Pitch Tracking

Sevag Hanssian

MUMT 621, Winter 2021

March 23, 2021

2021-03-14

Pitch Tracking

Pitch Tracking

Sevag Hanssian

MUMT 621, Winter 2021

March 23, 2021

Pitch as a perceptual phenomenon

Pitch is the perceptual correlate of frequency, and the aspect of auditory sensation whose variation is associated with musical melodies.[1]

Pitch of a pure tone is its frequency; pitch of a complex tone is its lowest (or fundamental) frequency.

Pitch can be quantified using fundamental frequency (or *f0*); interchangeable terms outside psychoacoustical studies.[2]

1. C.J. Plack. 2013. *The Sense of Hearing* [in English]. 2nd ed. 117–118. United Kingdom: Psychology Press Ltd.
2. Jong Wook Kim et al. 2018. *CREPE: A Convolutional Representation for Pitch Estimation*. arXiv: 1802.06182 [eess.AS]. https://arxiv.org/pdf/1802.06182.pdf.

Pitch as a perceptual phenomenon

Pitch is the perceptual correlate of frequency, and the aspect of auditory sensation whose variation is associated with musical melodies.[1]

Pitch of a pure tone is its frequency; pitch of a complex tone is its lowest (or fundamental) frequency.

Pitch can be quantified using fundamental frequency (or *f0*); interchangeable terms outside psychoacoustical studies.[2]

1. C.J. Plack. 2013. *The Sense of Hearing* [in English]. 2nd ed. 117–118. United Kingdom: Psychology Press Ltd.
2. Jong Wook Kim et al. 2018. *CREPE: A Convolutional Representation for Pitch Estimation*. arXiv: 1802.06182 [eess.AS]. https://arxiv.org/pdf/1802.06182.pdf.

- it's a bit more complicated. there's also a relationship between loudness and perceived pitch previous bad definition: the attribute of auditory sensation in terms of which sounds may be ordered on a scale extending from low to high
  It requires the words "low" and "high" to be associated with pitch or frequency rather than with loudness or intensity, for example.

Pitch tracking algorithms

- Candidate-generating function with pre- and post-processing to produce the pitch curve[3]
  - Cepstrum
  - Autocorrelation function
  - Average magnitude/square difference function (AMDF, ASDF)
  - Normalized cross-correlation function (RAPT, PRAAT)
  - Cumulative mean normalized difference (YIN)
- SWIPE: template matching with spectrum of sawtooth waveform
- pYIN: probabilistic YIN with Hidden Markov Models (HMM) to decode most probably pitch sequences (best performing[4],[5])
- WavePitch: based on the fast lifting wavelet transform (FLWT)

---

3. Kim et al. 2018.

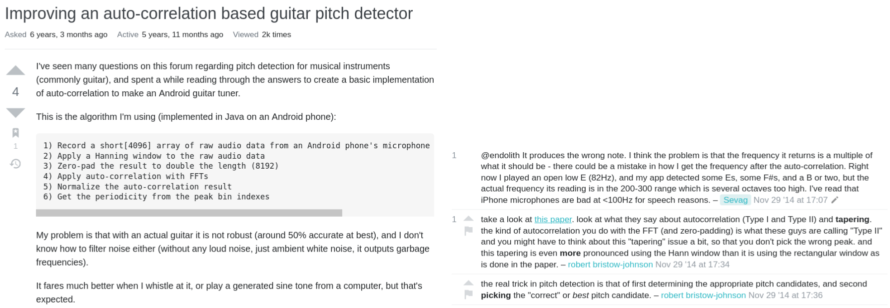
4. Onur Babacan et al. 2013. "A comparative study of pitch extraction algorithms on a large variety of singing sounds," 7815–7819. May. https://doi.org/10.1109/ICASSP.2013.6639185.

5. Adrian von dem Knesebeck and Udo Zölzer. 2010. "Comparison of pitch trackers for real-time guitar effects," 266–269. September.

