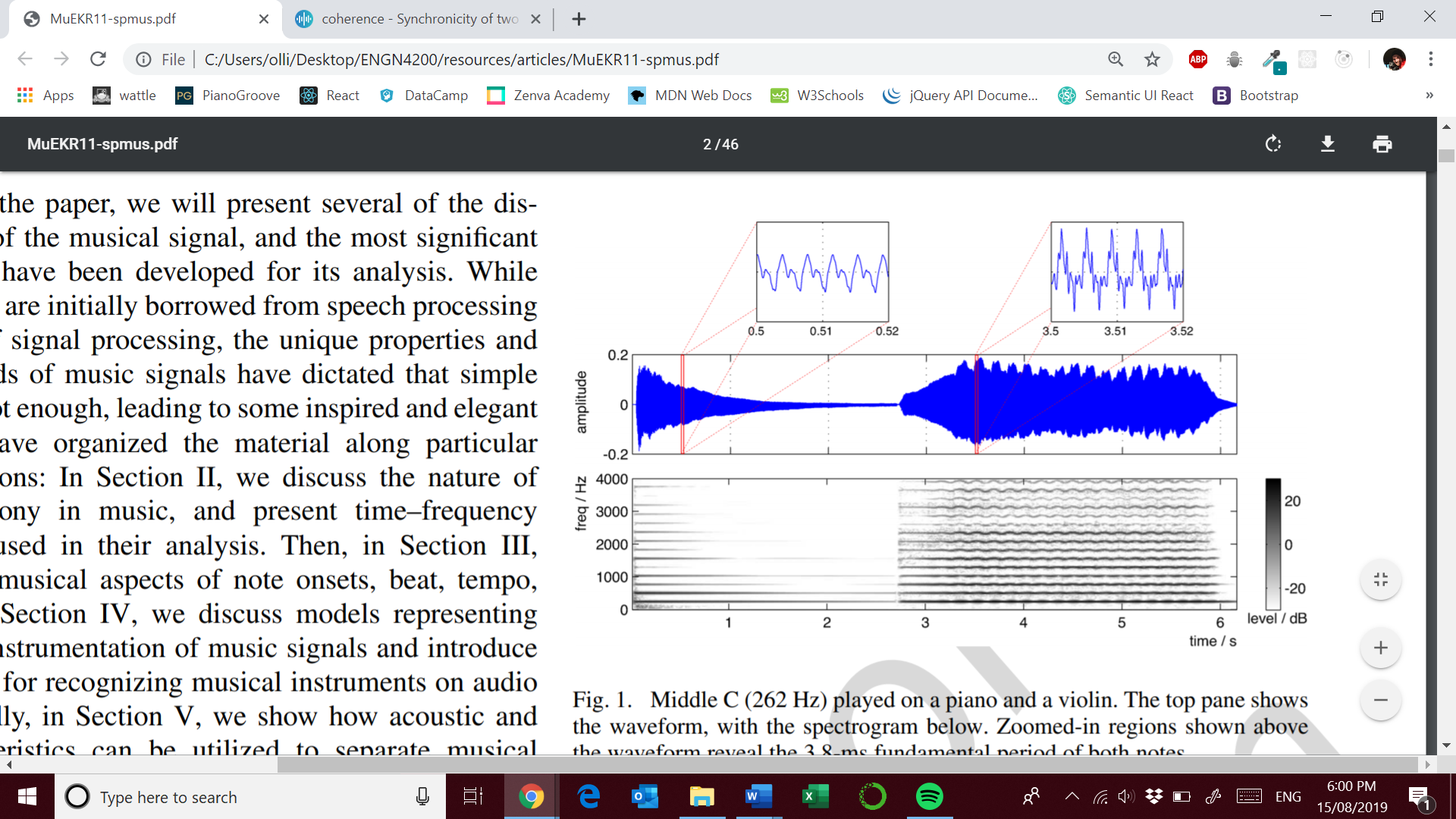
**SIGNAL PROCESSING FOR MUSIC ANALYSIS NOTES**

M. Müller, D. P. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011.

**PITCH AND HARMONY**

Pitch – refers to a sound wave with a well-defined fundamental frequency. The signals produced by most instruments consists of a harmonic series of sinusoids. This harmonic series results in a perceived pitch in the listeners ear.

Humans perceive frequencies that fall in a ratio of 2:1 (Octave) as highly similar due to the embedded sets of harmonics.



Source: M. Müller, D. P. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011.

