Ladies and gentlemen, our paper is divided into two parts:

In the first part I will explain some basic new details of our software for analysis of harmonic structures

In the second one I will present some statistical results, gained using our software on three sets of midi data – piano works by three composers – Mozart, Schubert and Brahms.

Twenty years ago, we started to build up the first version of MS DOS software for analysis of classical harmony with an interactive score editor. From that time many things were changed and we present the sixth version today.

The core idea is the same, but more details and solutions are modified, except for functional analysis. The main changes are 1. **platform shift**, 2. **MIDI input**, 3-4. **new chordal and tonal analysis**. Let me briefly comment upon them.

1. The software now works under much more comfortable operating system – **Microsoft Windows.**

2. The idea of automatic input is now implemented. This reduces input errors and allows for processing large number of compositions. We use **MIDI data** as a dataset for automatic style analysis in harmonic sense. We use to download them from a perfect website www.classicalarchives.com. They will be presented later.

3. The main idea how to detect the structure of chords in a **chordal analysis**, is modified with a strong improvement. In older versions there were excluded all doubled tones from the chord if they were a part of structure of thirds. The melodic tones, i.e. the nonchordal tones, were all other tones, which were not a part of structure of thirds and were excluded, too. Today **we count together the rhythmical values of given tone in one beat in all octaves**, than the software looks for structure of thirds first among the longer rhythmical values. Then we decide, which tones belong to the chord and which are nonchordal. The reason is, that for human perception the quantitative factors are very important. Those tones, which sound longer, or which are more times repeated, are also heard and considered in listener‘s mind as **more important**, as chordal or central tones. Those, which are only very short and are not repeated, are only embellishing. This theory is similar to some Schenker’s ideas.Next table shows the types of chords, their structure (in number of semitones from the root) and their signs:

Table of chords

Structure of the chord (in number of semitones from the root tone)	**Type of the chord** (name of the chord)	**Sign of the chord** (in output)
4 – 3	Major triad	+
3 – 4	Minor triad	-
4 – 4	Augmented triad	++
3 – 3	Diminished triad	--
4 – 3 –3	Dominant seventh chord	D7 (Maj-7)
3 – 3 – 3	Diminished seventh chord	Dim7
3 – 3 – 4	Diminished/minor seventh chord	Dm7
4 – 3 – 4	Major seventh chord	Maj+7
3 – 4 – 3	Minor seventh chord	Min-7
4 – 4 –3	Augmented seventh chord	Aug+7
3 – 4 – 4	Minor/maj seventh chord	Min+7

4. The difference in **tonal analysis** is great. In older versions the detection and determination of tonal key was made mainly from the key signature at the beginning of the composition. Than the key-change (modulation) was found with appearance of local accidentals. We have used that time also a procedure to decide, if the concrete accidental is only embellishing, or if it belongs to altered chord or if it signs also a change of tonal key.

The main problem started with the input-change, when we shifted from manual to MIDI input. In MIDI we are not able to distinguish the enharmonic tones, and some midi files have no key signature (in our midi sets by Mozart's, Schubert's and Brahms's compositions there are in some compositions similar problems, too.)

Our new idea was, that the listener doesn't see any key signature or accidental, and despite of it she/he is able to hear the tonality. In other words, she/he is able to **distinguish the centre.** So there should be some other way, how to make an algorithm for this detection and determination.

In listener's mind there is **active her/his memory**. This fact we have considered as very important. Therefore, there are three parallel processes in detection and determination of tonal key. One is the easiest – the user can **put in the main key signature** at the beginning of the analysis manually.

The second process is based on memory. We put together more (7 and more) various pitches, which occur at the beginning of the composition. Then they are sorted to create a scale. Then it is compared with the structure of scale of major or minor key. If it is not one of these two scale structures, following pitch in the order from the beginning of the composition is added to the scale. The structure of scale of the major and major key is considered as

different. This is because of the 7th step occurs only in harmonic minor key altered in ascending direction.
The third parallel process in detection and determination of tonal key is based on computing the so-called „**chordal weight**". It is the quantification in the occurrence of every triad, not only in rhythmical values, but also according to the position of this chord in the measure. The first bar beat is considered as more important, so we count with higher value for those chords, which stands on this position in the bar.
Sorry to say, the new tonal analysis doesn't work without problems up to now. It is in it's trial processing and we observe some problems and even mistakes.

