Automatic Music Transcription Tutorial

ISMIR 2015 - Malaga, Spain October 26, 2015

# Useful References

Pgs. 53 of tutorial slides

**MIREX Multi-F0 Estimation and Note Tracking task**

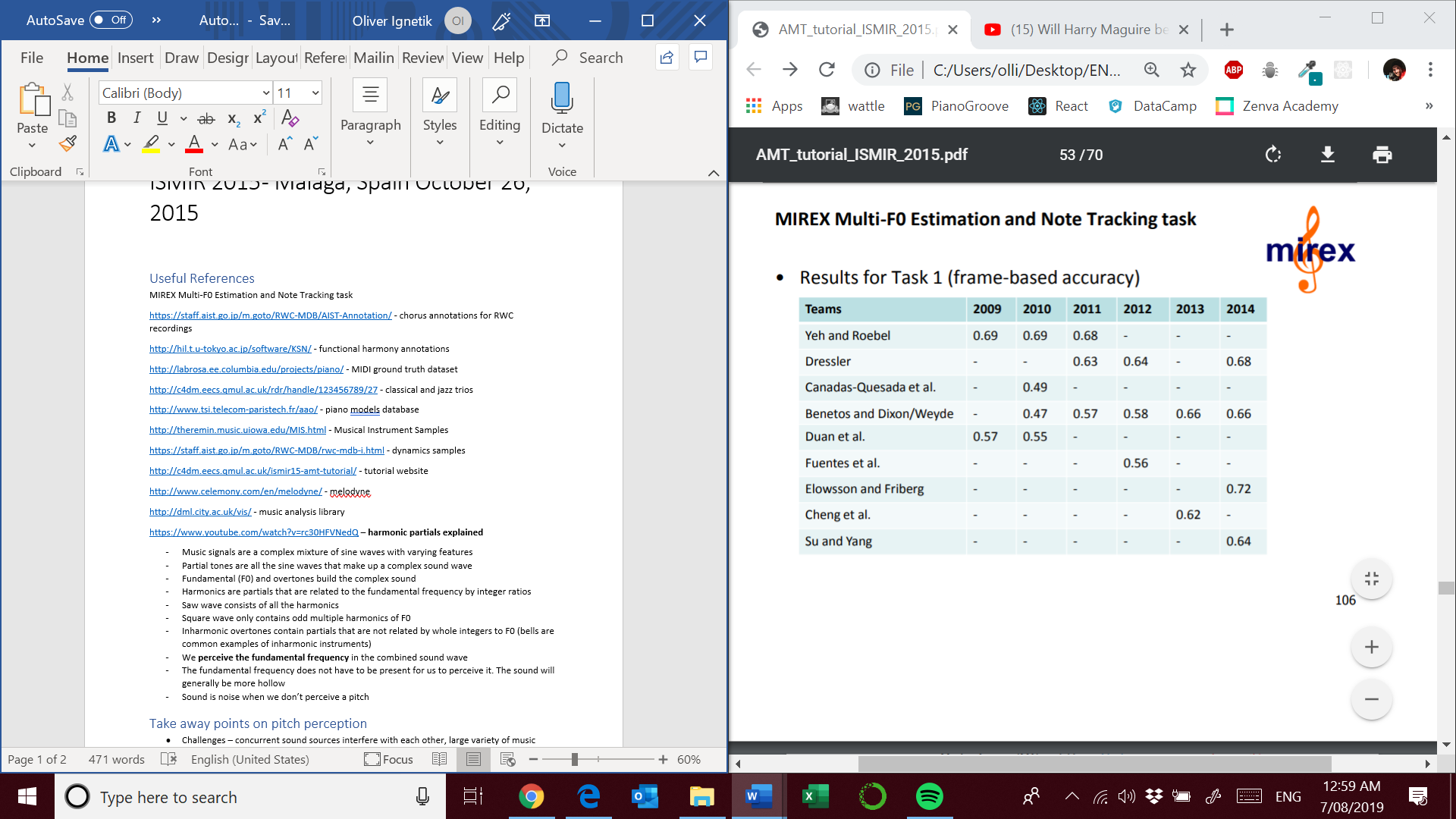
Task 1 : Frame based evaluation of multiple instruments

