Cumulative Occurrences of Notes as Basis for Key Classification

Dela Torre, Promentilla

Table of Contents

Introduction

Definitions

- Pitch - fundamental frequency, the highness/lowness of a sound that we perceive (440 Hz)
- Note - a name arbitrarily assigned to a particular frequency on the field of music (440 Hz pitch as A)
- Scale - ordered notes in increasing/decreasing fundamental frequency
- Key - a group of notes, the basis of musical composition
    - Key signatures with sharps: G, D, A, E, B, F#, C# (progressively each adds a sharp up to 7)

Key Classification

- Key classification is an important aspect in music analysis.
- Korzeniowski and Widmer (2018) also says that key classification is of importance because it is important to mix music harmonically well.
- Korzeniowski and Widmer (2018) also observes that deriving a key is a demanding task and that it would be costly even for experts to annotate large amounts of music classification.
- Therefore, there is an opportunity for research on key classification systems.

Review of Related Literature

Key Profile:

“The original key-profile model is based on a set of twelve-valued vectors, called key- profiles, representing the stability or compatibility of each pitch-class relative to each key.” (Temperley, 2002)

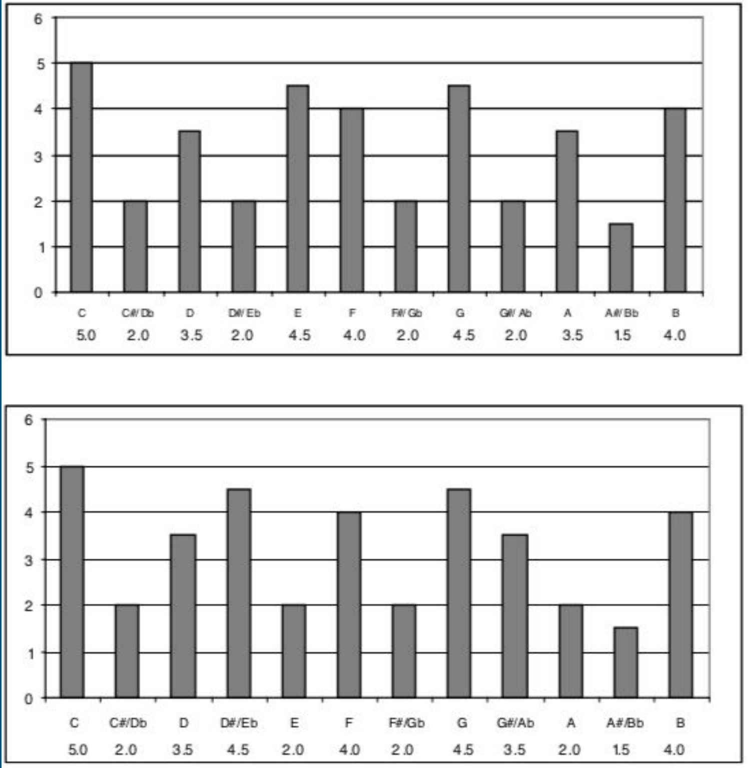


Fig. 1. Key-Profiles for C major (above) and C minor (below).

Neural Networks

"A neural network is a network or circuit of neurons, or in a modern sense, an artificial neural network, composed of artificial neurons or nodes." (Hopfield, 1982)

