Automatic Chord Recognition from Audio With the Use of Enhanced Pitch Class Profile (EPCP)

Varchasva Agarwal

Maharaja Agrasen Institute of Technology

Guru Gobind Singh Indraprastha University Delhi

Abstract

In this article, a feature vector called Enhanced Pitch Introduced a class profile (EPCP) to automatically recognize chords from the original audio. For this reason, harmonic products The spectrum is first obtained from the DFT of the input signal, Then there is the algorithm for calculating the 12-dimensional pitch The class profile is applied to it to give the EPCP feature vector. EPCP vectors are associated with predefined templates For the 24 major/minor triads, the template that produces the greatest correlation is identified as the chord of the input signal. Experimental results show that EPCP produces fewer errors It is superior to traditional PCP in terms of frame rate and chord recognition.

1 Introduction

A musical chord can be defined as a set of simultaneous The continuity of tones and chords over time or the progression of chords forms the core of harmony in a piece of music. therefore Analyze the overall harmony structure of the music Usually start by marking each chord in it. Even for experienced listeners, this is a difficult and tedious task. The score is in hand. The automation of chord marking can therefore be very Useful for those who want to perform harmonic analysis on music. Once the harmonic content of the music is known, it can be further used for higher-level structural analysis. It can also Good intermediate representation for this type of music signal Applications such as music segmentation, music similarity recognition and audio thumbnails. For these and other reasons, Automatic chord recognition has attracted some Researchers in the music information retrieval community.

Chromatogram or pitch level profile has become an option The feature set in automatic chord recognition or key extraction has been introduced by Fujishima (Fujishima 1999). The perception of pitch has two dimensions: height and chromaticity. The pitch moves vertically in octaves, telling which octave One note belongs to. On the other hand, chroma tells where it is Related to other people in the octave. Chromatogram Or the sound level profile is a 12-dimensional vector representation of chromaticity, which is expressed in Each of the twelve semitones in the chromatic scale. Due to chords Consists of a set of tones, whose labels are only determined Through the position of these hues in the chroma, no matter how they are Height, the chromaticity diagram seems to be the ideal characteristic of the representation A musical chord.

There are some changes in the 12-bin chromatogram obtained, But it usually follows the following steps. First, DFT Calculate the input signal X(k), calculate the constant Q from X(k) and transform XCQ, which uses the logarithmic interval frequency to reflect the frequency resolution Human ear (Brown 1990). Frequency resolution Constant Q transformation follows isothermal transformation Scale, so the k-th spectral component is defined as

$$f_k = (2^{1/B})^k f_{min},$$

Where fk changes from fmin to the upper limit frequency, both Which is set by the user, B is the number of boxes In the octave of constant Q transformation. Once XCQ(k) is By calculation, a chromaticity vector CH can be easily obtained As:

$$CH(b) = \sum_{m=0}^{M-1} \left| X_{CQ}(b + mB) \right|,$$

Where b = 1, 2, · · ·, B is the chromatogram bin index, and M is the number of octaves spanned in the constant Q spectrum. For chord recognition, only B = 12 is required, but B = 24 or B = 36 is also used for preprocessing like fine Tuning.

The following is how the rest of the paper is organised: Section 2 covers related work in this subject; Section 3 begins by describing the issues in past work caused by using the chromagram as the feature set, and then proposes a remedy by using the Enhanced Pitch Class Profile as a feature vector (EPCP). Section 4 presents a comparison of the two approaches using real-world recording instances, followed by commentary. In section 5, we draw findings and make recommendations for further research.

2 Related Work

Chord detection and key extraction algorithms from audio recordings have almost exclusively employed chromagram or pitch class profile (PCP) based features as a front end. Fujishima created a real-time chord identification system employing a 12-dimensional pitch class profile obtained from the DFT of an audio stream and pattern matching using binary chord type templates (Fujishima 1999). Gomez and Herrera suggested a method that extracts tonal metadata such as chord, key, scale, and cadence information from audio recordings automatically (Gomez and Herrera 2004).

