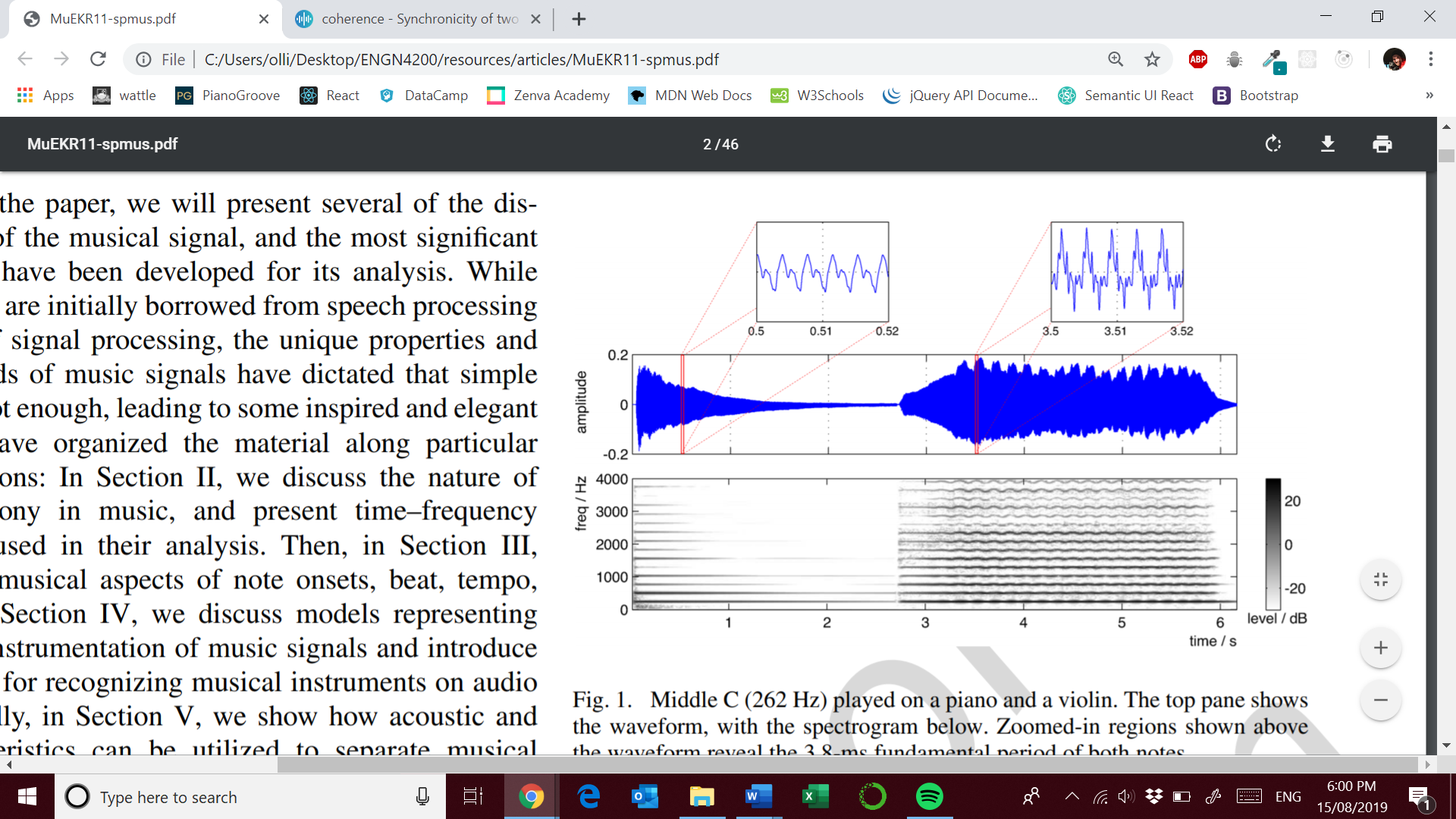
SIGNAL PROCESSING FOR MUSIC ANALYSIS NOTES

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.

