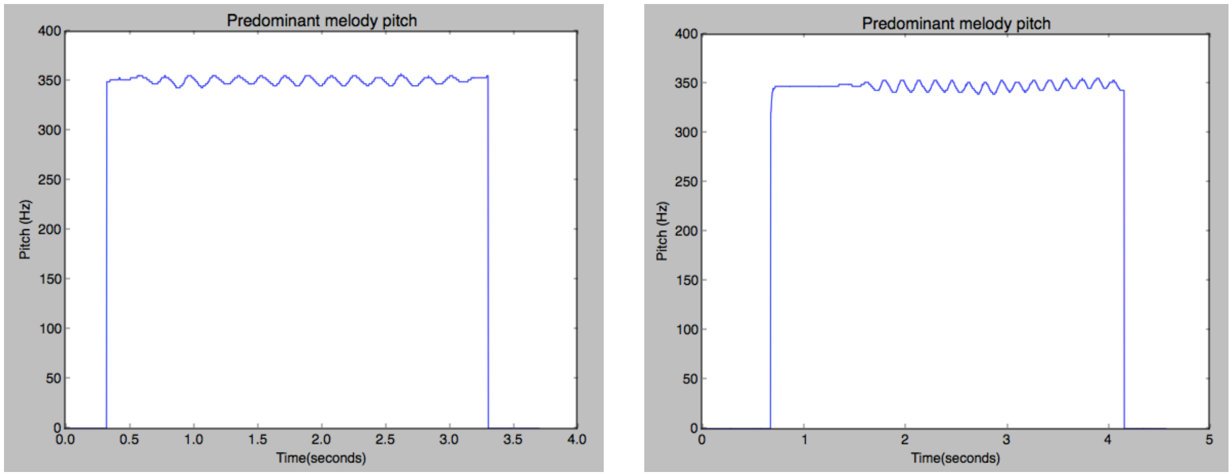

Figure 10: Pitch trajectory with standard rate vibrato and fast vibrato.

When vibrato rate is increased, the observability of the sinusoidal attenuations on the pitch trajectory is also more observable. As seen in Figure 7, the note with standard vibrato (on the left) and the note with faster vibrato (on the right) have better and more uniform sinusoidal forms compared to how slow vibrato appears on the trajectory as presented in the right plot of Figure 6. Another important aspect requires attention is that when the rate is increased, the violinist tends to play the vibrato towards to end of the note and more forte. This behavior is also observed by Gleiser, Friberg and Svante [5] in 1998 and discussed in their quarterly progress and status report. Similarly, when the rate is decreased, vibrato expands over the whole note while its strength is lower and its dynamic intention is close to piano. According to the player participated in the recordings, this type of dynamic change is necessary because it gives a more natural sensation of vibrato. This observation could be profoundly beneficial considering in the dataset, there are some recordings showing discontinuous fluctuation even though no vibrato is present.

While discontinuous wave forms of no vibrato pitch trajectories are the most common discrepancies, it is worth to discuss the rest inconsistent behaviors challenging the information retrieval studies on vibrato.

As mentioned above while summarizing the alto saxophone vibrato, one of the behavioral difference is the shape of the vibrato appearance on the pitch trajectory.

iii. Freesound's Dataset

Even though datasets discussed previously offer extensive collection, it is crucial to analyze different datasets to apply and test the methodologies discussed in this research. Therefore , three datasets from Freesound were implicated in the framework of the current research. Those datasets were created by a user named Carlos_Vaquero,a former master student from MTG. Within the wide collection of instrument recordings, one dataset for transverse flute,one dataset for violoncello and one dataset for violin, both consisting of vibrato and no vibrato audios were selected. For all three instruments, non vibrato and vibrato recordings are made according to tuning frequency (concert pitch) of 442 and dynamic intention of mezzopiano.

According to Vaquero’s comprehensive research on instrumental sound description, some instruments in his recordings have questionable non vibrato steadiness. This situation is observable in flute recordings included in the current vibrato research. He indicates the reason is most probably derived from the fluctuations in the energy of the sound. In consequences of this unexpected attenuations, some fallacious results may appear during non vibrato sound analysis. Potential misleading behavior of flute is not only because of its fluctuations in energy but also it is because of certain deviations in the timbre depending on the amount of air blown into for each note. Nonetheless, he also pointed out the fact that the number of samples are broader than other wind instruments. This has become one of the reasons why flute is selected for analysis.

For string instruments from his dataset, it is essential to note that vibrations on the left hand are transferred to the strings while playing vibrato. This situation resulted in fluctuating pitch and loudness. That is why vibrato templates have distinctions depending on the technique, position, player and whether it is played with movement of the finger, hand or arm or a combination of these. The inconsistency of vibrato behavior in violin can also be seen when pitch and derivatives of two different datasets are compared.