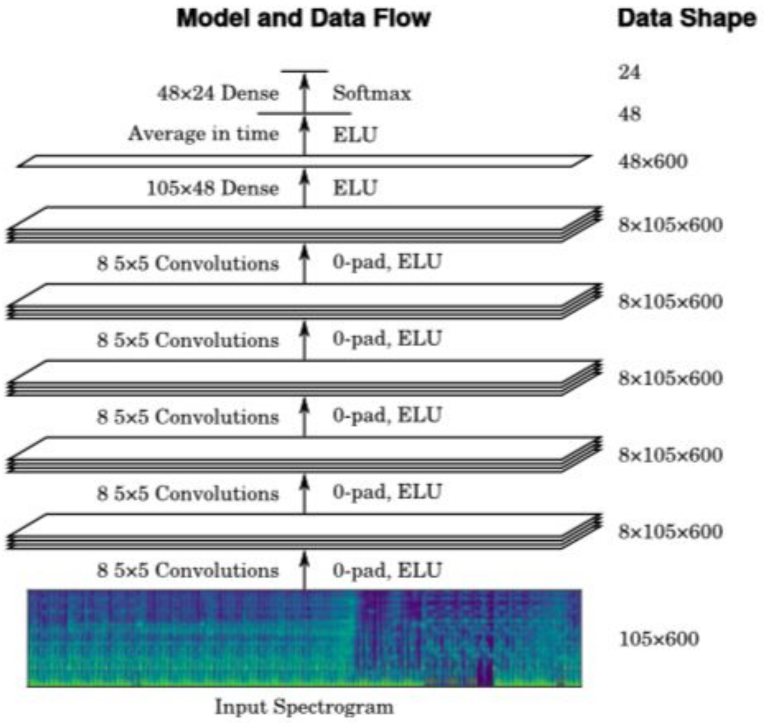


Fig. 1. Neural network for musical key estimation. Convolutions are 0-padded, i.e. the feature maps keep the input shape. All layers are followed by exponential-linear activations [15], except the last, which is followed by a softmax.

Two Methods of Key Classification

Profile Based Methods:

- Uses key profiles, pitch class profiles, harmonic pitch class profiles, etc.

General Steps

- Base the classification on the template of key profiles
- Process the signal or train the model
- Classify based on correlations with the profiles.

Neural Network Based Methods:

- Uses Neural Networks
    - CNN

General Steps

- Train neural network on a vast amount of data
- Classify keys

Summary of Performances For Each Reviewed Literature

Author	Method	Dataset	Performance
Temperley (2002)	Key Profile Bayesian Model	Kostka-Payne Corpus	77.1%
Pauws (2004)	Key Profile Chromatograms	Piano Sonatas	75.1%
Noland and Sandler (2006)	Key Profile HMM	Beatles Songs	91%
O ’brien and Lerch (2015)	Key Profile	GTZAN Dataset Different Music Genre	35.04%
Korzeniowski and Widmer (2017)	Neural Networks (CNN)	GiantSteps Key Dataset Electro Music	67.9% (62.5%)
Korzeniowski and Widmer (2018)	Neural Networks (CNN)	GiantSteps Key Dataset	67.9% (79.9%)

Disadvantages of Each Method

Profile Based Methods:

- Profiles are overfitted to their music genre.
- The drawback is that key templates differ for musical genres and favour one key mode over another (Korzeniowski & Widmer, 2018).
- Cannot handle other music genre.

Neural Network Based Methods:

- Needs vast amounts of training data
- It might also be the case that this training data needs to be annotated
- Needs extensive training time

Proposed Algorithm

- Profile Based but music Genre-Agnostic.
- Does not require any training data.
- Does not require training time.
- Aims to reach a reasonable performance.
  - Given the limitations.

Experimental Setup

Input Signal Dataset

- The database to be used is the GiantSteps Key Dataset.
- The dataset is comprised of 604 audio recordings each one with a duration of 2 minutes.
- The sampling rate is of standard audio quality which is 44100 samples per second or 44.1 kHz.
- Each recording has a corresponding ground truth value for the global key.
- Experimentation on this database also serves as a platform for direct comparison with neural networks-based key classification.