Pitch tracking algorithms
- Candidate-generating function with pre- and post-processing to produce the pitch curve[3]
  - Cepstrum
  - Autocorrelation function
  - Average magnitude/square difference function (AMDF, ASDF)
  - Normalized cross-correlation function (RAPT, PRAAT)
  - Cumulative mean normalized difference (YIN)
- SWIPE: template matching with spectrum of sawtooth waveform
- pYIN: probabilistic YIN with Hidden Markov Models (HMM) to decode most probably pitch sequences (best performing[4],[5])
- WavePitch: based on the fast lifting wavelet transform (FLWT)
3. Kim et al. 2018.
4. Onur Babacan et al. 2013. "A comparative study of pitch extraction algorithms on a large variety of singing sounds," 7815–7819. May. https://doi.org/10.1109/ICASSP.2013.6639185.
5. Adrian von dem Knesebeck and Udo Zölzer. 2010. "Comparison of pitch trackers for real-time guitar effects," 266–269. September.

- YIN = yin and yang. the interplay of autocorrelation and cancellation

Autocorrelation



Improving an auto-correlation based guitar pitch detector

Asked 6 years, 3 months ago Active 5 years, 11 months ago Viewed 2k times

4

I've seen many questions on this forum regarding pitch detection for musical instruments (commonly guitar), and spent a while reading through the answers to create a basic implementation of auto-correlation to make an Android guitar tuner.

This is the algorithm I'm using (implemented in Java on an Android phone):

```
1) Record a short[4096] array of raw audio data from an Android phone's microphone
2) Apply a Hanning window to the raw audio data
3) Zero-pad the result to double the length (8192)
4) Apply auto-correlation with FFTs
5) Normalize the auto-correlation result
6) Get the periodicity from the peak bin indexes
```

My problem is that with an actual guitar it is not robust (around 50% accurate at best), and I don't know how to filter noise either (without any loud noise, just ambient white noise, it outputs garbage frequencies).

It fares much better when I whistle at it, or play a generated sine tone from a computer, but that's expected.

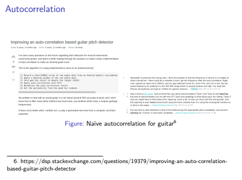
1 @endolith It produces the wrong note. I think the problem is that the frequency it returns is a multiple of what it should be - there could be a mistake in how I get the frequency after the auto-correlation. Right now I played an open low E (82Hz), and my app detected some Es, some F#s, and a B or two, but the actual frequency its reading is in the 200-300 range which is several octaves too high. I've read that iPhone microphones are bad at <100Hz for speech reasons. – Sevag Nov 29 '14 at 17:07

1 take a look at this paper. look at what they say about autocorrelation (Type I and Type II) and **tapering**. the kind of autocorrelation you do with the FFT (and zero-padding) is what these guys are calling "Type II" and you might have to think about this "tapering" issue a bit, so that you don't pick the wrong peak. and this tapering is even **more** pronounced using the Hann window than it is using the rectangular window as is done in the paper. – robert bristow-johnson Nov 29 '14 at 17:34

the real trick in pitch detection is that of first determining the appropriate pitch candidates, and second **picking** the "correct" or *best* pitch candidate. – robert bristow-johnson Nov 29 '14 at 17:36

Figure: Naïve autocorrelation for guitar[6]

6. https://dsp.stackexchange.com/questions/19379/improving-an-auto-correlation-based-guitar-pitch-detector



Autocorrelation

Autocorrelation: measure signal self-similarity by multiplying it with lagged copies of itself

Figure: Animation of sine wave autocorrelation[7]

7. http://qingkaikong.blogspot.com/2017/01/signal-processing-how-autocorrelation.html

Autocorrelation
Autocorrelation: measure signal self-similarity by multiplying it with lagged copies of itself
Figure: Animation of sine wave autocorrelation[7]
7. http://qingkaikong.blogspot.com/2017/01/signal-processing-how-autocorrelation.html

- YIN = yin and yang. the interplay of autocorrelation and cancellation

Autocorrelation – peak picking

Peak picking: measure signal periodicity by measuring inter-peak distance on autocorrelation

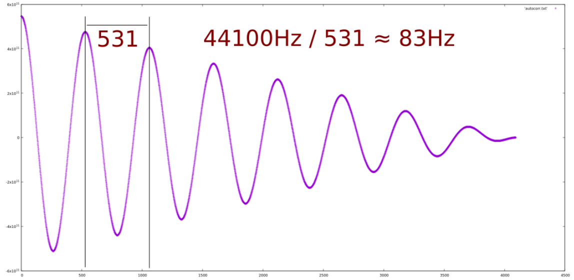


Figure: Autocorrelation peaks, example: 83Hz sine wave[8]

Peaks: [529, 1060, 1591, 2122, 2652, 3181, 3702]
Lag increment: ∼531 samples
Convert lag to pitch: 44100 Hz (sample rate) divided by 531 gives ∼83Hz, which is the frequency of the sine wave.