Task 2a: note based evaluation of multiple instruments

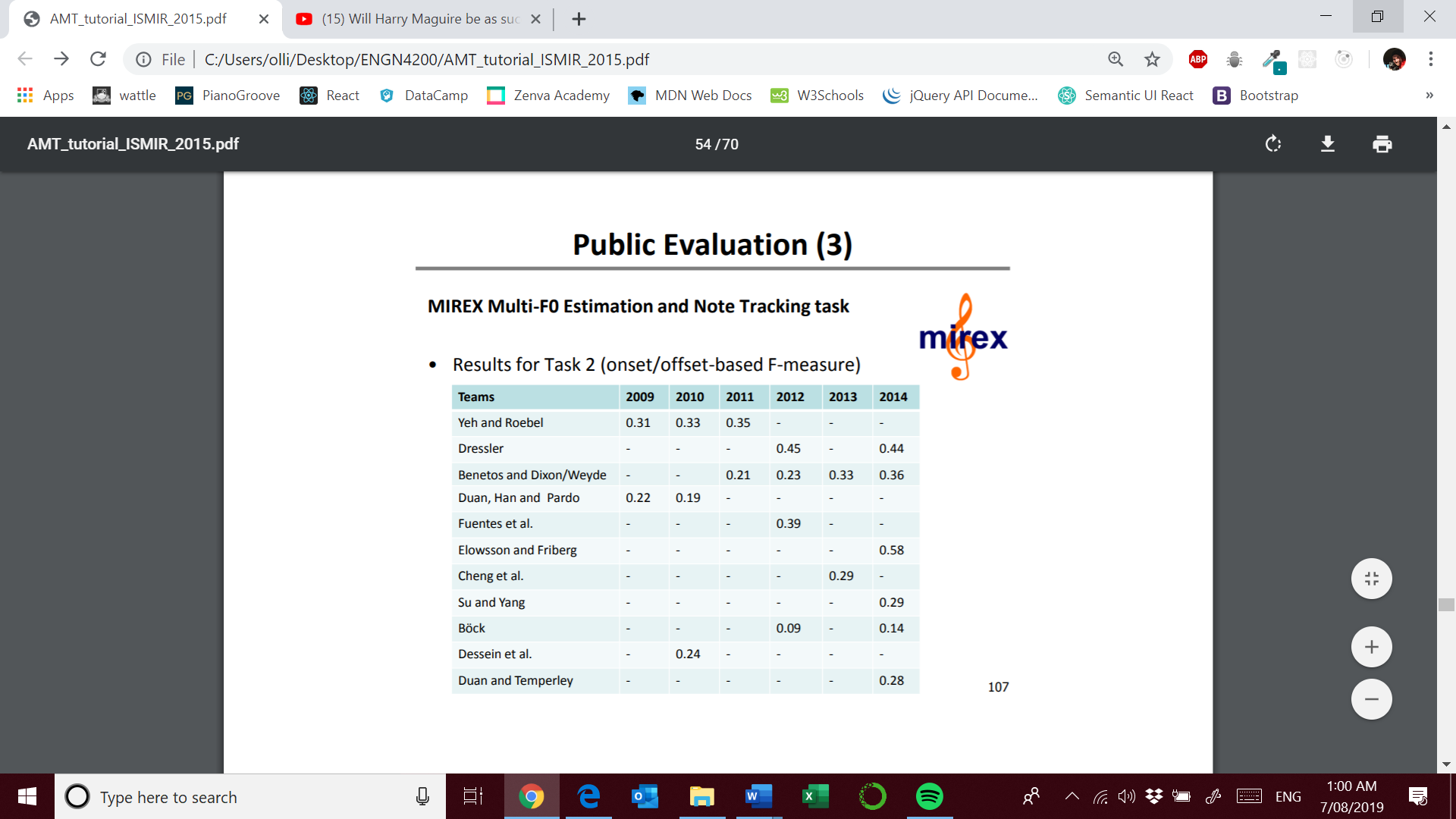
Task 2b: note based evaluation piano only