They linked a Harmonic Pitch Class Profile (HPCP), which is based on Fujishima's PCP, with a chord or key model adopted from Krumhansl's cognitive study, as the feature vector (Krumhansl 1990). Similarly, Pauws employed the maximum-key profile correlation approach to extract key from raw audio data, averaging chromagram characteristics over varying length fragments at various places and correlating them with Krumhansl and Kessler's 24 major/minor key profile vectors (Pauws 2004). Harte and Sandler utilised a 36-bin chromagram to identify the input audio's tuning value based on the distribution of peak locations, and then created a 12-bin, semitone-quantized chromagram that could be connected with binary chord templates (Harte and Sandler 2005). Sheh and Ellis provided a statistical learning technique for chord segmentation and recognition, in which they employed hidden Markov models (HMMs) trained using the Expectation Maximization (EM) algorithm, with chord labels considered as hidden values inside the EM framework (Sheh and Ellis 2003). Bello and Pickens used HMMs with the EM algorithm as well, but they added musical knowledge to the models by defining a state transition matrix based on the key distance in a circle of fifths and avoiding random initialization of a mean vector and a covariance matrix of the observation distribution, which was modelled by a single Gaussian (Bello and Pickens 2005). Furthermore, they selectively update the parameters of interest while training the model for parameter estimation, based on the premise that a chord template or distribution is almost universal, preventing distribution parameter change. The issues with the chromagram-based approach are discussed in the next section.

3 Enhanced Pitch Class Profile

All of the previously mentioned work on harmony acknowledgment or key extraction, while the subtleties of the calculations might differ, make them thing in like manner: they all utilization a chromagram as the element vector. To recognize a harmony, some utilization a format coordinating with calculation (Fujishima 1999; Gomez and Herrera 2004; Pauws 2004; Harte and Sandler 2005), though others utilize a probabilistic model like HMMs (Sheh and Ellis 2003; Bello and Pickens 2005). Albeit the Well have for some time been acknowledged by a discourse acknowledgment society for their fantastic exhibition, their presentation in a harmony acknowledgment task, best case scenario, only tantamount to that of simple design coordinating with calculations. Besides, it is a very tedious and drawn-out occupation to physically mark all the harmony limits in accounts with comparing harmony names to create the preparation information. Be that as it may, the regular 12-dimensional pitch class profile might create a few issues, especially when utilized with the layout coordinating with calculation.

3.1 Problems with Chroma-based Approach

In harmony acknowledgment frameworks with a layout coordinating algorithm, formats of harmony profiles are first characterized. These formats are 12-dimensional, and each receptacle relates to a contribute class a chromatic scale. They are generally parallel or all-or-none sort; i.e., each container is either 0 or 1. For example, since a C significant triad involves three notes at C (root), E (third), and G (fifth), the layout for a C significant set of three is [1,0,0,0,1,0,0,1,0,0,0,0], and [0,0,1,0,0,0,0,1,0,0,0,1] for a G significant set of three, where the format marking is [C,C#,D,D#,E,F,F#,G,G#,A,A#,B]. As can be found in these models, each layout in 12 significant sets of three will be only a moved variant of the other, and for the minor group of three, it will be as old as significant triad with its third moved by one to the left; e.g., [1,0,0,1,0,0,0,1,0,0,0,0] is a C minor triad layout. Formats for increased, decreased, or seventh harmonies can be characterized correspondingly.