8. https://github.com/sevagh/pitch-detection/tree/master/misc/mcleod

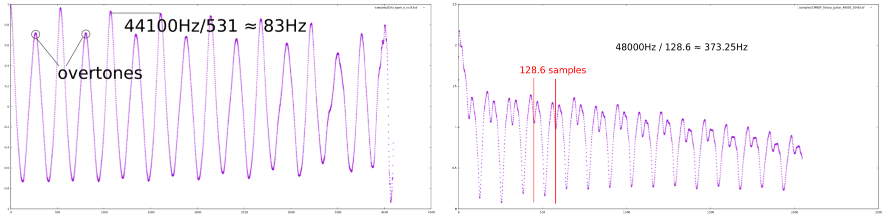
2021-03-14

Pitch Tracking

└─Autocorrelation – peak picking

Autocorrelation – peak picking
Peak picking: measure signal periodicity by measuring inter-peak distance on autocorrelation
531 44100Hz / 531 ≈ 83Hz
Figure: Autocorrelation peaks, example: 83Hz sine wave[8]
Peaks: [529, 1060, 1591, 2122, 2652, 3181, 3702]
Lag increment: ∼531 samples
Convert lag to pitch: 44100 Hz (sample rate) divided by 531 gives ∼83Hz, which is the frequency of the sine wave.
8. https://github.com/sevagh/pitch-detection/tree/master/misc/mcleod

YIN and McLeod Pitch Method – better autocorrelation



(a) MPM's[9] normalized square difference function

(b) YIN's[10] cumulative mean normalized difference function

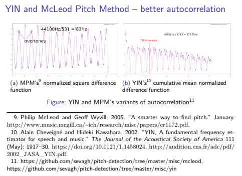
Figure: YIN and MPM's variants of autocorrelation[11]

---

9. Philip McLeod and Geoff Wyvill. 2005. "A smarter way to find pitch." January. http://www.music.mcgill.ca/~ich/research/misc/papers/cr1172.pdf.

10. Alain Cheveigné and Hideki Kawahara. 2002. "YIN, A fundamental frequency estimator for speech and music." *The Journal of the Acoustical Society of America* 111 (May): 1917–30. https://doi.org/10.1121/1.1458024. http://audition.ens.fr/adc/pdf/2002_JASA_YIN.pdf.

11. https://github.com/sevagh/pitch-detection/tree/master/misc/mcleod, https://github.com/sevagh/pitch-detection/tree/master/misc/yin



- Both functions related to autocorrelation, with the main correction that both discard the 0th lag – signal max self-similarity at no lag.

YIN and McLeod Pitch Method – better peak picking

Better peak picking: parabolic interpolation in both

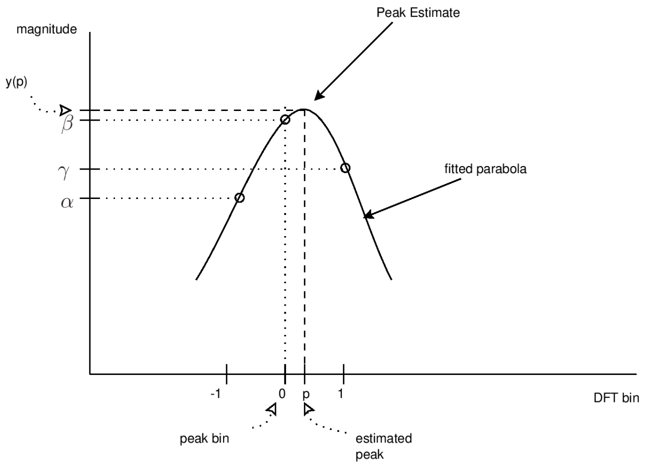
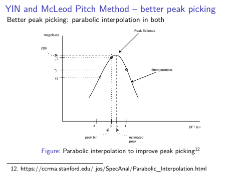
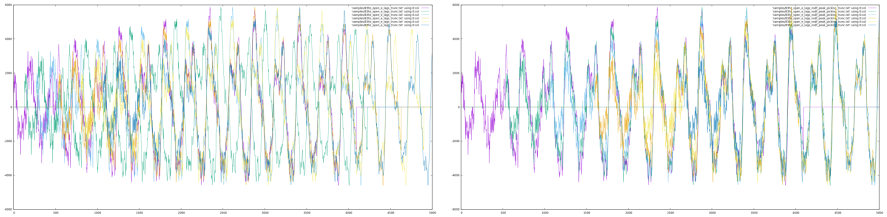