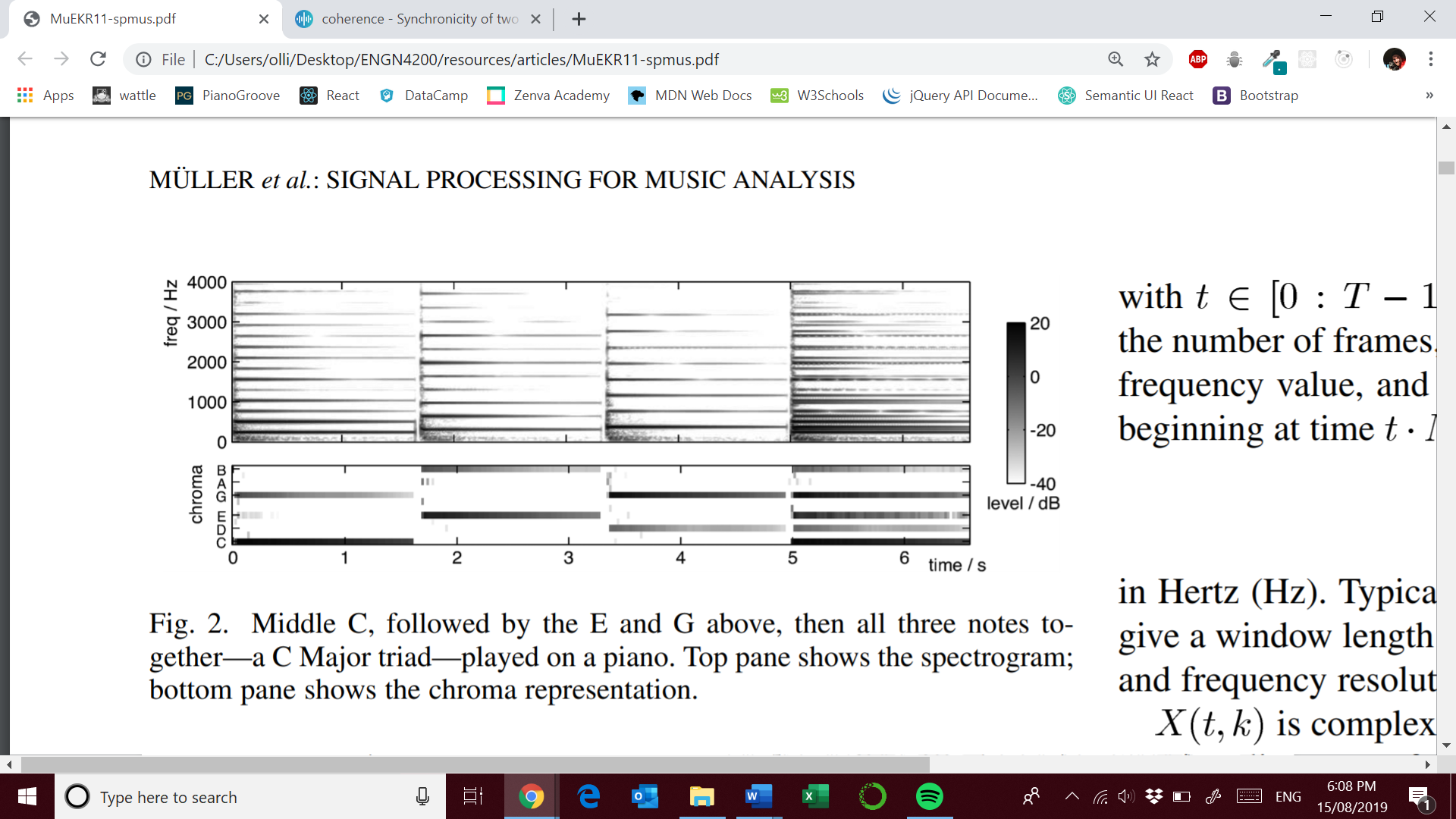
Humans perceive frequencies that fall in a ratio of 2:1 (Octave) as highly similar due to the embedded sets of harmonics.



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. Equal tempered scales allows an octave to be split into twelve equal steps on a logarithmic axis. Each note has a frequency 2^1/12 times larger ( semitone ).

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.

You can see that the chroma are shared in fifths. You can also seen 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 pitches is a major challenge to automatic music analysis.