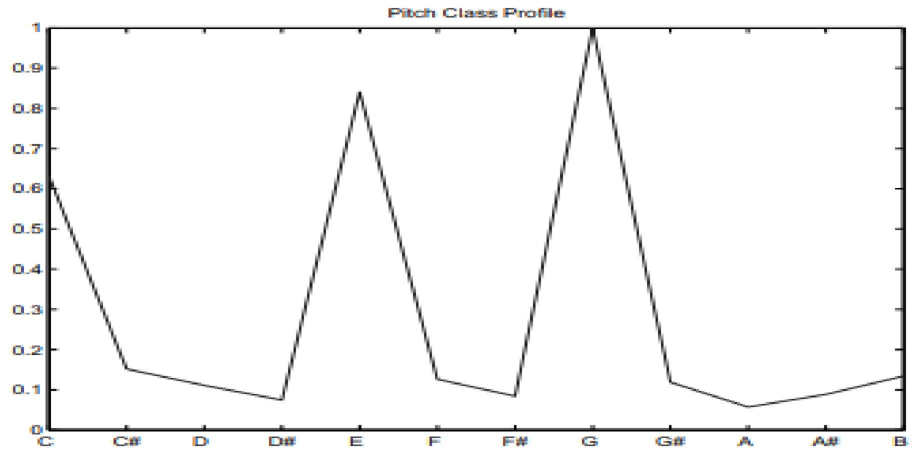


Figure 1: Chroma vector of a C major triad played by piano.

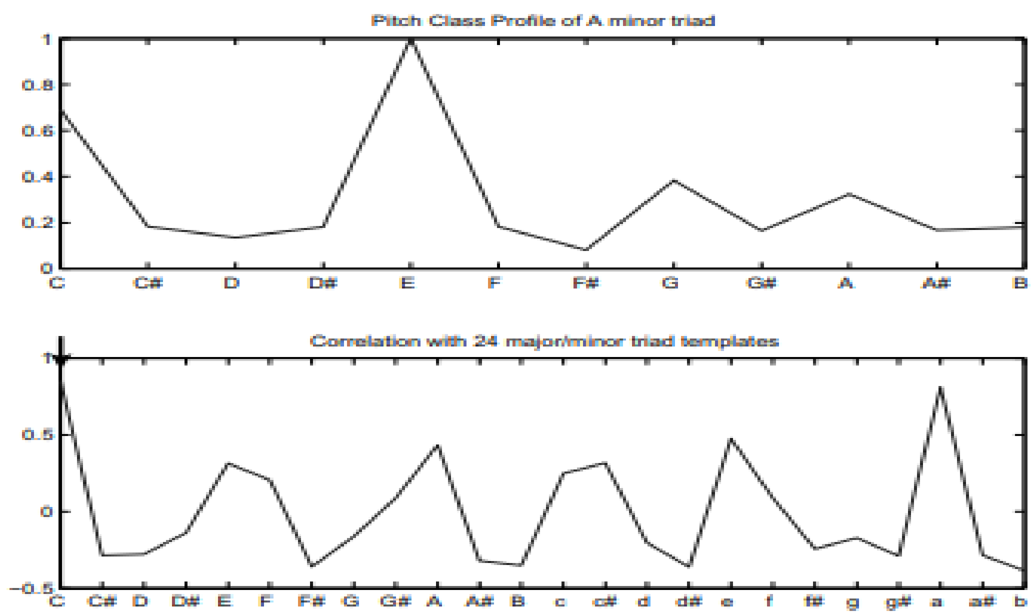
The chroma vector from genuine accounts can never be paired, nonetheless, in light of the fact that acoustic instruments produce overtones just as central notes. Figure 1 shows the chroma vector from C significant group of three played by a piano. As can be seen in Figure1, regardless of whether the most grounded tops are found at C, E, and G, the chroma vector has nonzero power at every one of the 12 pitch classes as a result of the suggestions produced by the harmony tones. This uproarious chroma vector might make confusion the acknowledgment frameworks with parallel kind formats. The issue can be more genuine between the harmonies that share at least one notes like a significant set of three and its equal minor or its relative minor; e.g., a C significant set of three and a C minor triad share two harmony notes C and G, and a C significant set of three what's more, A minor set of three share notes C and E for all intents and purpose. A minor set of three from a genuine model and its connection with 24 significant/minor triad layouts. Like the past model in Figure 1, the chroma vector has most energy at the harmony tones, i.e., at A, C, and E, however the pitch class G, which isn't a harmony tone, has more energy than the pitch class A. This might be brought about by non-harmony tones as well as by hints of different tones. The high energy in G accordingly gives the most extreme relationship with a C significant group of three, which is an overall major of a minor set of three, as indicated by a bolt in the lower figure, what's more, in this way the framework recognizes the harmony as a C significant triad. Comparative blunders are made between the equal major and minor sets of three.