Figure: Parabolic interpolation to improve peak picking[12]

12. https://ccrma.stanford.edu/ jos/SpecAnal/Parabolic_Interpolation.html



Autocorrelation vs. MPM

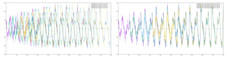


(a) Autocorrelation (b) MPM

Figure: Guitar signal superimposed on itself at peak lags[13]

13. https://github.com/sevagh/pitch-detection/tree/master/misc/mcleod

Autocorrelation vs. MPM



(a) Autocorrelation (b) MPM
Figure: Guitar signal superimposed on itself at peak lags[13]
13. https://github.com/sevagh/pitch-detection/tree/master/misc/mcleod

Musical importance of pitch

Pitch is musically important to humans:[14]

- Relationships between pitches (relative pitch) are more important than the absolute value. Sequence of pitch changes is the melodic contour
- Pitches separated by an octave have the same pitch chroma. Most (known) music depends on pitch relations defined by octaves
- Most (known) music in the world come from a discrete set of five to seven pitches arranged within an octave range

14. Josh McDermott and Marc Hauser. 2005. "The origins of music: Innateness, uniqueness, and evolution." *Music Perception - MUSIC PERCEPT* 23 (September): 29–59. https://doi.org/10.1525/mp.2005.23.1.29. https://web.mit.edu/jhm/www/Pubs/McDermott_2005_music_evolution.pdf.

Musical importance of pitch

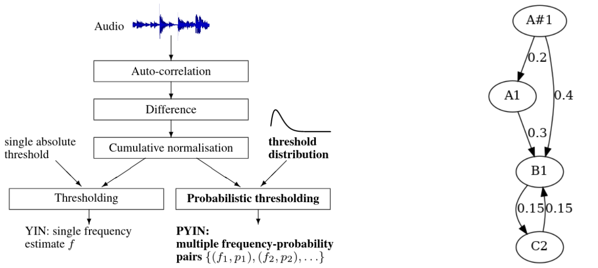
Pitch is musically important to humans:[14]

- Relationships between pitches (relative pitch) are more important than the absolute value. Sequence of pitch changes is the melodic contour
- Pitches separated by an octave have the same pitch chroma. Most (known) music depends on pitch relations defined by octaves
- Most (known) music in the world come from a discrete set of five to seven pitches arranged within an octave range

14. Josh McDermott and Marc Hauser. 2005. "The origins of music: Innateness, uniqueness, and evolution." *Music Perception - MUSIC PERCEPT* 23 (September): 29–59. https://doi.org/10.1525/mp.2005.23.1.29. https://web.mit.edu/jhm/www/Pubs/McDermott_2005_music_evolution.pdf.