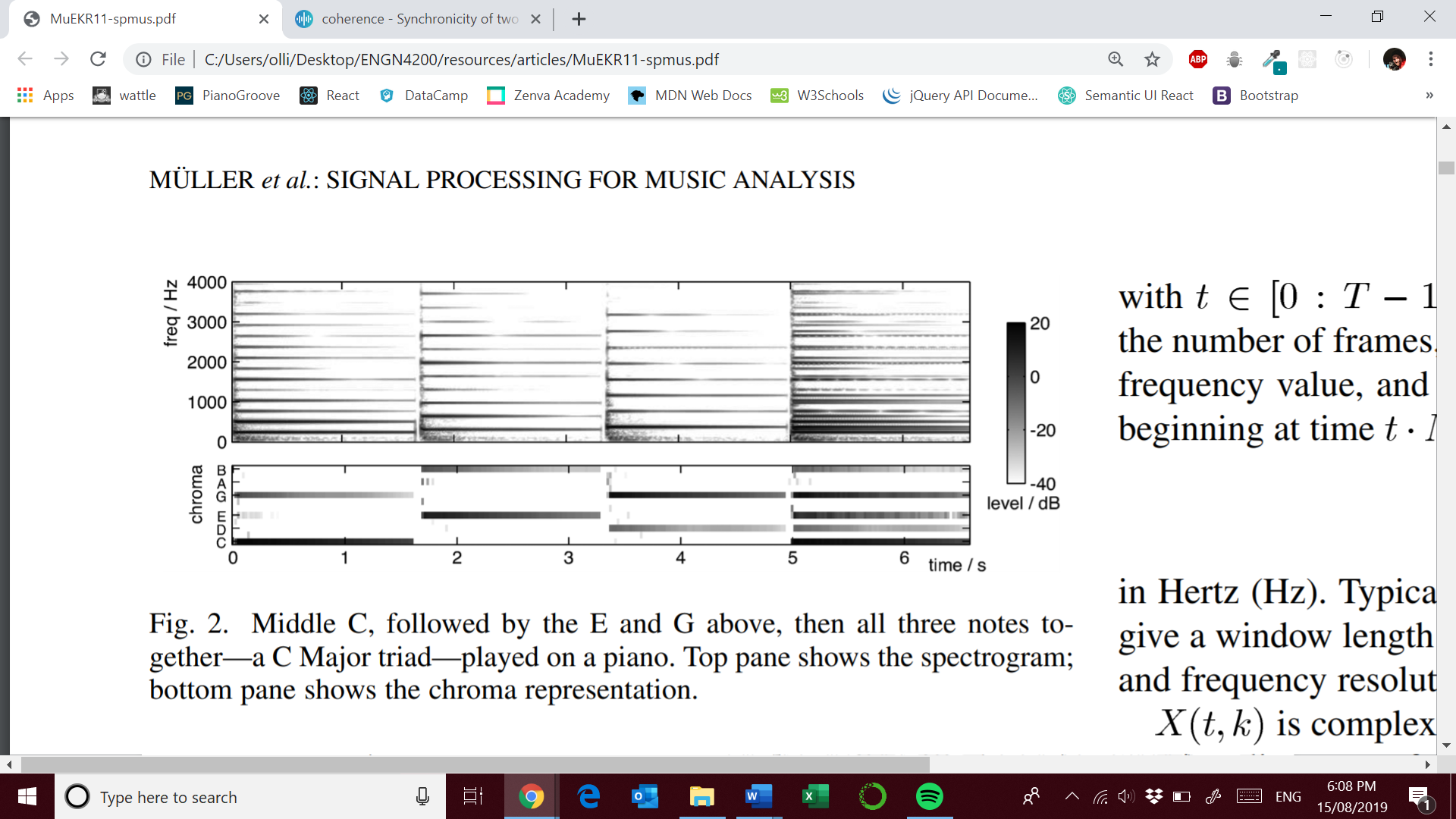
Middle C played on a piano left and violin right. You can see evidence of vibrato ( slight frequency modulation ) in the spectrogram of the violin compared to the clear decay in the piano sound wave.

* Middle C: 240Hz – 262Hz ( there is a range of tunings)

Equal tempered scales allow an octave to be split into twelve equal steps on a logarithmic axis. Each note has a frequency 2^1/12 times larger ( semitone ). (2^2/12)^7 = 3/2 ie. G has a frequency that is 1.5 times larger than C

The lowest note on a piano is A0 (27.5 Hz), the highest note is C8 (4186 Hz), and middle C (262 Hz) is C4.

An octave degree is known as its chroma as seen below;



Source: M. Müller, D. P. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011.

They use parameters :

* Sample frequency: 44.1kHz
* Window length: 46ms -> 2028 samples

You can see that the chroma are shared in fifths. You can also see how much more complex the spectrogram is. Chroma can be thought of as a spectral template of a pitch. Consonant harmonies involve pitches with simple frequency ratios indicating many shared harmonics. The ubiquity of such consonance is a major challenge to automatic music analysis.