In **functional analysis,** it is not possible to determine the tonal function before detection the tonal key. Therefore, this is the last part of the processing.

Having all analyzed results, we proceed with **statistic evaluation** – we create and verify or falsify relevant hypothesae. The most interesting ones will be discussed later.
Let me conclude this software-ideas introduction with the following. It is important to say, that only the first part of the software was used in our demonstration of analysis and statistical evaluation. We made only chordal analysis of three sets of piano composition by three world-famous composers:
There are 27 movements of piano-sonatas by Wolfgang Amadeus Mozart, 33 movements of piano-sonatas by Franz Schubert and 32 piano pieces by Johannes Brahms.

Our hypothesis is, that in a compositional style of each composer it is possible to find some chords, which are **preferred in the composer's subconsciousness.** These chords he used more frequently and/or preferably. For detection of these style features or fingerprint of chordal use we need to search through the complete analysis of more compositions, or, optimally, in the whole amount of all his compositions. We tried it first in three sets of piano compositions.
The results are of various values. Obviously, some are **trivial** and not worth for presentation. Another results are **expectable.** For instance, it is a well-known fact, that the evolution of harmonic speech and harmonic norms developed from the classicism of young Mozart to the new, more complicated harmonic structures in composing of old Brahms. The verification of such expectations is now possible.
Finally, some results are **perhaps of new quality**, which demonstrate the possible preference of some chords of separate composers, even **they show perhaps style features of the compositional style**.

Let me illustrate these findings.
In the first case study, we aimed to test the occurrence of every single chord, then the couples of chords and the software computes also the triples of chords.
First two sets of compositions by Mozart and Schubert are also statistically evaluated in singles and couples. In couples of chords in Brahms music there appeared a problem. The results of chordal analysis are either signs of 11 types of chords or signs of other sounds – signed CC (as compound chord, that means some more complicated, but structure of thirds) and signed X as a beat, where is no structure of thirds. The couple of chords is distinguished only, if they are neighbours. If there is another sound between them, there is no couple. Many compositions by Brahms for this reason had only a few couples. So we didn't reflect these couples.

Next tables and results are therefore the results of statistical evaluations of single chords in music by each of these three composers, the statistical evaluations of couples are computed only for compositions by Mozart and Schubert.
The comparison of the occurrence of chords in these sets brought next results.

RESULTS. There are 16 influences on variables (single and couple of chords) - in percentage of frequencies - of Mozart and Schubert in the Table 1.

- **In the 1st column they are sorted according to Pearson's parametric correlation from the strongest one in the top row to the weakest one in the bottom row. Their influences are higher than 1% level of signification. The positive values mean higher occurrence in Mozart's music, the negative ones are for higher occurrence in Schubert's music. For comparison,**
- **in the 2nd column there are Spearman's correlations,**
- **in the 3rd column - values of F parametric analysis of dispersion,**
- **in the 4th column - values of Z nonparametric analysis of dispersion – the so-called Mann-Whitney's test.**

TABLE 1

	Pearson(R)	Spearman(R)	Analysis of dispersion (F)	Mann-Whitney (Z)
1. /D7/	-0.47	-0.47	16.78	-3.60
2. /D7,D7/	-0.47	-0.46	16.78	-3.57

3. /Maj+7/	0.44	0.44	14.21	3.38
4. /--,Maj+7/	0.41	0.37	11.93	2.86
5. /--/	0.41	0.38	11.61	2.92
6. /+,Dm7/	-0.37	-0.43	9.21	-3.30
7. /Dm7/	-0.37	-0.35	8.97	-2.68
8. /Maj+7,Min-7/	0.36	0.33	8.87	2.55
9. /--,+/	0.36	0.29	8.82	2.25
10. /Maj+7,+/	0.35	0.33	8.26	2.53
11. /D7,Dim7/	-0.34	-0.37	7.62	-2.83
12. /--,--/	0.34	0.28	7.68	2.14
13. /-,Dm7/	-0.33	-0.39	7.06	-3.03
14. /-,--/	0.33	0.29	7.01	2.21
15. /Dm7,D7/	-0.30	-0.37	5.92	-2.83
16. /-,-/	-0.27	-0.34	4.68	-2.59

Arithmetic averages and medians of the same 16 categories of single and couples of chords are compared for two compositional styles – Mozart's and Schubert's in the Table 2.