Key Signatures

C major or A minor	C	D	E	F	G	A	B
C# major or A# minor	C#	D#	F	F#	G#	A#	C
D major or B minor	D	E	F#	G	A	B	C#
D# major or C minor	D#	F	G	G#	A#	C	D
E major or C# minor	E	F#	G#	A	B	C#	D#
F major or D minor	F	G	A	A#	C	D	E
F# major or D# minor	F#	G#	A#	B	C#	D#	F
G major or E minor	G	A	B	C	D	E	F#
G# major or F minor	G#	A#	C	C#	D#	F	G
A major or F# minor	A	B	C#	D	E	F#	G#
A# major or G minor	A#	C	D	D#	F	G	A
B major or G# minor	B	C#	D#	E	F#	G#	A#

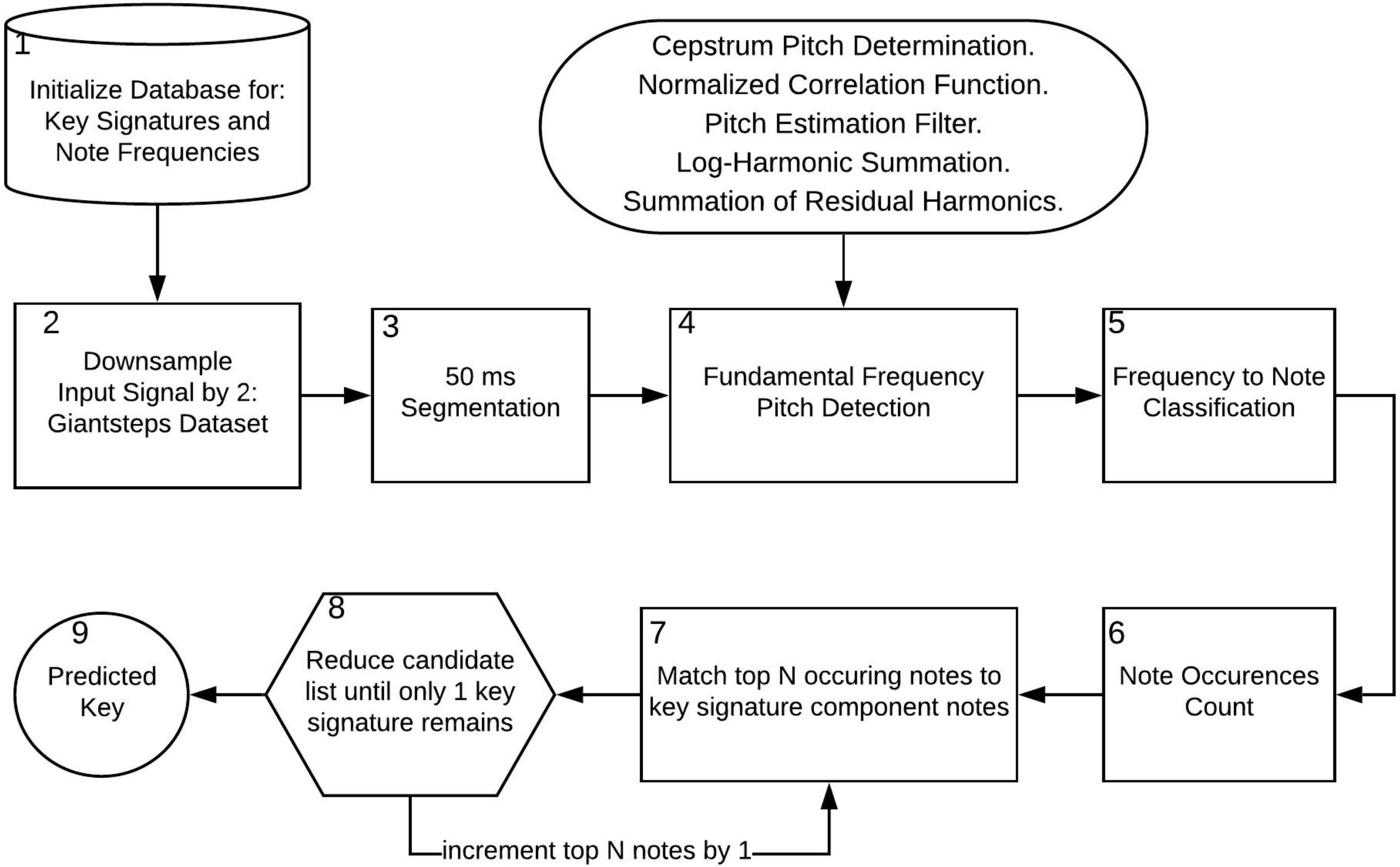
Note Frequencies

Octave Note	1	2	3	4	5	6	7
C	32.703 (24)	65.406 (36)	130.81 (48)	261.63 (60)	523.25 (72)	1046.5 (84)	2093.0 (96)
C♯/D♭	34.648 (25)	69.296 (37)	138.59 (49)	277.18 (61)	554.37 (73)	1108.7 (85)	2217.5 (97)
D	36.708 (26)	73.416 (38)	146.83 (50)	293.66 (62)	587.33 (74)	1174.7 (86)	2349.3 (98)
E♭/D♯	38.891 (27)	77.782 (39)	155.56 (51)	311.13 (63)	622.25 (75)	1244.5 (87)	2489.0 (99)
E	41.203 (28)	82.407 (40)	164.81 (52)	329.63 (64)	659.26 (76)	1318.5 (88)	2637.0 (100)
F	43.654 (29)	87.307 (41)	174.61 (53)	349.23 (65)	698.46 (77)	1396.9 (89)	2793.8 (101)
F♯/G♭	46.249 (30)	92.499 (42)	185.00 (54)	369.99 (66)	739.99 (78)	1480.0 (90)	2960.0 (102)
G	48.999 (31)	97.999 (43)	196.00 (55)	392.00 (67)	783.99 (79)	1568.0 (91)	3136.0 (103)
A♭/G♯	51.913 (32)	103.83 (44)	207.65 (56)	415.30 (68)	830.61 (80)	1661.2 (92)	3322.4 (104)
A	55.000 (33)	110.00 (45)	220.00 (57)	440.00 (69)	880.00 (81)	1760.0 (93)	3520.0 (105)
B♭/A♯	58.270 (34)	116.54 (46)	233.08 (58)	466.16 (70)	932.33 (82)	1864.7 (94)	3729.3 (106)
B	61.735 (35)	123.47 (47)	246.94 (59)	493.88 (71)	987.77 (83)	1975.5 (95)	3951.1 (107)

Pitch Detection Methods