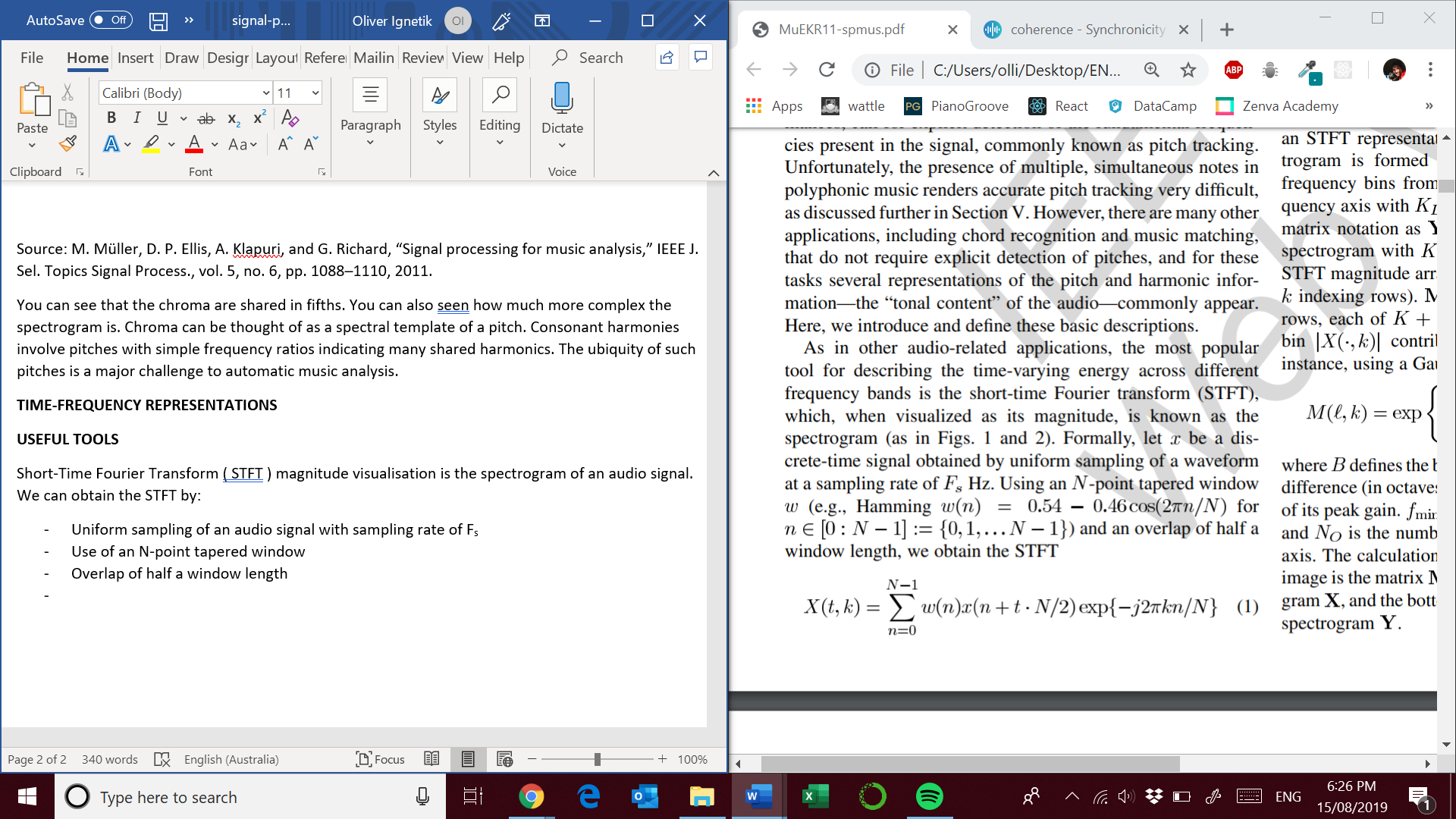
**TIME-FREQUENCY REPRESENTATIONS**

**USEFUL TOOLS**

**Short-Time Fourier Transform ( STFT ) – Spectrogram**

The magnitude visualisation is the spectrogram of an audio signal. We can obtain the STFT by:

* Uniform sampling of an audio signal with sampling rate of F­s
* Use of an N-point tapered window.
  + w[n] = 0.54 – 0.46cos(2\*pi\*n/N)
* Overlap of half a window length
  + How does this translate to hop length ?



* The time resolution is : *N/2Fs*
* The frequency resolution is : *Fs/N*
* Other important parameters include the number of frames *T*
* *K/2* is the last unique frequency value ( Nyquist Theorem )
* *X( t, k )* is complex-valued with the phase depending on the precise alignment of each short-time analysis window.

**Typical parameters**

* Fs = 44100 Hz
* N = 4096 samples is the window size

Magnitude only representations do not model phase interactions which can effect amplitude modulations.