3. Derivative Analysis Method and Statistical Feature Extraction

The numerical differentiation of digital signals is a widely used implication in music and speech processing. Since voiced speech also involves periodic vibration of vocal folds, derivation of the pitch is frequently applied in the linguistic studies [9,14]. Similarly, the focus of a previous research on emotion recognition in speech is based on feature extraction by pitch derivative statistics [14]. The literature, titled as " Recognizing Emotion In Speech" by Dellaert, Polzin and Waibel [1], reveals seventeen features by only using the pitch information in two different extraction ways. As novel as their first method where they implemented various statistical analysis on the pitch contour, an akin approach is applied in the current research.
In this derivative analysis, first derivatives of pitch contours from recordings with different vibrato rates are compared. A general behavior of vibrato other than the attenuation behavior on the pitch trajectory was searched out. Apart from few exceptions, the first derivative of pitch trajectory showed certain common characteristics regarding rate, extend and duration of vibrato, dynamic intentions of playing, recording quality and some others.

Briefly stated, it is observed that there are more peak values in the derivative trajectory as the vibrato rate increases. Ideally, there is no fluctuation in the no vibrato recordings' pitch derivatives, however either due to noise or due to short attacks resulting from some dynamics such as pizzicato. Furthermore, certain facts such that non vibrato sounds are not always as steady as expected and that some fluctuations on the pitch is inevitable due to the certain extended techniques of some instruments cannot be discarded.

Four different vibrato rate for F#4 is presented in Figure 11.1,11.2,11.3,11.4 with their first derivative plots. By taking the first derivative of the pitch contours, three features are extracted. First one is the percentage of how many peaks appear in the derivative trajectory. This feature is the main one to distinguish the vibrato presence. However, since there are quite a few exceptions, this feature is not enough by itself. Second one is the mean value of the distances between peaks. Third one is the max distance between each peaks. The last two features are computed to understand the reason whether the peak percentage is high.To be more precise if it is because vibrato is present or because there are some fluctuations due to certain constrictions explained in the following sections. Overall, this method works mostly for violin tested on our recordings and Carlos's recordings. Also, with this method, some compatible results are obtained for Carlos's flute and violoncello recordings. For brass instruments such as alto saxophone, tremolo could be more significant.

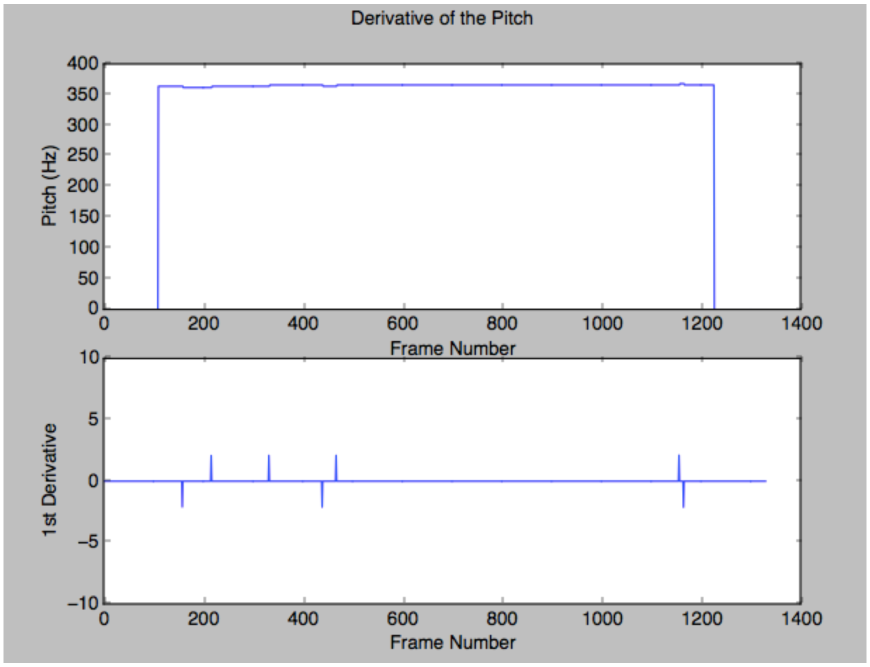

Figure 11.1

Type of vibrato: no vibrato
Mean of the differences: 167.83
Max of the differences: 689
Peak Percentage: 0.53

In the presence of no vibrato, there observed less peaks in the derivative trajectory than recordings with vibrato as expected. Another important observation is that when notes are played without vibrato, there is a big gap between the peaks at the beginning and end of the sound.