Task 3: Timbre tracking (not often run)

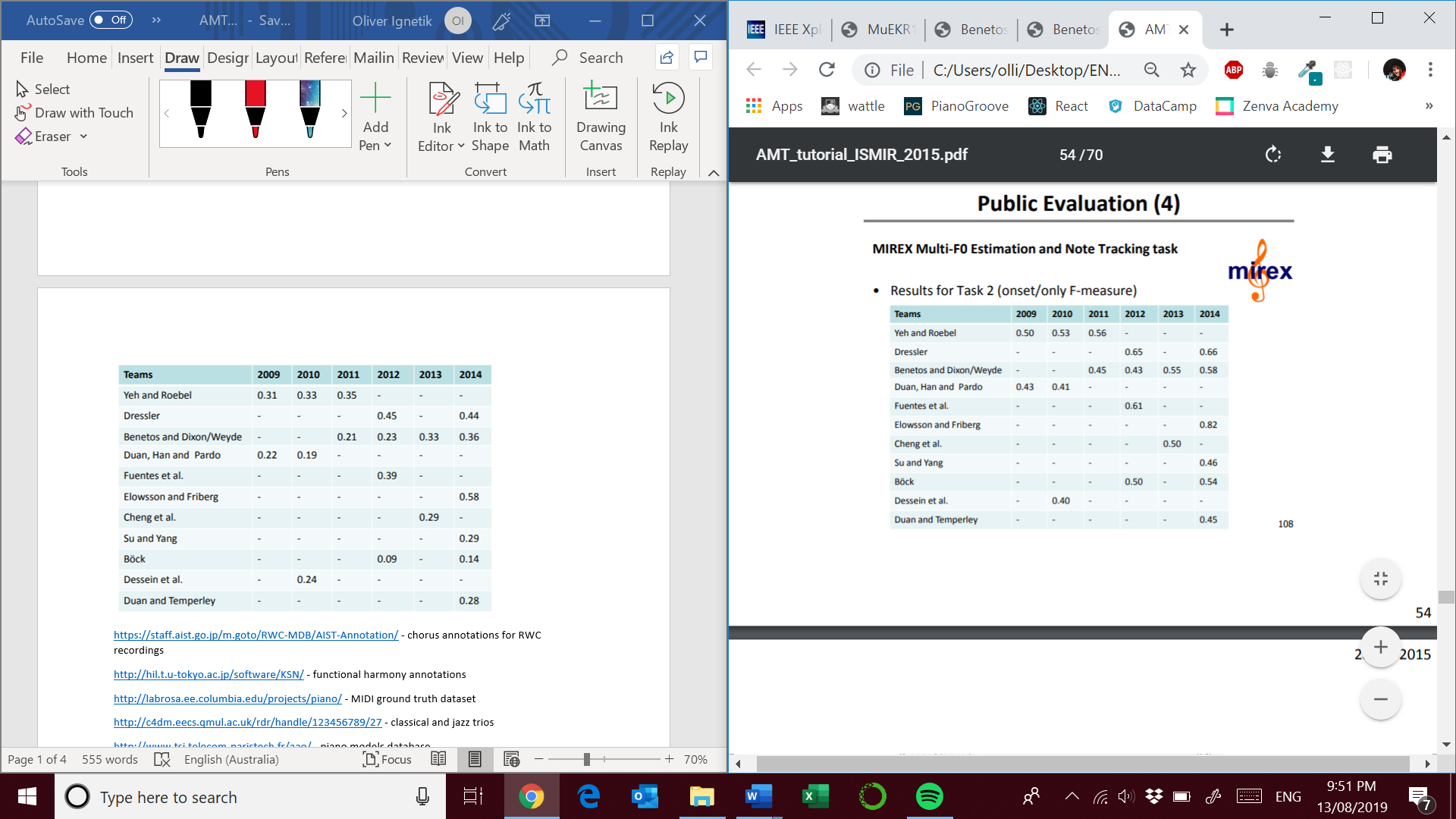
**Task 1 : Frame based accuracy**



**Task 2 : onset/offset-based-F-measure**



**Task 2 : onset/only F-measure**



# DATASETS

<https://staff.aist.go.jp/m.goto/RWC-MDB/AIST-Annotation/> - chorus annotations for RWC recordings