**Log-Frequency Spectrogram**

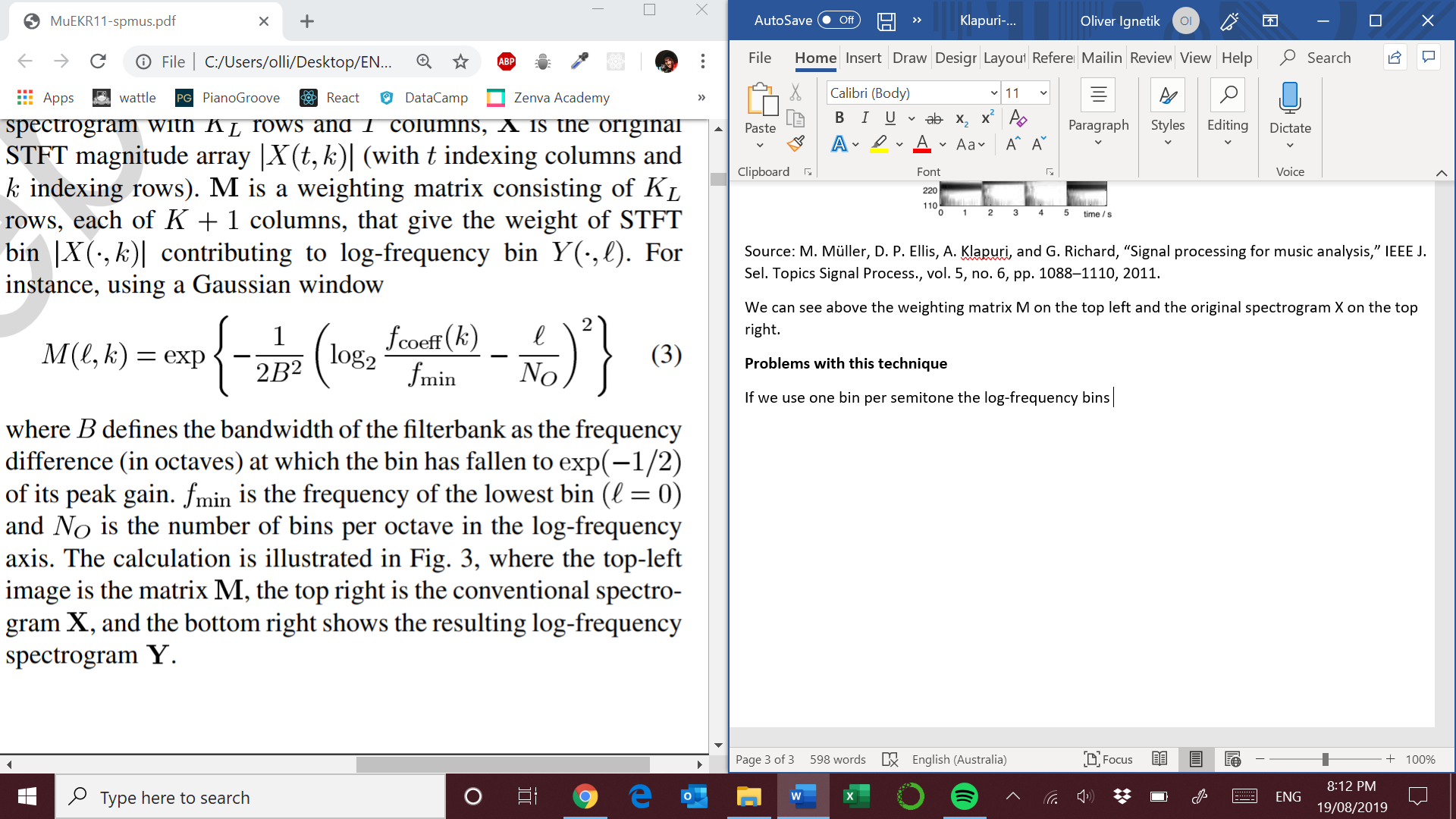
Since human perception of music defines a logarithmic frequency scale it makes sense to use the same scale in music analysis.

**Constant-Q transforms** – the bandwidth of each bin varies in proportion to its centre frequency. The Q-ratio centre frequency: bandwidth remains the same. For example, if we use 12 frequency bins per octave, each bin represents a semitone of an equal tempered scale.

We can transform a STFT representation using a weighting matrix that gives each bin of the STFT a coefficient contributing to a log-frequency bin.

**Y** = **MX**,

Where M is the weighting matrix X is the original STFT magnitude array and Y is the log-frequency spectrogram. Below is an example of a gaussian filterbank;