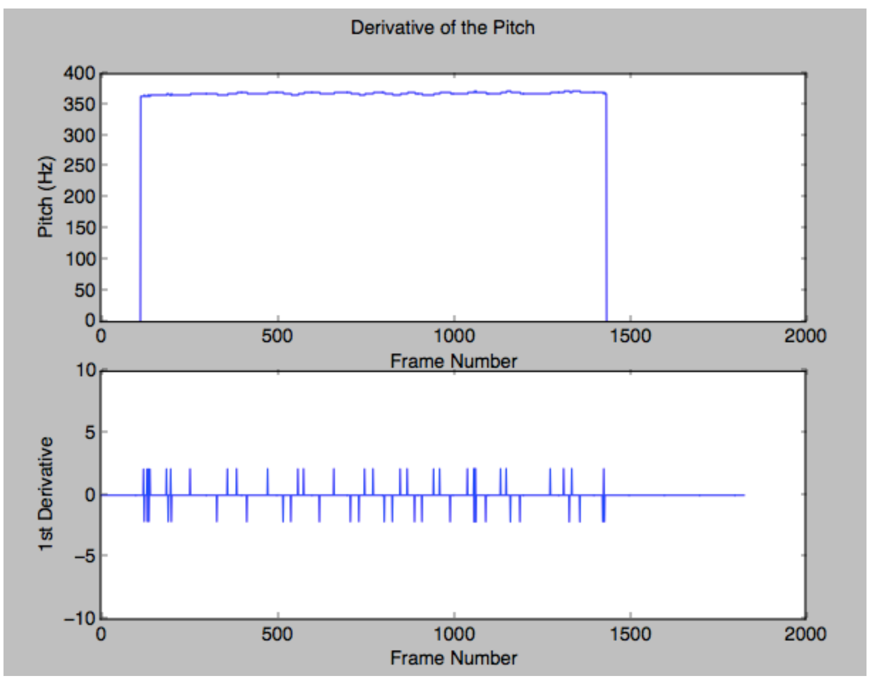

Figure 11.2

Type of vibrato: slow vibrato
Mean of the differences: 22.95
Max of the differences: 86
Peak Percentage: 3.17

When vibrato is performed in a slower rate, on the pitch trajectory it tends to show a more extended waveform with a lower frequency whereas on the derivative trajectory, peaks tend to appear in a smaller quantity similar to some non vibrato trajectories. However, the continuity of the peak trajectory resembles the regular vibrato behavior which is more easily observable in standard and fast vibrato playings.

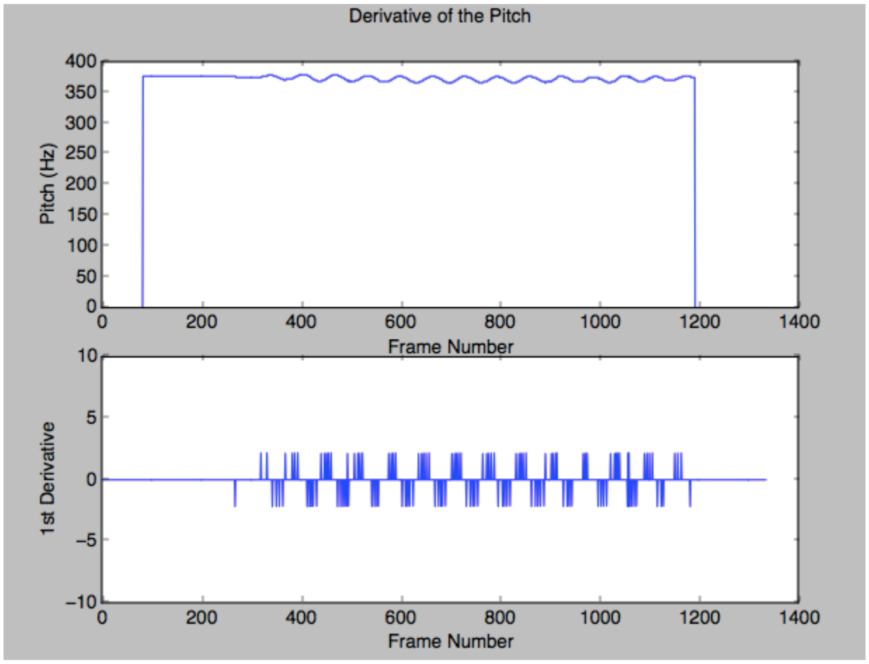

Figure 11.3

Type of vibrato: vibrato at a standard rate
Mean of the differences: 7.38
Max of the differences: 52
Peak Percentage: 9.36

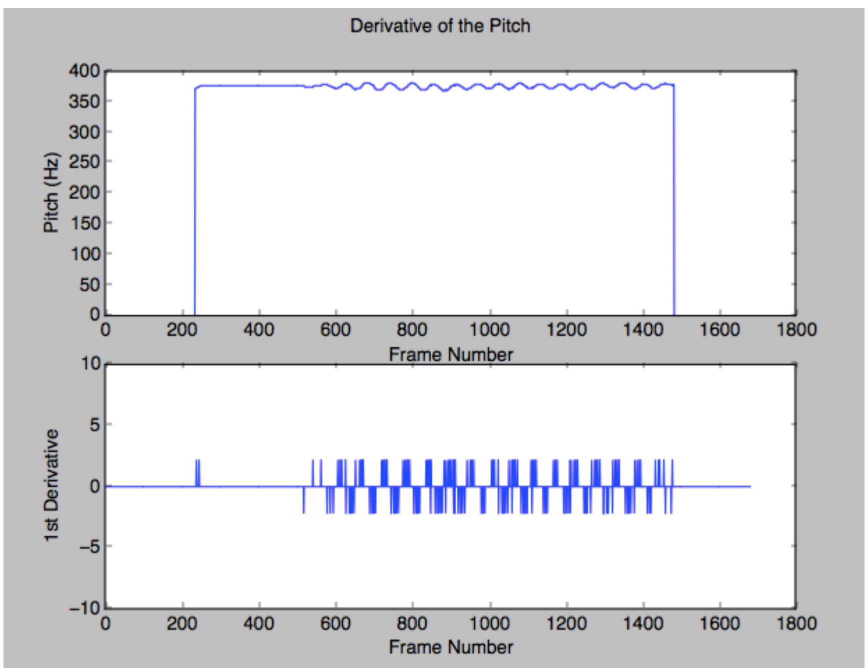

Figure 11.4

Type of vibrato: fast vibrato
Mean of the differences: 7.96
Max of the differences: 273
Peak Percentage: 9.33