3.2 Enhancing the Chroma Vector

The issues exemplified in the above area can be addressed by upgrading the ordinary chroma vectors with the goal that they can turn out to be to a greater degree a paired kind, actually like their tem plates utilized for design coordinating. To this end, the Harmonic Product Spectrum (HPS) was first gotten from the DFT of the information signal, and the Enhanced Pitch Class Profile was then registered from the HPS, rather than the first DFT. The Harmonic Product Spectrum has been utilized for de tecting key recurrence in an intermittent sign or for de termining contribute human discourse (Schroeder 1968; Noll 1969). The calculation for registering the HPS is exceptionally basic and depends on the harmonicity of the sign. Since most acoustic instruments and human voice produce a sound that has harmonics at the whole number products of its major recurrence, destroying the first greatness range by a number will likewise yield a top at its major recurrence.

$$
\begin{aligned}
HPS(\omega) &= \prod_{m=1}^{M} |X(m\omega)| \\
F_0 &= \arg\max_{\omega}\{HPS(\omega)\},
\end{aligned}
$$

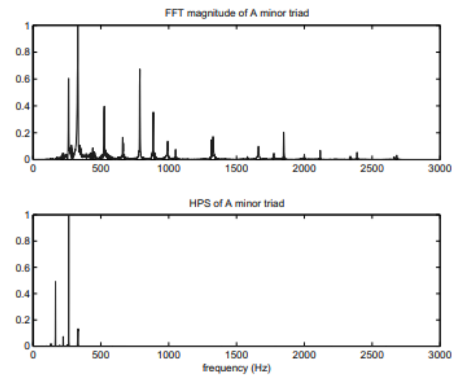
The last HPS is acquired by increasing the spectra, and the top in the still up in the air as the basic recurrence. The calculation is summed up in the accompanying conditions: where HP S(ω)is the Harmonic Product Spectrum, X(ω) is the DFT of the sign, M is the quantity of sounds to be thought of, and F0 is the assessed key recurrence.



This calculation was demonstrated to function admirably in monophonic signals, yet incidentally, it likewise works for assessing multiple contributes polyphonic signs. On account of harmony recognition application, notwithstanding, wrecking the first range by the powers of 2 ended up working better compared to decimating by number numbers. This is on the grounds that music not at the force of 2 or at the octave counterparts of the crucial recurrence might add to producing some energy at other pitch classes than the individuals who involve harmony tones, along these lines preventing upgrading the range. For instance, the fifth harmonic of A3 is C6#, which isn't a harmony tone in A minor triad. Thusly, Equation 3 is adjusted as follows to reflect this: When the HPS is processed from the DFT, the EPCP vector is gotten just by registering the chroma vector from the HPS rather than the DFT. In Figure 4 are shown the EPCP vector from a similar model, and its connection with the 24 major/minor group of three formats. Overlaid are the customary PCP vector and its relationship in specked lines for examination. We can plainly see from the figure that non-harmony tones are sufficiently smothered to underline the harmony tones as it were, which are A, C, and E in this model. This eliminates the vagueness between its general significant set of three, and the subsequent connection has a greatest worth at the A minor set of three. For explanation, Table 1 shows connection results between the PCP vector and the EPCP vector and the triad formats.

$$HPS(\omega) = \prod_{m=0}^{M} |X(2^m\omega)| \qquad (5)$$

Figure 3 shows the DFT of the same example as in Figure 2, and the corresponding HPS obtained using $M = 3$.