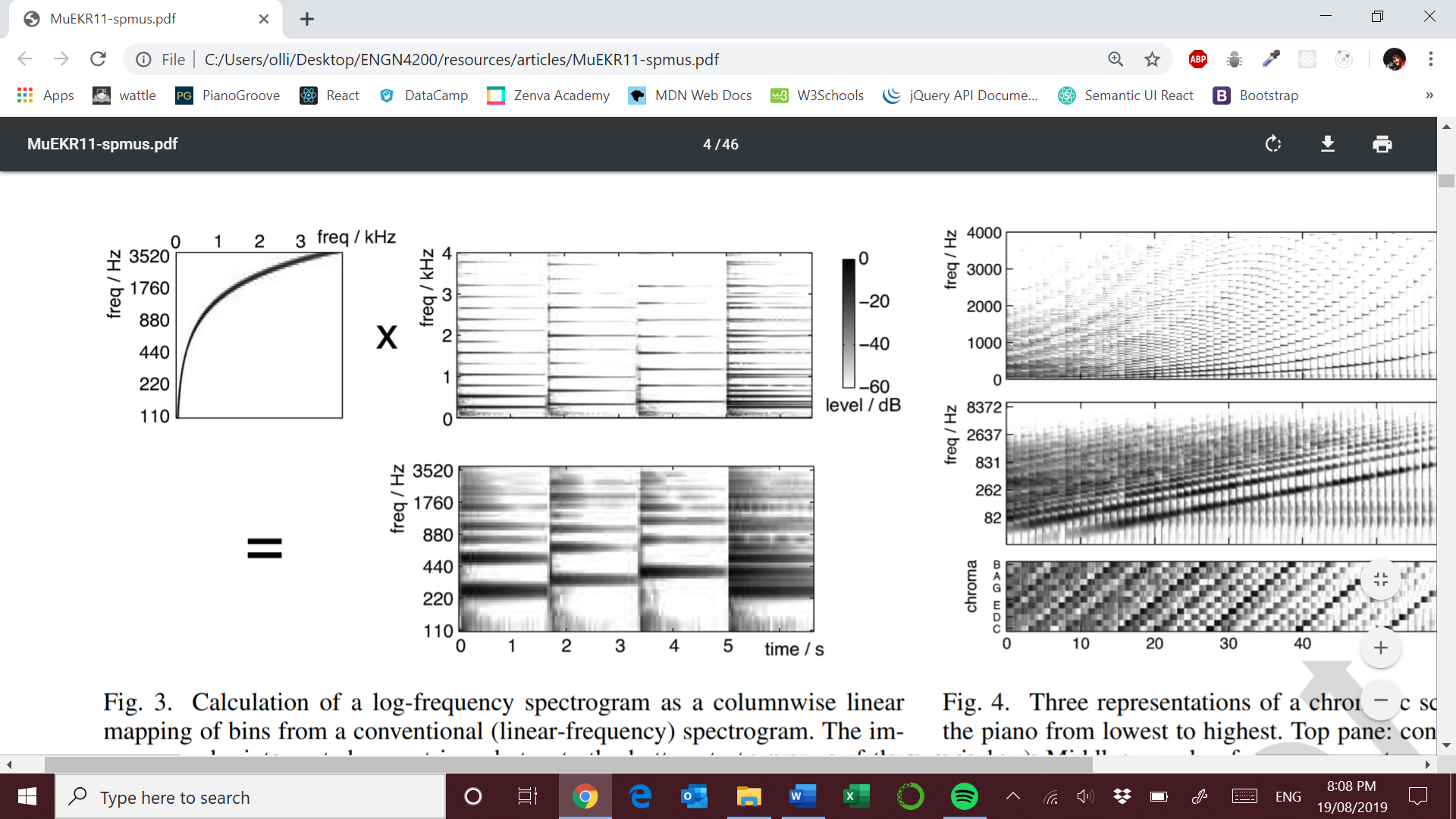


B = frequency difference ( in octaves ) at which the bin has fallen to exp(-1/2) of its peak gain.

fmin = frequency of the lowest bin ( l = 0 )

NO = number of bins per octave in the log-frequency axis

**VISUAL INTERPRETATION OF LOG FREQUENCY SPECTROGRAM**



Source: M. Müller, D. P. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011.

We can see above the weighting matrix M on the top left and the original spectrogram X on the top right.

**KEY FEATURE NEEDED – semitone resolution**

**Problems with the log frequency mapping**

* **Blurring of the frequency bins**: If we use one bin per semitone the log-frequency bins ie. NO = 12 with fmin  = 110 Hz ( A2 ) the centres of the log-frequency bins are only 6.5Hz apart. **???**
* To have these centred on distinct STFT bins would require a window of 153ms with a sampling rate of 44100 Hz.
* **SPECTRO-TEMPORAL TRADE OFF**: We need long time windows to achieve semitone resolution at low frequencies. This is a serious problem because human perception of rhythm can discriminate changes les then 10ms. **Thus we can lose information !**
* The high frequency bins average together many STFT bins!
* The low frequency bins are only 6.5 Hz apart!

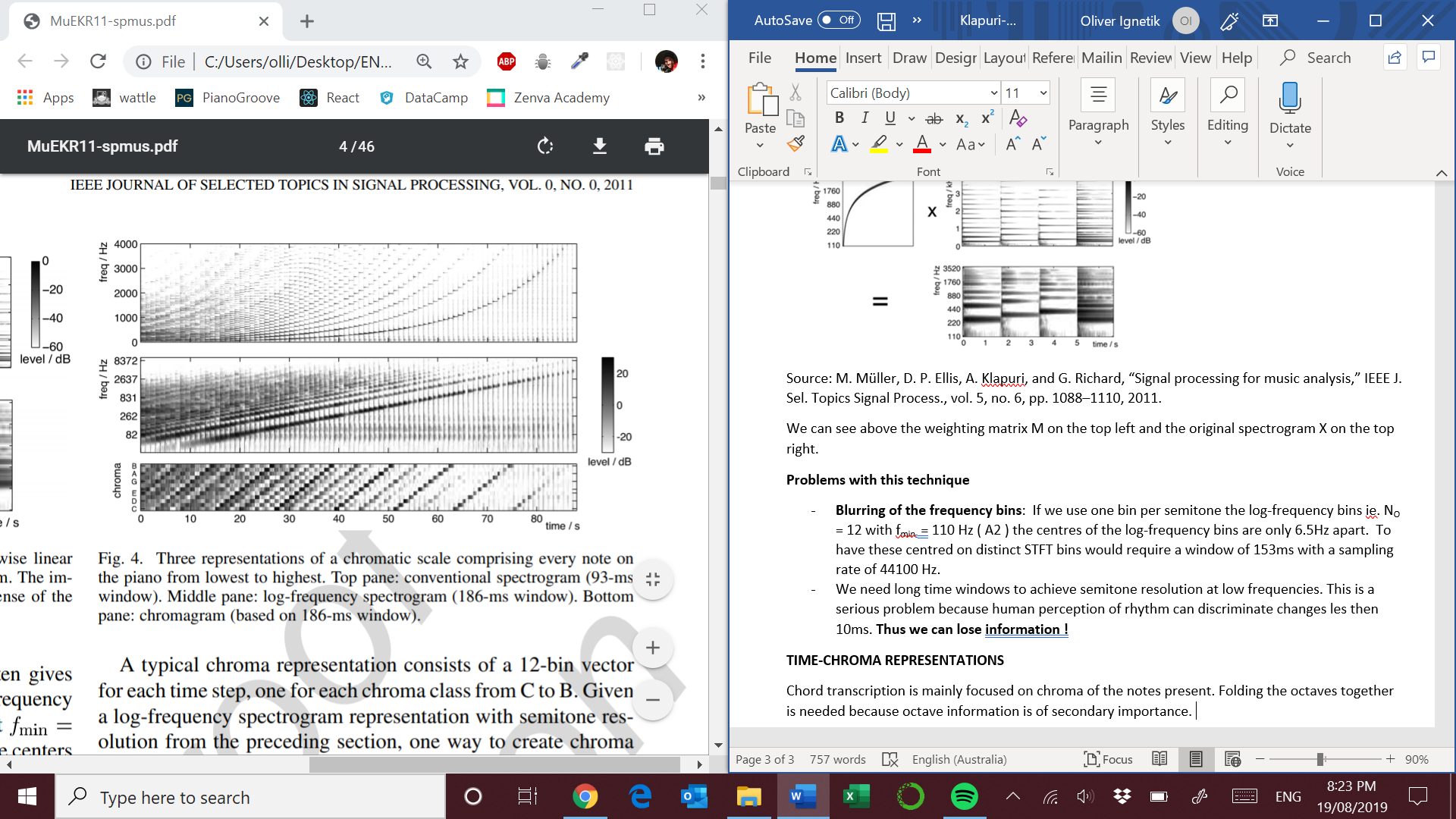
**Alternative to log frequency mapping – filter bank approach**