Compared to non vibrato and slow vibrato trajectories of the first derivative, vibrato is present in a more continuos and resembling way in standard and fast forms. Despite being highly similar to each other, fast vibrato has a slight difference in the way that performers usually start playing vibrato as approaching to the end of the note. Consequently, peaks resulting from the derivative of the vibrato waveform starts to appear on the trajectory certain frame later than the time instance when the note is played. This is because it is perceived as more natural technique of playing vibrato, thus it is clearly observable in melodies with vibrato.

Having analyzed more sounds from various datasets, it is understood that the first three features are not sufficient by themselves to determine the existence of vibrato in the recordings. Firstly, in the violin recordings, it is observed that fast vibrato derivatives shows a continuous and more periodic peak trajectory but a high maximum difference value at the same time. Hence, this observation required to seek for another feature to locate the position of the maximum difference on the derivative trajectory. By the help of this analysis, it became possible to distinguish a non vibrato sound and a fast vibrato sound. A high positional difference occurs in the beginning of sounds in the case of fast vibrato while it appears in the middle when there is no vibrato. This is the most significant feature of the non vibrato sound since it has no continuous or periodic waveform.

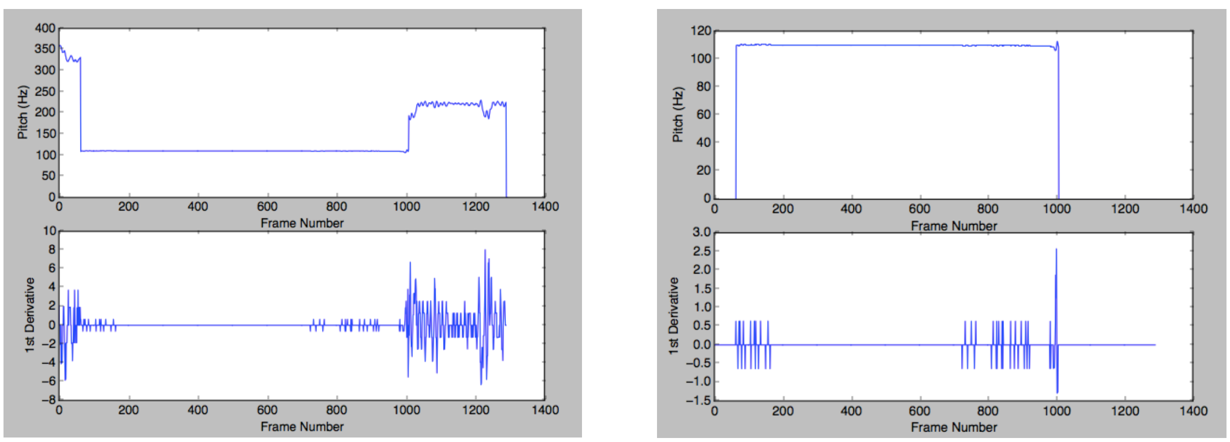

Figure 12: Violoncello single note of D4 with no vibrato. Pitch and statvative trajectories are plotted without doing any filtering presented on the left, also with eliminating the recording noise with the help of pitch confidence values presented on the right.

The other distinctive feature of non vibrato sounds resulting in faulty detection is the noise in the beginning and in the end of the recordings. This is very significant in Carlos's transverse flute and some violoncello recordings played with no vibrato. The same discrepancy is found in MTG's violin recordings, resulted from attacks explaining why there are more irregular peaks in the derivative trajectories of single notes rather than melodies. As Vaquero explained in his thesis, vibrato is highly dependent on the performance abilities.

In order to eliminate the noise at the edges of the pitch trajectory, values with low confidence value are assigned to zero so that the algorithm detected no abrupt change when taken its derivative. An example of eliminating noise as explained is presented in the Figure 12.