1. Cepstrum Pitch Determination (CEP).
2. Normalized Correlation Function (NCF).
3. Pitch Estimation Filter (PEF).
4. Log-Harmonic Summation (LHS).
5. Summation of Residual Harmonics (SRH).

Proposed Algorithm


1
Initialize Database for:
Key Signatures and
Note Frequencies
Cepstrum Pitch Determination.
Normalized Correlation Function.
Pitch Estimation Filter.
Log-Harmonic Summation.
Summation of Residual Harmonics.
2
Downsample
Input Signal by 2:
Giantsteps Dataset
3
50 ms
Segmentation
4
Fundamental Frequency
Pitch Detection
5
Frequency to Note
Classification
6
Note Occurences
Count
7
Match top N occuring notes to
key signature component notes
8
Reduce candidate
list until only 1 key
signature remains
9
Predicted
Key
increment top N notes by 1

Proposed Algorithm

1. Initialize the databases for key signatures and frequencies for each note.
2. Downsample the input signal by a factor of 2.
3. Segment each frame into 52 ms (millisecond) windows with an overlap of 42 ms.
4. Estimate fundamental frequency/pitch for each frame.
5. Classify each frequency into one of the 12 notes.
6. Count the occurrences of each note for the entire input signal.
7. Set a list of candidate keys (initially 12 keys).
8. Starting from 1, get the top *n* occurring notes.
    a. Match the top *n* occurring notes with the 7 note components of each key signature.
    b. Reset the list of candidate keys to include only those with the highest matches with the top *n* occurring notes.
    c. Increment the top *n* occurring notes by 1.
    d. Repeat until only 1 candidate key remains.
9. The remaining candidate key is selected to be the predicted key for classification.