Construct a bank of individual bandpass filters, one per semitone, each tuned to the appropriate bandwidth with minimal temporal support.

Source: C. Schörkhuber and A. Klapuri, “Constant-Q transform toolbox for music processing,” in Proc. Sound Music Comput. Conf. (SMC), Barcelona, Pain, 2010.

**TIME-CHROMA REPRESENTATIONS**

Chord transcription is mainly focused on chroma of the notes present. Folding the octaves together is needed because octave information is of secondary importance. The chroma can be though of as a vector.



**How to construct a chromagram ?**

1. STFT with semitone resolution ie. Fres = 1.059 Hz , L = 41624 samples at Sr = 44.1kHz
2. Multiple approaches:
   1. Add together all the bins corresponding to each chroma [6]
   2. **Include energy only from strong sinusoidal components in the audio and exclude non-tonal energy such as percussion and noise**

Source: M. Müller, D. P. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011.

Many musical notes have the highest energy in the fundamental harmonic. With a weak fundamental the root chroma is the bin into which the greatest number of low-order harmonics fall into. Lower frequency ranges still tend to clutter the chromogram because they have energy dispersed across a large number of harmonics.

**Chroma applications**

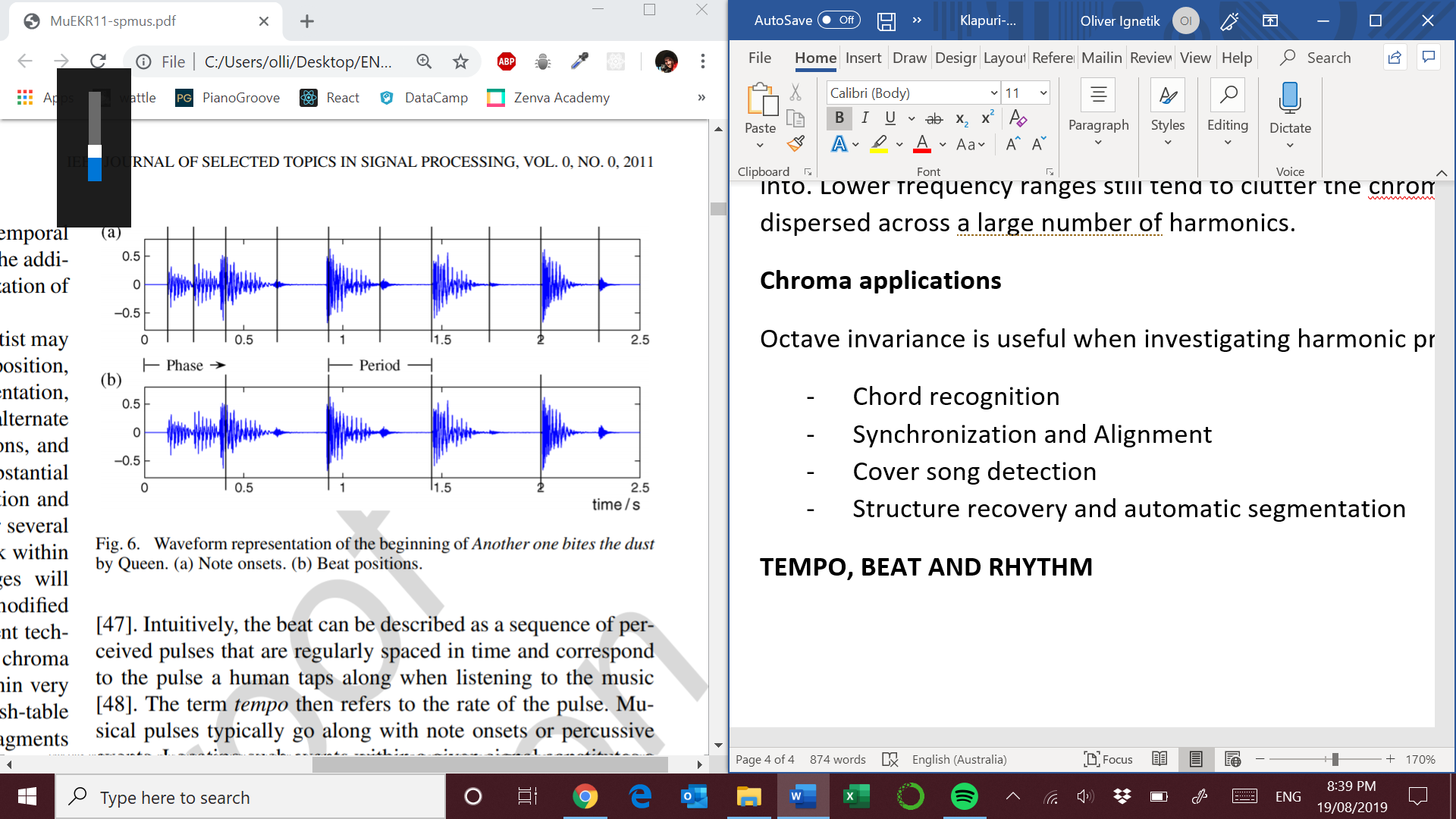
Octave invariance is useful when investigating harmonic progressions