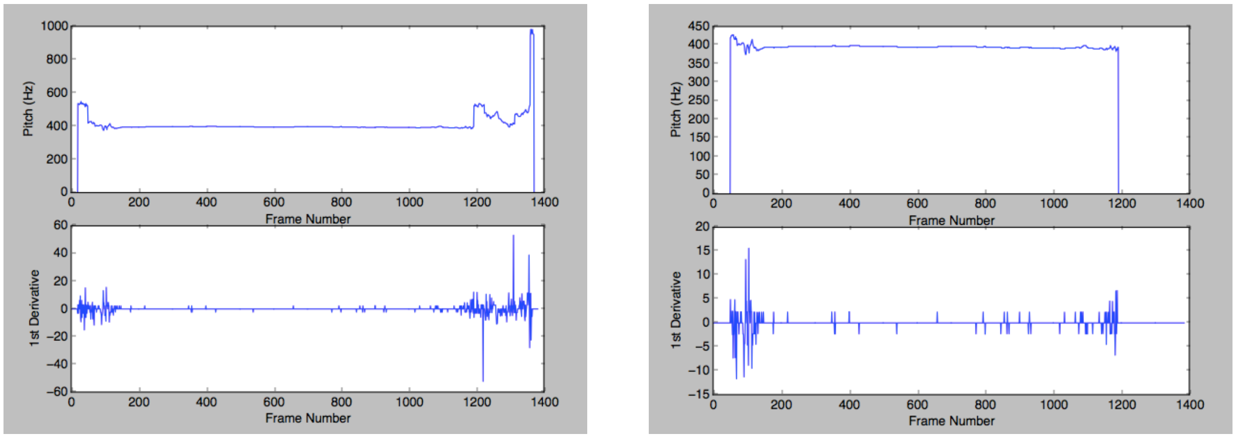

Figure 13: Transverse flute single note G4 with no vibrato. Pitch and derivative trajectories are plotted without doing any filtering presented on the left, also with eliminating the recording noise with the help of pitch confidence values presented on the right.

By eliminating the environmental noise, peak percentage of the derivative is reduced to 4.49 % from 24.11 %. The mean value of the differences between peaks is increased to 16.53 from 4.15. The maximum value of the differences between peaks remained same as expected.
While the peak percentage is 19.6 when there is no filtering, it reduces to 8.59 with noise elimination. However, it can be seen that it is not enough to prevent the noise due to recording sessions to disturb the detection analysis completely. The problem could be solved by either using a different algorithm or cutting the noisy parts manually. The code segment for noise elimination is presented below. It basically calculates the pitch confidence value which appears the most and assigns zero to the values at the edges whose confidence is different than the most common confidence.

```
# Eliminating the noise at the beginning and at the end
abs_confidence = abs(confidence)
m_common = cll.Counter(abs_confidence).most_common()[0]
first =[index for index, val in enumerate(abs_confidence) if val == m_common[0]][0]
last = [index for index, val in enumerate(abs_confidence) if val == m_common[0]][-1]
pitch[0:first+1] = 0
pitch[last: ] = 0
```

Even though removal of the noise reduces deceptively high peak percentage, it is not sufficient to build an algorithm which could detect existence of vibrato in those exceptional samples without any error. An algorithm which generally result in correct estimation is dependent on some threshold values and could be developed as follows.

Since some irregular non vibrato sounds are much close to slow rate vibrato sounds, the distinction of slow rate is discarded in the following methodology. Since the first distinctive property is the quantity of the derivative peaks, several thresholds are set to treat the sounds whether more analysis on them is necessary or the presence of vibrato could be determined. From the analysis results which are accessible in annotation files, it is observed that when sound is played with vibrato, the peak percentage is generally higher than 5 %. Those sound files are mostly the ones played in a faster vibrato rate in MTG's violin dataset. Similarly, when there is no vibrato present on the sound, the average percentage of peaks is 1.15 % in MTG's violin dataset, 1.52 % in Carlos's violin dataset and the maximum values are in between 3-4 % excluding some exceptional recordings. Considering there are too much erroneous analysis results due to noise in non vibrato sounds of transverse flute and violoncello datasets, average peak percentages are higher than the intended thresholds. Discarding those, it is sufficient to say recordings with peak percentage lower than 3 % has no vibrato whereas recordings with peak percentage higher than 5 % are played with vibrato. For those having percentage value in the range of 3-5 %, further analysis with more detailed parameters is needed.

To go further in understanding the derivative behavior of the pitch contour, differences between the peaks are calculated and saved into an array. Afterwards, three other parameters are extracted from difference array. The most appropriate ones at the first glance are mean and maximum values of difference trajectory. For better understanding of the presence of vibrato, it is necessary to analyze them at the same time. The mean value gives the average separation of peaks in the derivative which should not be higher than 100 in general since when the notes are played with vibrato peak waveforms are uniform and periodic. Moreover, a high value of mean of the difference usually leads that there is one or more gaps on the derivative trajectory. That is why it is urgent to analyze the maximum value of differences between peak values. When there is no vibrato present on the note, peaks generally result from erroneous aspects like recording noise or overshoots on the pitch trajectory. This non periodicity of faulty peaks is detected after observing a large gap on the difference array thanks to mean and max value analysis. Despite the fact that all these parameters create an image on vibrato presence, there is still an exceptional case. Fast vibrato recordings are observed to show large gaps on their derivative trajectories due to delayed vibrato playing. Finally, to eliminate the confusion between non vibrato recordings and fast vibrato recordings, the location of the maximum difference value is taken into account as another parameter. Because, it is musically more correct to play vibrato at the end of the note, especially if it is played with "forte" dynamic nuance as in fast vibrato recordings, the large gap between peak locations is appeared in the begging of the note. The close coordination between vibrato and the rhythm is a profound effect in the expressiveness of vibrato depending on the tempo, i.e. fast vibrato resulting an increasing in vibrato rate [19]. On the contrary, because non vibrato sounds contain no fluctuations due to vibrato on the pitch trajectory, there is no uniform or periodic peak attenuations on the derivative trajectory as well. Hence, detected large maximum difference values appears either in the middle of the note or in the end.