- Although rhythm is arguably just as important, if not more so, to many cultures' music, pitch has received far more attention in the literature we will review, likely due to its importance in Western music and the resultant theoretical ideas about how pitch functions in music.
- Pitch is universal, but the centrality of relative pitch indicates some innate auditory mechanism for encoding sound as pitch distances. Humans are good at "melodic contour", i.e. recognizing a melody whether it is transposed in tempo or octave.
- Limitations of memory and categorization, or sensory or computational bias to have intervals that approximate simple integer ratios

pYIN – musically probable pitch sequences

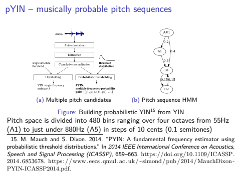


(a) Multiple pitch candidates (b) Pitch sequence HMM

Figure: Building probabilistic YIN[15] from YIN

Pitch space is divided into 480 bins ranging over four octaves from 55Hz (A1) to just under 880Hz (A5) in steps of 10 cents (0.1 semitones)

15. M. Mauch and S. Dixon. 2014. "PYIN: A fundamental frequency estimator using probabilistic threshold distributions." In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 659–663. https://doi.org/10.1109/ICASSP.2014.6853678. https://www.eecs.qmul.ac.uk/~simond/pub/2014/MauchDixon-PYIN-ICASSP2014.pdf.



- the pitch candidates are used as observation probabilities. the transition probabilities are used to come up with the final pitch estimate

CREPE – current state-of-the-art

Best performing techniques such as the pYIN algorithm, are based on a combination of DSP pipelines and heuristics. [...] we propose a data-driven pitch tracking algorithm, CREPE, which is based on a deep convolutional neural network that operates directly on the time-domain waveform.[16]

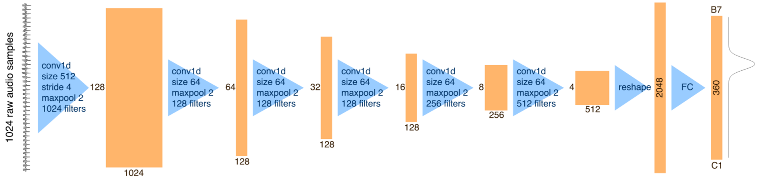


Figure: CREPE network architecture

360 pitch values are are selected so that they cover six octaves with 20-cent intervals between C1 and B7, corresponding to 32.70 Hz and 1975.5 Hz.

16. Kim et al. 2018.

CREPE – current state-of-the-art

Best performing techniques such as the pYIN algorithm, are based on a combination of DSP pipelines and heuristics. [...] we propose a data-driven pitch tracking algorithm, CREPE, which is based on a deep convolutional neural network that operates directly on the time-domain waveform.[16]

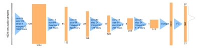


Figure: CREPE network architecture

360 pitch values are are selected so that they cover six octaves with 20-cent intervals between C1 and B7, corresponding to 32.70 Hz and 1975.5 Hz.

16. Kim et al. 2018.

Human pitch perception

Initially two theories of human pitch perception: place and temporal.[17]

In the place theory (place coding), spectral analysis is done in the cochlea, so that the *resolved* harmonics of a sound excite different parts of the basilar membrane (BM), firing neurons with different characteristic frequency.

In the temporal theory (temporal coding, phase locking), the *unresolved* harmonics form a complex waveform in the BM, and firing neurons lock to the phase of the envelope of the complex waveform.

17. Brian C. J. Moore. 2013. *An Introduction to the Psychology of Hearing* [in English]. 6th ed. 203–242. United Kingdom: Emerald Group Publishing Limited.

Human pitch perception

Initially two theories of human pitch perception: place and temporal.[17]

In the place theory (place coding), spectral analysis is done in the cochlea, so that the *resolved* harmonics of a sound excite different parts of the basilar membrane (BM), firing neurons with different characteristic frequency.

In the temporal theory (temporal coding, phase locking), the *unresolved* harmonics form a complex waveform in the BM, and firing neurons lock to the phase of the envelope of the complex waveform.

17. Brian C. J. Moore. 2013. *An Introduction to the Psychology of Hearing* [in English]. 6th ed. 203–242. United Kingdom: Emerald Group Publishing Limited.

- First part is related to the tonotopic organization of the inner ear – well established and independently confirmed in different studies. 100,200,300hz example. Second is still a matter of dispute
- gabor tf uncertainty principle. Our real sensitivity to <1% differences in fundamental frequencies for resolved harmonics is better than what would be possible with firing rate/place coding, suggesting temporal coding
- More sophisticated models (which explain or account for human experimental data) require both place and temporal analyses (pitch can be recognized even with unresolved harmonics). Tentative neural component, processed after the ears – two harmonics presented to different ears create the correct fundamental.