* Chord recognition
* Synchronization and Alignment
* Cover song detection
* Structure recovery and automatic segmentation

**TEMPO, BEAT AND RHYTHM**

Tempo is the rate of the pulse of the music. Has several challenges:

* Onset detection
* Offset detection
* Temporal evolution of sound events



1. Note onsets
2. Beat positions

Source: M. Müller, D. P. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011.

**Onset detection**

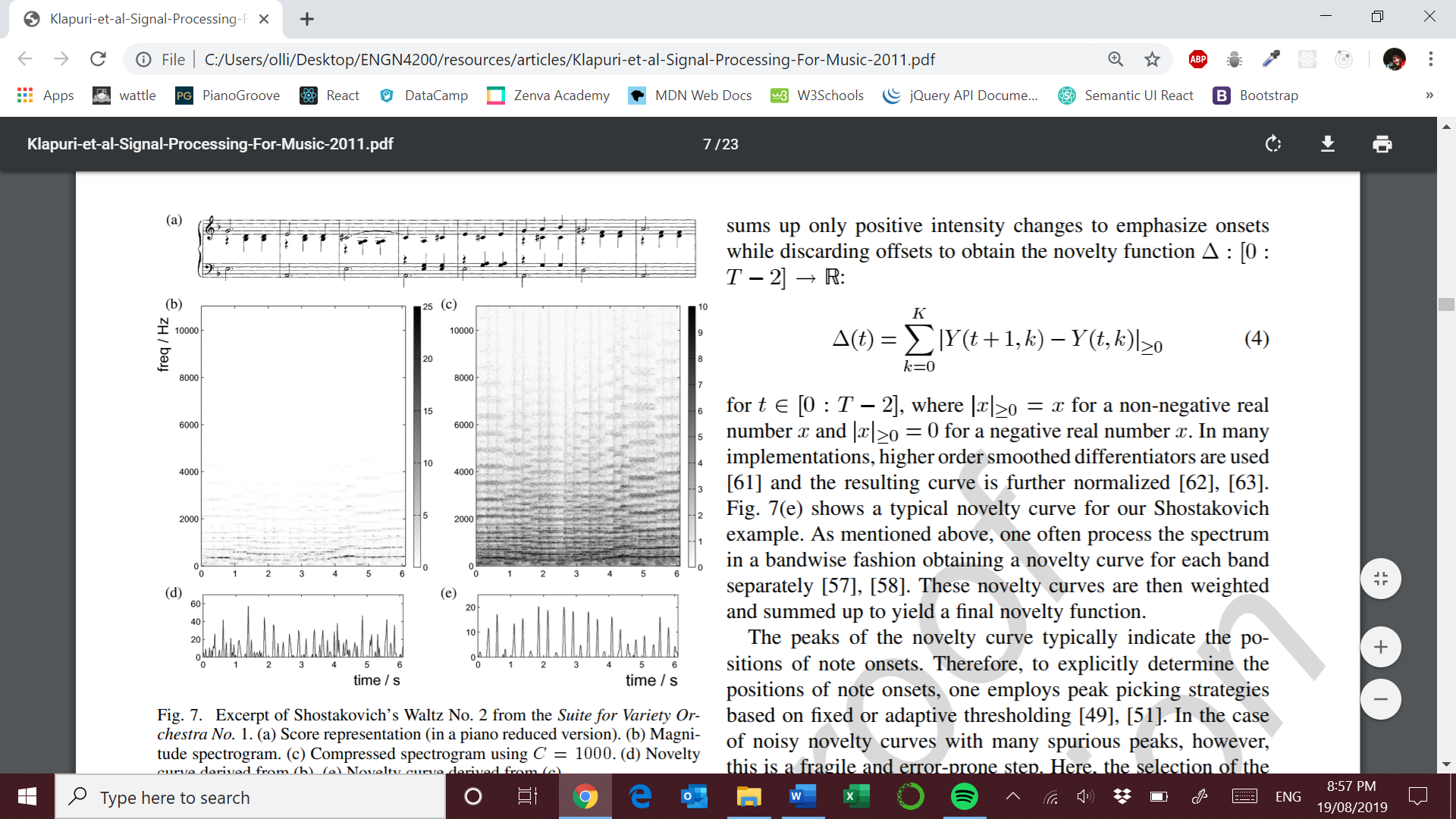
The general idea is to capture sudden changes in a music signal. By doing so you will obtain a *Novely Curve*. Non-percussive instruments have a number of challenges such as :

* Soft onset ( flutes etc. )
* Blurred note transitions ( sliding up to notes )
* Masking effects

Some methods for computing refined novelty curves (pg. 6)

* Spectral content analysis
* Pitch analysis
* Harmony and phase
* Instrument specific
* Resolution of masking effects with band wise analysis

**Spectral Flux**



1. Score representation
2. Magnitude spectrogram
3. Compressed spectrogram
4. Novelty curves b)
5. Novelty curve for c)

You can see in this novelty curve that there are blurred onsets in the first bar as the non-percussive instruments are playing the notes softly

Source: M. Müller, D. P. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011.