Overall, analysis on the first derivate of pitch provides extraction for four different parameters. All reveal different properties of recordings such as presence at first, then how fast it is played, and even rate with a little more analysis on area under peak oscillations.

4. Template based vibrato detection algorithm and its implementation

Based on the approach proposed in the paper,Template-Based Vibrato Analysis of Music Signals, Driedger, Balke, Ewert and Muller [3] present a novel vibrato analysis method in order to avoid the previous analysis based on F0-estimation due to its erroneous practice. As previously discussed in Section 2, it is evident that vibrato shows itself in the form of characteristics spectra-temporal pattern in signal's spectrogram,particularly in the higher harmonics. The authors of the paper further assert that F0 trajectory estimation as the first step of the vibrato analysis may not be trivial, practical and accurate enough to reflect the frequency modulation patterns. In the direction of achieving a moderately novel approach, they aim to detect spectra temporal patterns directly in the spectrogram representation by locally comparing it with a set of predefined vibrato templates. The main idea is represented in Figure 14.

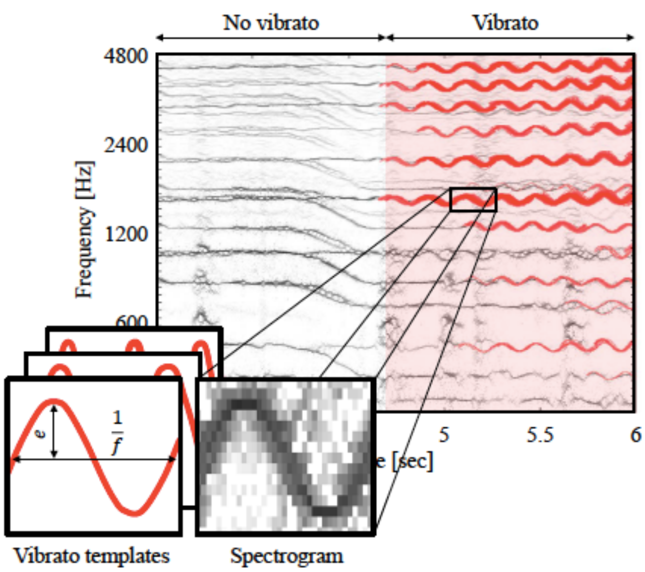

Figure 14: Template-based vibrato analysis. Vibrato templates with different extents and rates are compared to find a matching spectra-temporal pattern in the signal's spectrogram representation. Reprinted from "Template-Based Vibrato Analysis of Music Signals" [3].

In the paper, it is proposed to create a set of sinusoidal trajectories with predefined duration, predefined ranges of frequencies and extends. Moreover, the template waveforms are discretized to have a matching time-frequency resolution with binarized log frequency spectrogram [3].

During the implementation of the proposed method, the templates are created in Python by a simple sinusoidal wave formula

*y = np.sin(2 * np.pi * freq * x / Fs)*

with a frequency range from 5 Hz to 8 Hz with steps of 0.25 Hz. The extend is arranged to have in cents from 50 to 210 with 10 cents steps by

amp = pow(2, (cent / 1200.)).

By the formula given above, cents are converted to frequency values to have a compatible comparison with spectrogram representation obtained by Python. Having a set 225 templates with different rate and extend combinations required a computational complexity of a double nested loop. However, to be able to compare these templates with signal's spectra-temporal trajectories, it will be necessary to use a quadruple nested loop whose computational complexity is so hight that overall analysis for one minute signal takes at least 40 seconds on a standard computer.[*] Yet, being computationally demanding is not the only obstacle encountered throughout the limited implementation period. Computation of a binarized log-frequency spectrogram was not straightforward enough to complete the comparison of all template pattern with signal's spectra-temporal vibrato patterns on its spectrogram. Hence, the detailed correlation wise computation is left to be implemented properly after this dissertation.

Although the template based analysis is not fully concluded, studying the proposed approach created another perspective to follow on the way of vibrato detection. A slight modification to the template based approach from spectrogram to pitch contours could also help to detect vibrato with different waveforms rather than sinusoidal ones. In the previously proposed method, the vibrato template shapes were assumed to be sinusoidal for all time. However, the shape of the waveform changes depending on instrument, sometimes on pitch. Horri made this clarification of possible vibrato shapes into sinusoidal, triangular, trapezoidal, and unidentifiable [28]. Some non sinusoidal vibrato shapes were presented in Section 2 and more are to be presented in Figure 15,16 and 17.

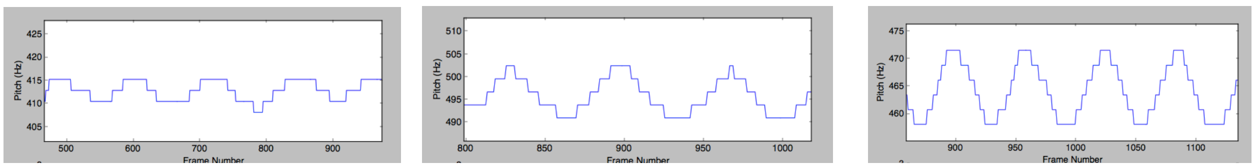

Figure 15: Some of the violin vibrato templates extracted from pitch trajectory. From left to right, slow, standard and fast vibrato rates are presented.