<http://hil.t.u-tokyo.ac.jp/software/KSN/> - functional harmony annotations

<http://labrosa.ee.columbia.edu/projects/piano/> - MIDI ground truth dataset

<http://c4dm.eecs.qmul.ac.uk/rdr/handle/123456789/27> - classical and jazz trios

<http://www.tsi.telecom-paristech.fr/aao/> - piano models database

<http://theremin.music.uiowa.edu/MIS.html> - Musical Instrument Samples

<https://staff.aist.go.jp/m.goto/RWC-MDB/rwc-mdb-i.html> - dynamics samples

<http://c4dm.eecs.qmul.ac.uk/ismir15-amt-tutorial/> - tutorial website

<http://www.celemony.com/en/melodyne/> - melodyne

<http://dml.city.ac.uk/vis/> - music analysis library

# Harmonics and Partials explained

<https://www.youtube.com/watch?v=rc30HFVNedQ> – **harmonic partials explained**

* Music signals are a complex mixture of sine waves with varying features
* Partial tones are all the sine waves that make up a complex sound wave
* Fundamental (F0) and overtones build the complex sound
* Harmonics are partials that are related to the fundamental frequency by integer ratios
* Saw wave consists of all the harmonics
* Square wave only contains odd multiple harmonics of F0
* Inharmonic overtones contain partials that are not related by whole integers to F0 (bells are common examples of inharmonic instruments)
* We **perceive the fundamental frequency** in the combined sound wave
* The fundamental frequency does not have to be present for us to perceive it. The sound will generally be more hollow
* Sound is noise when we don’t perceive a pitch