Results

Pitch Detection Methods

Pitch Detection Methods:	Accuracy Rate:
PEF	0%
LHS	12.5%
NCF	18.75%
SRH	34.375%
CEP	46.875%

Cepstrum Pitch Determination Parameters

Cepstrum Parameters	Accuracy Rate
Default Cepstrum Parameters • Segmented 52 ms • Overlap 42 ms	43.75%
Setting Frequency Range from 20 Hz ~ 4000 Hz	15.625%
Segmented by 100 ms	46.875%
Segmented by 200 ms	46.875%
Overlap 0 ms / No Overlap	46.875%
Segmented by 100 ms and No Overlap	40.625%

GiantSteps Key Dataset

CEP Conditions	Accuracy Rate
Downsampled by 2 Segmentation 52 ms Overlap 42 ms	45.03%
Downsampled by 4 Segmentation 52 ms Overlap 42 ms	34.60%
Downsampled by 2 Segmentation 52 ms Overlap 0 ms / No Overlap	42.55%

5.781 s

Key Classification Time - time it took to classify one 2 minute audio recording

Discussion

Comparative Results Among Key Classification Methods and Proposed Algorithm

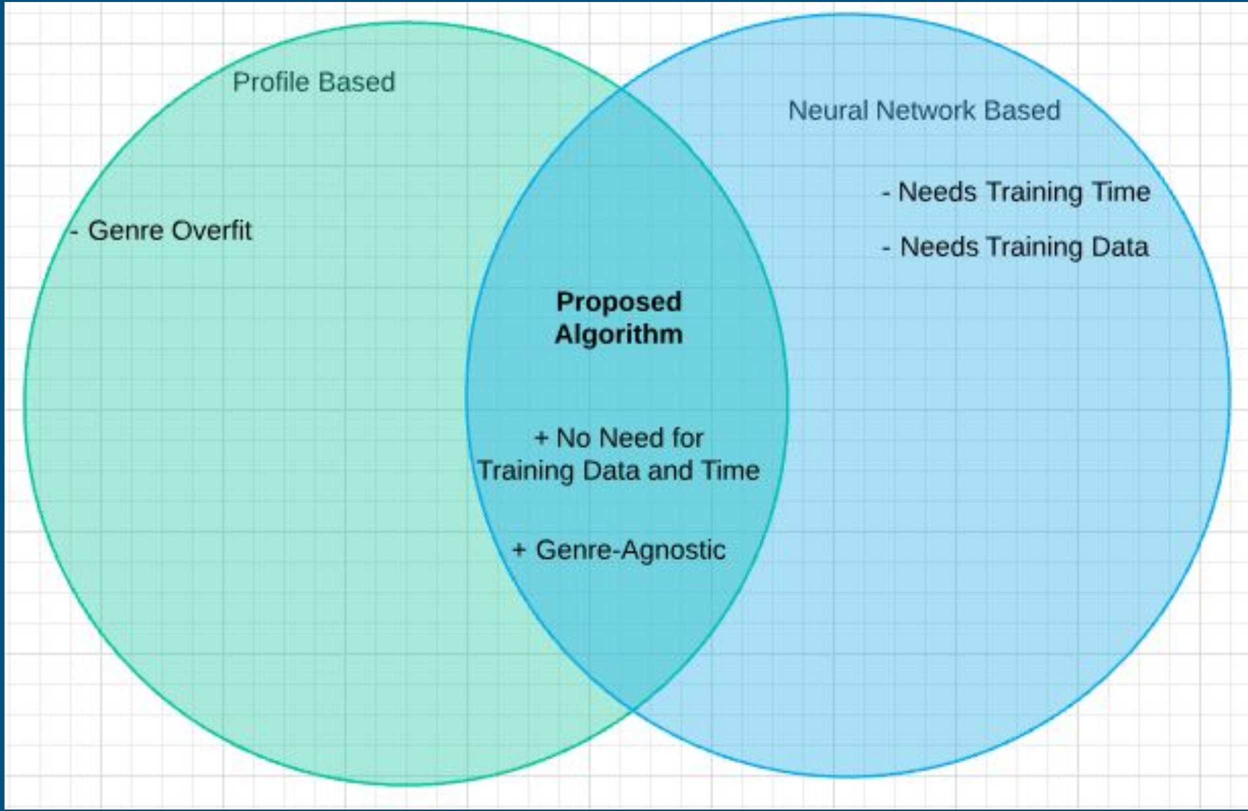
Methods	Accuracy Rate
Profile Based Method	75.1%
Profile Under Different Music Genres	35.04%
Neural Network Based Method	67.9%
Proposed Algorithm	45.03%