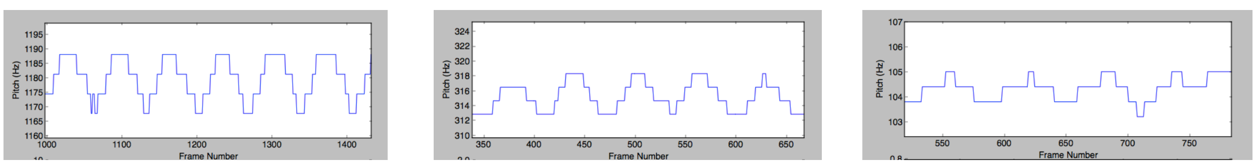

Figure 16: Vibrato templates from flute,violin and violoncello are presented from left to right. On the right hand side, violoncello template is observed to give more similar waveforms when frame size is selected as 4096 instead of 2048 which is the default one for all others.

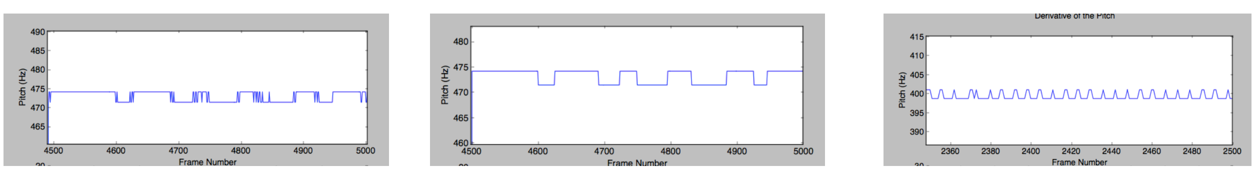

Figure 17: Vibrato templates from alto saxophone with more square or triangular shapes. The first two template are from the same recordings analyzed with different frame size. Frame size 512 is selected for the left one, frame size 2048 is selected for the middle one.

As seen on the alto saxophone templates, they do not have similar fluctuations on the pitch trajectory. The waveforms observed slightly could be deceptive to use as a templates considering some non vibrato recordings also present similar forms in a non periodic way.

Apart from the template shapes, templates may be selected both from recordings with common behavior and with some exceptional ones. By this way, a comprehensive dataset of vibrato templates could be obtained since all the annotations include the information of vibrato presence.

5. Conclusion

Since there is still a lot to uncover, to minimize the need for human involvement, to develop less complex but more novel algorithms, in this study, it was aimed to build a framework for automated analysis of vibrato in monophonic sounds.

To this end, various approaches on many literatures are studied while proposed methods are modified to certain recordings in the current algorithmic framework. Moreover, in order to observe and have a better understanding of vibrato behavior, numerous sound recordings are analyzed with different tools. One of the main tool for the analysis was Essentia standard algorithms. Especially, algorithms calculating spectral peaks and estimation predominant melody and pitch gave useful outputs. Another important tool which is mostly used in the current study is Sonic Visualizer. It's compatibility with Essentia tools and functionality of spectrogram representation made the analysis process much easier. In addition to those, statistical and signal processing functions of Python are used.

The analysis of recordings revealed some conclusions that behavior of vibrato is highly dependent on the instrument, performance, environment where recording are made and even the pitch. Although previous studies analyze and compare various aspects of vibrato, mainly by F0 trajectory tracking or spectral analysis, little attention has been paid to the fact that vibrato does not have a regular shape which sinusoidal forms are mostly assumed to be. The other aspect which is ignored many times is the need to treat vibrato as inseparably with tremolo.

This fairly irregular and divergent behavior of vibrato lead the research to further analysis with differentiation methods and first derivative of the pitch trajectory disclosed some significant results. From the derivative trajectory, the best conclusion happens to be the ease of observing the periodicity in the case of vibrato. By further study, this method could result in a more generalized detection algorithm.

With all the analysis, two recently proposed approaches could be carried out in the future works. More concrete and well experimented conclusion could be drawn with novel methodologies. It is obvious that there will be many limitations to eliminate and corrections to be made as they are implemented broadly which will carry the current research further.

6. References

[1] F. Dellaert, T. Polzin, and A. Waibel. (1996). "Recognizing Emotion in Speech". Proc. ICSLP, Philadelphia, PA, USA, 1970-1973.

[2] A. DiCristofano and A. Manaster (2016). "Vibrato and Tremolo Analysis". Physics 406 L1. Retrieved September 20, 2016 from https://courses.physics.illinois.edu/phys406/Student_Projects/Spring16/A_DiCristofano_A_Manaster_Physics_406_Final_Report_Sp16.pdf

[3] J. Driedger, S. Balke, S. Ewert, and M. Muller. (2016). "Template-Based Vibrato Analysis of Music Signals". Paper presented at the International Conference on Music Information Retrieval (ISMIR).