TABLE 2

Single or couple of chords	average (median) Mozart (%)	average (median) Schubert (%)
1. /D7/	10.73 (10.74)	15.74 (15.89)
2. /D7,D7/	3.35 (2.68)	7.13 (5.95)
3. /Maj+7/	7.90 (7.11)	4.01 (3.38)
4. /--,Maj+7/	0.69 (0.34)	0.11 (0.00)
5. /--/	8.88 (6.45)	4.83 (4.05)
6. /+,Dm7/	0.04 (0.00)	0.47 (0.00)
7. /Dm7/	0.88 (0.90)	1.88 (1.68)
8. /Maj+7,Min-7/	0.60 (0.00)	0.06 (0.00)
9. /--,+/	1.84 (1.55)	0.83 (0.65)
10. /Maj+7,+/	2.85 (1.68)	1.12 (0.85)
11. /D7,Dim7/	0.19 (0.00)	0.59 (0.33)

12. /--,--/	4.03 (3.57)	1.87 (1.10)
13. /-,Dm7/	0.11 (0.00)	0.38 (0.21)
14. /-,--/	1.21 (0.76)	0.58 (0.29)
15. /Dm7,D7/	0.05 (0.00)	0.32 (0.00)
16. /-,-/	8.26 (6.16)	11.81 (9.66)

The statistical differences of both composers are evident:

Mozart's music is typical with higher occurrence of Major 7th chord (Maj+7) and diminished triad (--) both single and in couples. Schubert's music is typical with higher occurrence of Dominant 7th chord (D7) and diminished-minor 7th chord (Dm7) both single and in couples.

CONCLUSION. These evident differences could be interpreted as possible differences between musical style of Vienna classicism and early romanticism, or could be seen as a tendency of evolution of harmony in usage of certain types of chords, too.

Statistical evaluation of percentage of the frequencies of single chords in piano works.

1. Comparison of three composer-styles: Mozart, Schubert and Brahms according to the occurrence of single chords.

Table 3

Sign of Chord	Whole set(N=92) Average(%)	Mozart(N=27) average(%)	Schubert(N=33) average(%)	Brahms (N=32) average(%)
+	38.17	**43.01**	**43.89**	*28.20*
-	19.53	*16.06*	19.80	**22.19**
D7	13.02	*10.73*	**15.74**	12.14
--	7.13	8.88	*4.83*	8.01
Maj+7	6.31	**7.90**	*4.00*	7.33
Min-7	5.61	5.34	*3.46*	8.05
Dim7	3.05	2.52	3.15	3.38
Dm7	2.85	*0.88*	1.88	**5.51**
++	2.56	2.74	1.88	3.12

Min+7	**1.07**	**1.32**	**0.77**	**0.99**
Aug+7	**0.76**	**0.61**	**0.58**	**1.07**

MAXIMUM – *MINIMUM:* **both are strongest style indicators**

These results are vizualized also on the graphs.

For music **by Mozart and Schubert** are significant h**igher frequencies of major triads,**

for music by **Brahms higher frequencies of diminished-minor** seventhchord, **minor–minor** seventhchord and **augmented triad**.