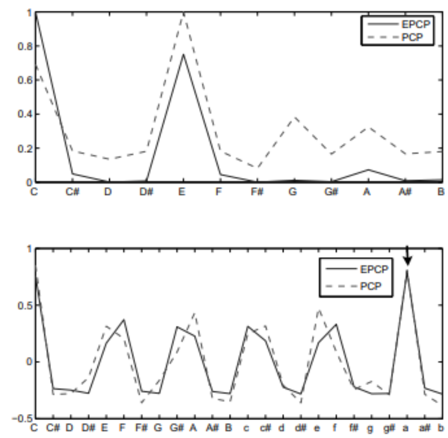


Figure 4: EPCP vector of an A minor triad, and its correlation with 24 major/minor triad templates. Arrow in the lower figure represents the maximum correlation.

4 Experimental Results

We looked at the EPCP include vectors against the con ventional pitch class profile utilizing genuine recording models. The sound documents are downsampled to 11025 Hz, and 36 canisters for every octave consistent Q change was performed between fmin = 96 Hz and fmax = 5250 Hz. The PCP/EPCP vec pinnacle of 36 aspects gives goal sufficiently fine to distin guish contiguous semitone frequencies paying little mind to tuning. A window length of 8192 examples was utilized with the hopsize of 1024 examples, which compare to 743 ms and 92.3 ms, separately. A generally long window is vital for cap turing consonant data saw in a melodic entry like arpeggios. A 36-dimensional PCP/EPCP vector is then processed from the consistent Q change as depicted in Equation 2. At last, a tuning calculation proposed by Harte and Sandler is applied to a 36-receptacle PCP/EPCP vector to com pensate mistuning that might be available in accounts, yielding a 12-dimensional PCP/EPCP vector. The subtleties on the tun ing calculation can be seen as in (Harte and Sandler 2005). Figure 5 shows outline level acknowledgment of initial 20 seconds of Bach's Prelude in C major performed by Glenn Gould. The strong line with X's addresses outline level harmony acknowledgment utilizing the EPCP vector, and the ran line with circles us ing the regular PCP vector. It is obvious from the figure that the EPCP vector makes less mistakes than the PCP vector. Especially, there are many mistakes with the

PCP vectors in recognizing a D minor harmony, a large portion of which are created by a turmoil between its equal major, or D major. Comparable blunders are displayed in perceiving

Table 1: Correlation results of PCP/EPCP vectors used in Figure 2 and 4 with 24 major/minor triad templates. Capital letters represent major triads, and small letters minor triads. Maximum correlation values are in boldface.

Template	**C**	C#	D	D#	E	F	F#	G	G#	A	A#	B
PCP	**0.8528**	-0.2861	-0.2786	-0.1379	0.3144	0.2061	-0.3592	-0.1621	0.0877	0.4342	-0.3202	-0.3510
Template	c	c#	d	d#	e	f	f#	g	g#	a	a#	b
PCP	0.2489	0.3167	-0.2041	-0.3607	0.4757	0.0886	-0.2432	-0.1719	-0.2895	0.8090	-0.2847	-0.3849
Template	C	C#	D	D#	E	F	F#	G	G#	A	A#	B
EPCP	0.7574	-0.2348	-0.2481	-0.2761	0.1652	0.3732	-0.2577	-0.2766	0.3092	0.2275	-0.2598	-0.2794
Template	c	c#	d	d#	e	f	f#	g	g#	**a**	a#	b
EPCP	0.3142	0.1859	-0.2217	-0.2825	0.1702	0.3316	-0.2196	-0.2797	-0.2780	**0.7939**	-0.2313	-0.2831