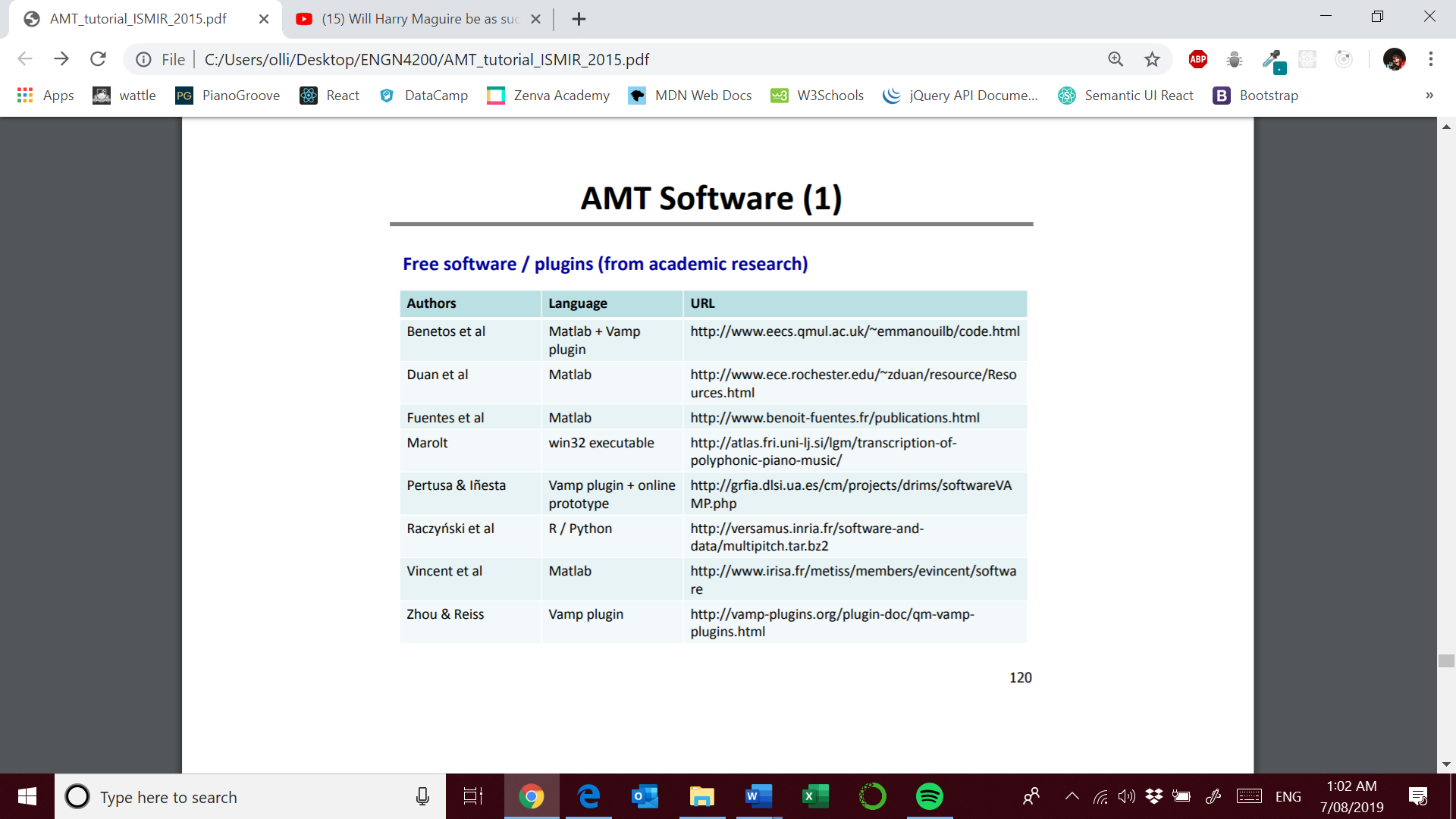
# Take away points on pitch perception

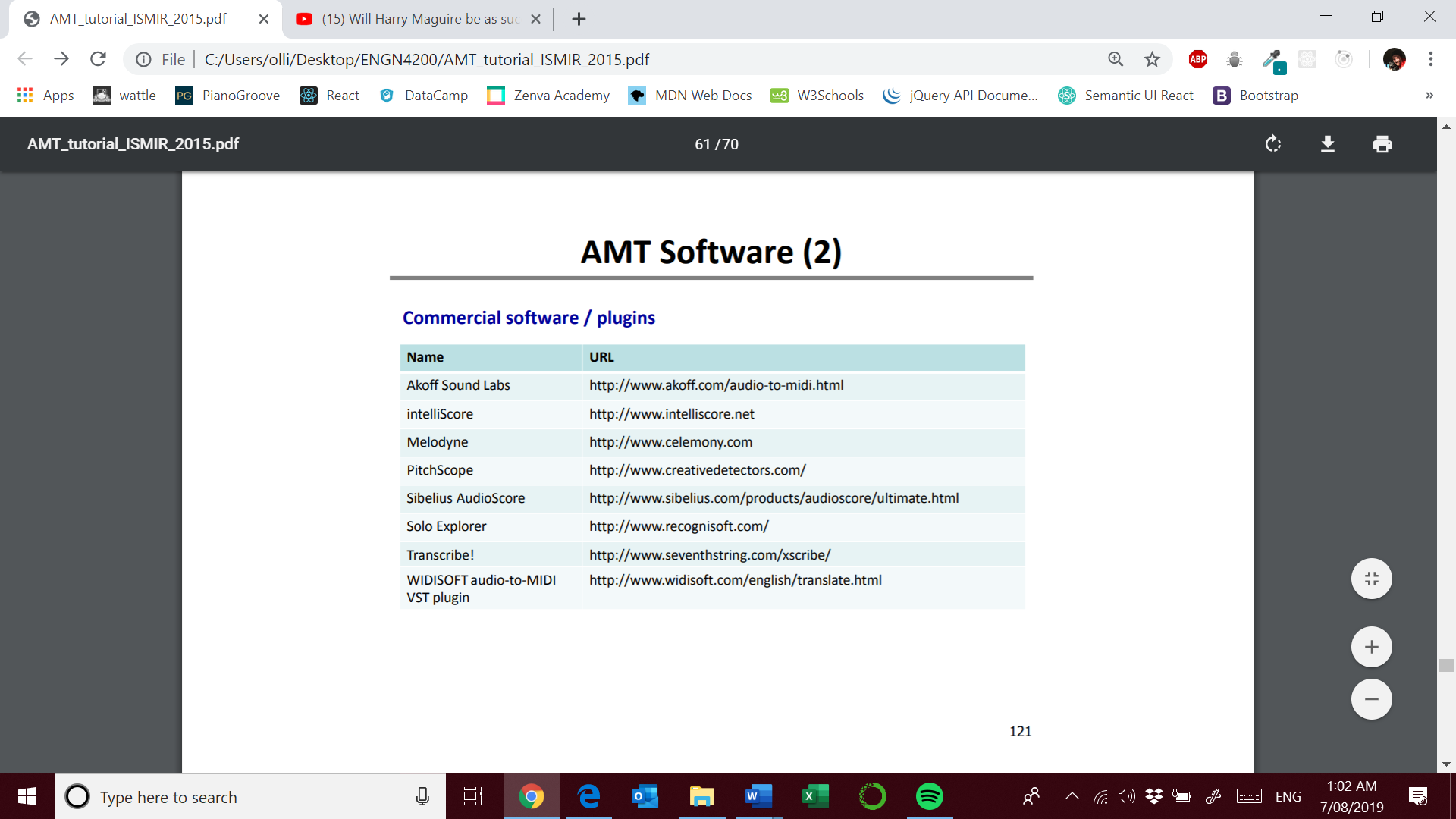
* Challenges – concurrent sound sources interfere with each other, large variety of music
* Performance is clearly far below of a human expert (especially in polyphonic music)
* No unified methodology
* Fundamental frequency (F0) – doesn’t have to be present for a pitch to be perceived, partials can suggest a fundamental frequency. For sounds to be perceived as a musical tone, its spectrum must consist of 3 successive harmonics [de Cheveigné, 2006; Houtsma, 1995]
* Pitch perception – harmonics make tones more pleasant but confuse pitch perception
* Pitch Perception theories – pattern matching and autocorrelation models
* How people transcribe – rough sketch -> harmony -> melody

# Take away points on methods

* MPS is the least explored level of AMT
* Instruments will play notes that are harmonically similar but may also play the same notes !
* MPE
  + Representation – STFT/CQT spectrum, ERB filterbank, specmurt,spectral peaks
  + Core algorithm – maximum likelihood, Bayesian, spectrogram decomposition, sparse coding, classification based
  + Iterative vs joint estimation methods
  + Time domain methods – Detailed simulation of human auditory system, adaptive oscillators (not adaptive) , probabilistic modelling (computationally expensive)
  + Frequency domain methods – iterative spectral subtraction (hard to subtract the right amount of energy), iterative bispectral subtraction, Spectral Peak modeling (difficulty with multiple sources of different loudness), spectral peak modeling ML (soft notes get masked),
  + Spectrogram Decomposition – Find the activation matrix if the note dictionary is known such that DA = V ( models perform poorly if test audio doesn’t match the dictionary). Convolutive models can model frequency modulations.
  + Classification based methods – View music transcription as multi-label classification, HMM (computationally expensive)

# SOFTWARE





# Future work

* Most existing transcription approaches are data-driven (bottom-up) which cause many errors that are not musically meaningful
* [Raczynski et al., 2013] used dynamic Bayesian networks to model higher level musical theory knowledge

Pg. 63 Musical Knowledge Incorporation