**Spectral based approach for computing novelty curves**

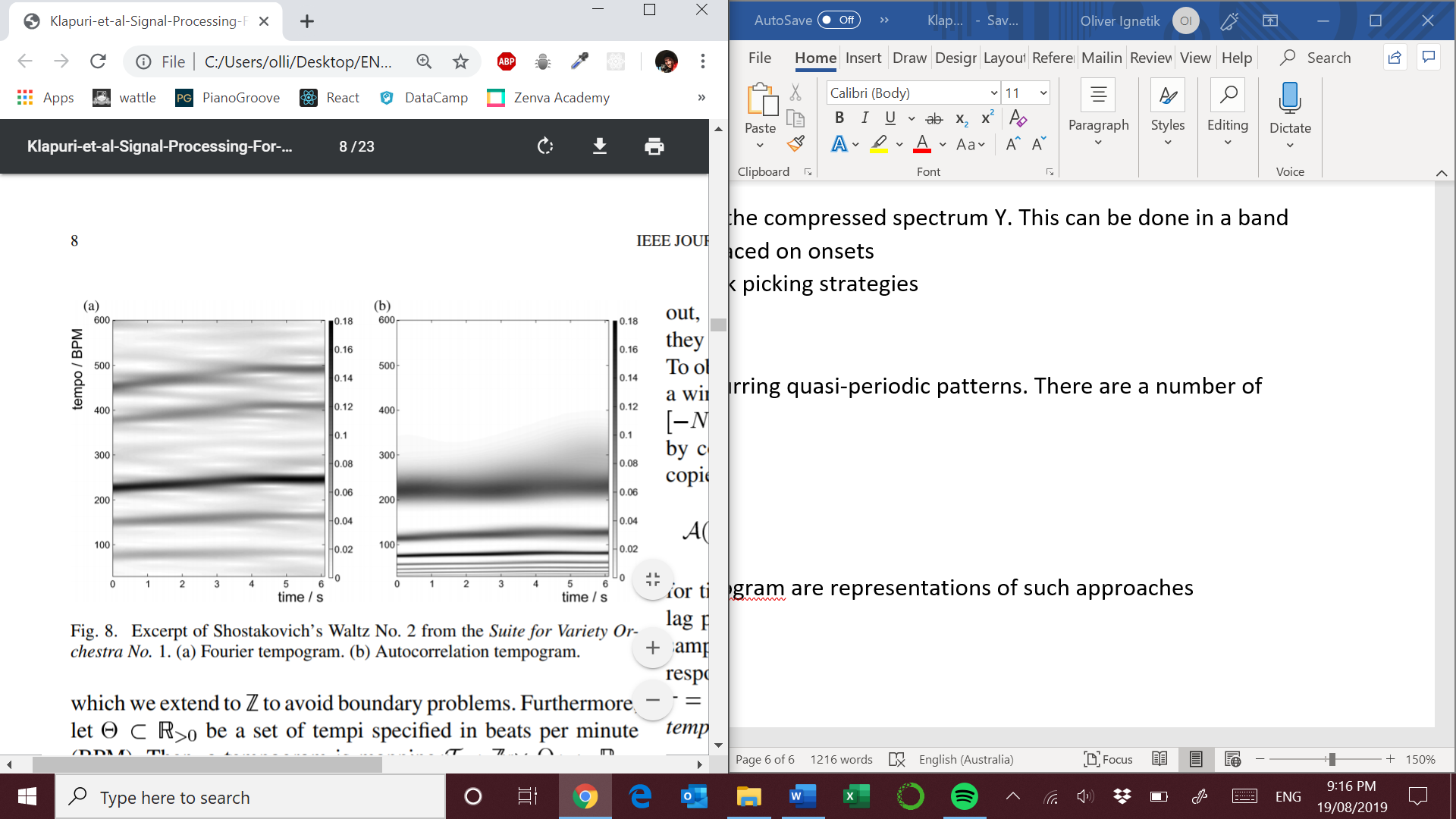
1. STFT of music recording
2. Binning strategies to switch over to a logarithmically spaced frequency axis
3. Logarithmic compression – apply a logarithm to the magnitude spectrogram of the signal ( corresponds to c) above
4. Compute the discrete derivative of the compressed spectrum Y. This can be done in a band wise fashion and the emphasis is placed on onsets
5. Use adaptive thresholding with peak picking strategies

**Periodicity and Tempo Estimation**

Novelty curves can be used to detect reoccurring quasi-periodic patterns. There are a number of methods used :

* Autocorrelation method
* Bank of filter resonators
* STFT novelty curve representation

**Beat spectrogram**, tempogram and rhythmogram are representations of such approaches



1. Fourier tempogram
2. Autocorrelation tempogram

Source: M. Müller, D. P. Ellis, A. Klapuri, and G. Richard, “Signal processing for music analysis,” IEEE J. Sel. Topics Signal Process., vol. 5, no. 6, pp. 1088–1110, 2011.

# References

Pg. 19 – 23 of pdf document