A minor group of three, confounded by its relative major, or C significant triad. Then again, the EPCP vector makes no blunder between the ideal harmony and harmonies that are pleasingly firmly identified with. Additionally displayed in Figure 6 are acknowledgment results for smoothed PCP/EPCP vectors across 11 casings. This smoothing system decreases blunders because of unexpected changes in signals brought about by transient and commotion like sounds, which can cloud consonant parts. Most blunders are revised in the event of the EPCP vectors while there are still a significant number mistakes with the PCP vectors. One more model is displayed in Figure 7. As displayed in the past model, utilizing the PCP vector, a few edges of a D significant group of three were misidentified as a B minor triad in the selection on account of their relative major-minor relationship. Additionally observable is an absolute misidentification of a B minor set of three as an E significant triad, which have no nearby consonant relationship with one another. The EPCP vector again makes no blunder in the two cases, and undeniably less mistakes happen overall. Figure 8 shows acknowledgment results after a smoothing professional cess more than 11 edges. Some prompt mistakes are amended in the two cases, however the blunders found in the center and toward the finish of the selection were not remedied by a straightforward smoothing process if there should arise an occurrence of the PCP vector.

5 Conclusions

We have introduced another element vector for programmed harmony acknowledgment from the crude sound which is more appropriate than the regular chroma vector when utilized with design coordinating with calculations with the twofold sort harmony tem plates. The new component vector, the Enhanced Pitch Class Pro record, or the EPCP vector was processed from the symphonious item range of an info signal rather than the DFT to curb the powers at pitch classes involved by suggestions of the harmony tones. Trial results with genuine recording models show the EPCP vector beats the regular PCP vector in distinguishing harmonies both at the casing rate and in smoothed portrayal. The distinction in execution between the two element vectors ends up being unmistakable when there is a more prominent level of disarray between pleasingly firmly related harmonies, for example, relative or standard allel major/minor harmonies. The outcomes show that the EPCP vector is substantially less touchy to such disarrays. Later on, we intend to utilize the EPCP vector as a front finish to the AI models, for example, stowed away Markov models (HMMs) or Support Vector Machines (SVMs), which have been as of late utilized in a harmony acknowledgment application. Moreover, more complex calculations for upgrading the chroma vector are additionally being thought of.

References

Schroeder, M. R. (1968). Period histogram and product spectrum: New methods for fundamental-frequency measurement. Journal of the Acoustical Society of America 43(4), 829–834.

Krumhansl, C. L. (1990). Cognitive Foundations of Musical Pitch. Oxford University Press. 310 Noll, A. M. (1969). Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate. In Proceedings of the Symposium on Computer Processing ing Communications, New York, pp. 779–797.

Brown, J. C. (1990). Calculation of a constant q spectral transform. Journal of the Acoustical Society of America 89(1), 425–434.

Fujishima, T. (1999). Realtime chord recognition of musical sound: A system using Common Lisp Music. In Proceedings of the International Computer Music Conference, Beijing. International Computer Music Association.

Sheh, A. and D. P. Ellis (2003). Chord segmentation and recognition using em-trained hidden Markov models. In Proceedings of the International Symposium on Music Information Retrieval, Baltimore, MD. Pauws, S. (2004). Musical key extraction from audio. In Proceedings of the International Symposium on Music Information Retrieval, Barcelona, Spain.

Gomez, E. and P. Herrera (2004). Automatic extraction of tonal metadata from polyphonic audio recordings. In Proceedings of the Audio Engineering Society, London. Audio Engineering Society.

Harte, C. A. and M. B. Sandler (2005). Automatic chord identification using a quantised chromagram. In Proceedings of the Audio Engineering Society, Spain. Audio Engineering Society.

Bello, J. P. and J. Pickens (2005). A robust mid-level representation for harmonic content in music signals. In Proceedings of the International Symposium on Music Information Retrieval, London, UK.