Performance on GiantSteps Key Dataset

According to the website by Knees et al. (2015) detailing their study and dataset, a table of the performances for published key classifiers, including commercial ones, on the GiantSteps Key dataset was shown.

Algorithm	Accuracy Rate
Beatport (2014 website)	29.14
QM-Key	39.40
Essentia (2015)	30.46
Faraldo et al., ECIR 2016 – edmt	33.40
Faraldo et al., ECIR 2016 – edma	58.10
Faraldo et al., ECIR 2016 – edma	58.10
Faraldo et al., ECIR 2016 – edmm	64.20
KeyFinder	45.36
Mixed-In-Key 7	67.22
Pioneer Rekordbox v3.2.2	71.85

Proposed Algorithm



Conclusion

Conclusion

Profile based methods suffer from overfitting its music genre while neural network based methods require training data and time. The proposed algorithm aims to reach a compromise between the two methods by being genre-agnostic and not requiring any sort of training. The proposed algorithm uses cepstrum pitch determination and the iterative elimination of candidate keys to classify the keys from the GiantSteps Key dataset. The proposed algorithm reaches a reasonable performance achieving its initial objective.

References

- [1] Pauws, S. (2004, 01). Musical key extraction from audio..
- [2] Korzeniowski, F., & Widmer, G. (2018, 08). Genre-agnostic key classification with convolutional neural networks.
- [3] Temperley, D. (2002, 01). A bayesian approach to key-finding. In (p. 149-155). doi: 10.1007/3-540-45722-418
- [4] Noland, K., & Sandler, M. (2006, 01). Key estimation using a hidden markov model. In (p. 121-126).
- [5] O ’brien, C., & Lerch, A. (2015, 09). Genre-specific key profiles..
- [6] Korzeniowski, F., & Widmer, G. (2017). End-to-end musical key estimation using a convolutional neural network.CoRR,abs/1706.02921. Retrieved from http://arxiv.org/abs/1706.02921
- [7] Knees, P., Faraldo,´A., Herrera, P., Vogl, R., B¨ock, S., H¨orschl¨ager, F., & Le Goff,M. (2015, October). Two data sets for tempo estimation and key detection in electronic dance music annotated from user corrections. In Proceedings of the 16th international society for music information retrieval conference(ISMIR’15). M´alaga, Spain.
- [8] Noll, Michael A. "Cepstrum Pitch Determination." The Journal of the Acoustical Society of America. Vol. 31, No. 2, 1967, pp. 293–309.
- [9] Atal, B.S. "Automatic Speaker Recognition Based on Pitch Contours." The Journal of the Acoustical Society of America. Vol. 52, No. 6B, 1972, pp. 1687–1697.
- [10] Gonzalez, Sira, and Mike Brookes. "A Pitch Estimation Filter robust to high levels of noise (PEFAC)." 19th European Signal Processing Conference. Barcelona, 2011, pp. 451–455.
- [11] Hermes, Dik J. "Measurement of Pitch by Subharmonic Summation." The Journal of the Acoustical Society of America. Vol. 83, No. 1, 1988, pp. 257–264.
- [12] Drugman, Thomas, and Abeer Alwan. "Joint Robust Voicing Detection and Pitch Estimation Based on Residual Harmonics." Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH. 2011, pp. 1973–1976.

Thank You for Listening.