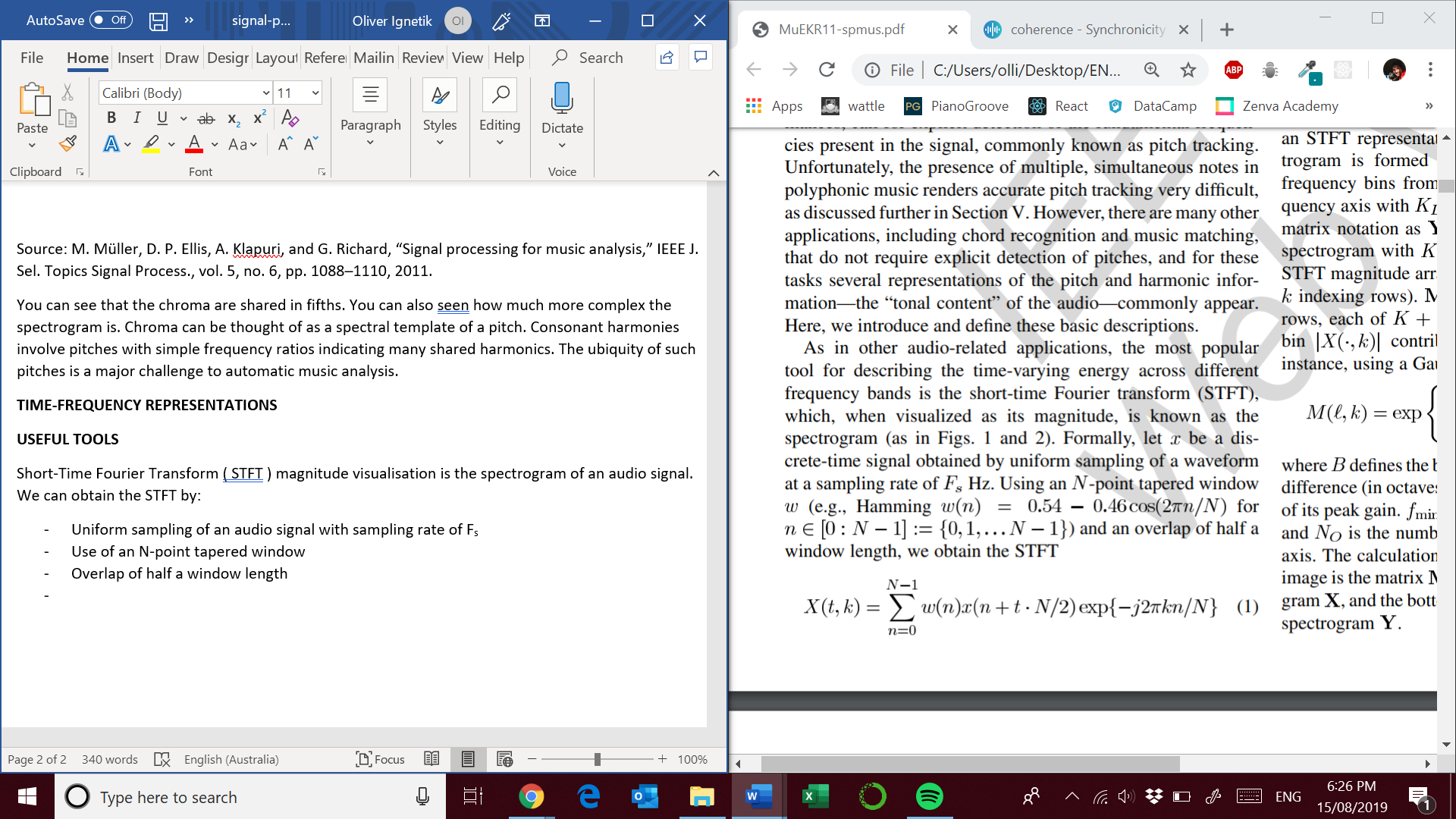
**TIME-FREQUENCY REPRESENTATIONS**

**USEFUL TOOLS**

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

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
* Overlap of half a window 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
* *X( t, k )* is complex-valued with the phase depending on the precise alignment of each short-time analysis window.

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 through the use of a weighting matrix that gives each bin of the STFT a coefficient contributing to a log-frequency bin.