[4] I. Ferrante. (2011). "Vibrato rate and extent in soprano voice: A survey on one century of singing". Journal of Acoustical Society of America, Vol. 130, No. 3, 2011, p. 1683-1688.

[5] J. Gleiser, A. Friberg, and S. Granqvis. (1998). "A method for extracting vibrato parameters applied to violin performance". *TMH-QPSR 4/1998*. Retrieved September 20, 2016 from http://www.speech.kth.se/prod/publications/files/2945.pdf

[6] M. Zhang, M. Bocko, and J. Beauchamp. (2015). "Measurement and analysis of musical vibrato parameters". 169th Meeting of the Acoustical Society of America, Pittsburgh, Pennsylvania. 2015. Proceedings of Meetings on Acoustics, Vol. 23.

[7] D. Bogdanov, N. Wack, E. Gmez, S. Gulati, P. Herrera, O. Mayor, et al. (2013). "ESSENTIA:an Audio Analysis Library for Music Information Retrieval" International Society for Music Information Retrieval Conference (ISMIR'13). p. 493-498.

[8] H.N. Fletcher. (2001). "Vibrato in Music. Acoustics Australia". Vol.29 No.3. p.97-102.

[9] C. Hsu and J. R. Jang, "Singing pitch extraction by voice vibrato/tremolo estimation and instrument partial deletion," in Proc. 11th Int. Soc. for Music Inform. Retrieval Conf., Utrecht, The Netherlands, Aug. 2010, pp. 525–530.

[10] J. Salamon and E. Gomez. (2012). "Melody Extraction From Polyphonic Music Signals Using Pitch Contour Characteristics". IEEE Transactions on Audio Speech and Language Processing. Aug. 2012, 20(6):1759-1770.

[11] B.A. Mesz and M.C. Eguia. " The Pitch of Vibrato Tones A Model Based on Instantaneous Frequency Decomposition". The Neurosciences and Music III—Disorders and Plasticity: Ann. N.Y. Acad. Sci. 1169: 126–130 (2009).

[12] Timmers, R., and Desain, P. (2000). Vibrato: questions and answers from musicians and science. Proceedings of the sixth ICMPC. Keele.

[13] F. Font, G. Roma, and X. Serra. (2013)."Freesound technical demo." Proceedings of the 21st ACM international conference on Multimedia. ACM, 2013.

[14] A. Gholipour, M.H. Seaaghi and M. Shamsi. (2012). The contribution of prosody to the identification of Persian regional accents. Industrial Electronics and Applications (ISIEA), 2012 IEEE Symposium.

[15] C. E. Seashore, “The vibrato,” in Iowa City: University of Iowa, vol. 1, 1932.

[16] C. E. Seashore, “The vibrato: What is it?”, The Psychology of Music, Music Educators Journal, Vol. 23, No. 4, 1937.

[17] E. Prame, “Vibrato extent and intonation in professional western lyric singing,” J. Acoust. Soc. Am., vol. 102, pp. 616–621, March 1997.

[18] J. Bretos and J. Sundberg, “Measurements of Vibrato Parameters in Long Sustained Crescendo Notes as Sung by Ten Sopranos,” Journal of Voice, vol. 17, pp. 343–352, September 2003.

[19] P. Desain, H. Honing, R. Aarts, and R. Timmers, “Rhythmic aspects of vibrato,” In Proceedings of the 1998 Rhythm Perception and Production Workshop, vol. 34, pp. 203–216, 1999.

[20] S. W. Lee, M. Dong and P. Y. Chan, “Analysis for vibrato with arbitrary shape and its applications to music”,Proc. of APSIPA ASC 2011.

[21] P. Herrera and J. Bonada, “Vibrato extraction and parameterization in the spectral modeling synthesis framework,” in Proc. Digital Audio Effects Workshop, Nov. 1998.

[22] L. Yang, E. Chew, and K. Z. Rajab, “Vibrato performance style: A case study comparing erhu and violin”. In Proc. of the 10th International Conference on Computer Music Multidisciplinary Research, 2013.

[23] H. Pang and D. Yoon, “Automatic detection of vibrato in monophonic music”. Pattern Recognition, 38(7):1135–1138, 2005.

[24] S. Rossignol, P. Depalle, J. Soumagne,X. Rodet, and J. Collette, “Vibrato: detection, estimation, extraction, modification”. In Proc. of the Digital Audio Effects Workshop, 1999.

[25] L. Yang and E. Chew, “Ava : an Interactive System for Visual and Quantitative Analyses of Vibrato and Portamento Performance Styles”. Paper presented at the International Conference on Music Information Retrieval (ISMIR), 2016.

[26] V.A. Mandelshtam, “FDM: The filter diagonalization method for data processing in NMR experiments”,Progress in Nuclear Magnetic Resonance Spectroscopy,2001.

[27] L. Yang, K. Z. Rajab, and E. Chew, “Filter diagonalisation method for music signal analysis: Frame-wise vibrato detection and estimation”. Journal of Mathematics and Music, 2016. Under revision.

[28] Y. Horii, “Frequency modulation characteristics of sustained /a/ sung in vocal vibrato”. Journal of Speech and Hearing Research, 32, 829-836, 1989.