For music **by Mozart** are significant **lower frequencies of dominant seventhchord and minor triad,**

for Schuberts compositions are typical **lower frenquencies of major-major seventhchord and diminished triad.**

For music **by Brahms** are typical **lower frequencies of major triad.**

2. Comparison of two types of keys: major – minor

Occurrence of chords in major and minor compositions

Table 4

Sign of Chord	Whole set (N=92) average	medián	major (N=59) average	medián	minor (N=33) average	medián
+	**38.17**	**39.75**	**42.38**	**45.11**	**30.66**	**30.96**
-	**19.53**	**18.80**	*16.84*	*15.35*	**24.35**	**22.85**
D7	**13.02**	**12.26**	**13.21**	**12.82**	**12.68**	**10.85**
--	**7.13**	**5.73**	*6.20*	*5.12*	**8.77**	**8.70**
Maj+7	**6.31**	**4.88**	**6.52**	**5.02**	**5.93**	**4.24**
Min-7	**5.61**	**4.60**	**6.27**	**4.66**	**4.42**	**2.99**
Dim7	**3.05**	**2.32**	*2.39*	*1.51*	**4.22**	**3.57**
Dm7	**2.85**	**1.97**	**2.02**	**1.26**	**4.34**	**3.95**
++	**2.56**	**1.84**	**2.39**	**1.81**	**2.87**	**1.96**
Min+7	**1.07**	**0.80**	**1.12**	**0.94**	**0.81**	**0.68**

Aug+7	**0.76**	**0.45**	**0.65**	**0.35**	**0.95**	**0.47**

MAXIMUM *-MINIMUM*

One of the interesting facts is well visible in this table – the frequency of major triads is high also in minor compositions. The next research should compute the frequency of major triads in minor compositions by Brahms, while his usage of major triads is lower in whole.

There are many other tables, which indicates the significant influences of some non independent variables on other independent ones.

The next table shows results of stepwise logistic regression.

Table 5

Sign of chord	Mozart	Schubert	Brahms	major	history
konstant	-4.37	-70.76	10.20	-13.30	1.85
+	*	0.84	-0.19	0.20	-0.04
-	*	0.80	*	*	*
D7	*	1.08	-0.27	0.19	-0.04
--	0.44	*	-0.21	0.22	*
Maj+7	*	0.38	*	*	*
Min-7	0.38	*	*	0.60	*
Dim7	*	0.65	-0.32	*	-0.24
Dm7	-1.46	0.64	0.77	-0.38	0.66
++	*	0.60	*	*	-0.15
Min+7	0.82	-1.25	*	0.76	*

Aug+7	*	*	*	*	*
Regres. determ.	**62.30%**	**80.40%**	**73.40%**	**62.60%**	**45.80%**
Correct/incorrect determined comp.	**80/12**	**86/6**	**82/10**	**81/11**	**72/20**

The table shows, that important elements for determination of **Mozart's style** are **quantity presence** of **diminished triad, minor-minor** seventh chord an presence of **minor-major seventh chord**, and strongly quantity of **non /presence of diminished-minor** seventh.

Determination of **Schubert's style depends** mostly on **quantity** of **D7**, and also of **major and minor triads**, and strongly on quantity of **non/presence of minor-major seventh chord.**

Determination of **Brahms style** depends mostly on **quantity of diminishes-minor seventh** and on quantity of **non/presence of major triad, D7, diminished triad and diminished seventh chord**.

There are also some software products, which help to formulate some hypothesis according to gained data. We have used one of these software products named GUHA (implements rules of associations and implications. It was developed in Czechoslovak academy of sciences in Praque by Petr Hájek and other two authors). Next tables bring further hypotheses.

Table 6

approximation **of Mozarts compositions**	approximation **of Schuberts compositions**	approximation **of Brahms comositions**	the best approximation **in 79 hypotheses without composer**
	- in combination with -		
major	**major**	**major**	**major**
even		**even**	**even**
order	**order**	**order**	**order**
allegro	**allegro**		**allegro**

This table shows four approximations of composer's style in combination with major key, meter, historical order and tempo.

The hypotheses according to this combinations are visible in next two tables. The first one shows the conditions (or assumptions) of frequencies of single chords, the second one shows conclusions (combinations).

Table 7: Combinations of determination conditions of compositional style

Sign of Chord	Mozart antecedent:	Schubert antecedent:	Brahms antecedent:	The best antecedent:
+	> 33.5%	> 33.5%	< 45.5%	< 45.5%
-	< 22.2%	< 22.2%	> 15%	< 22.2%
++	< 1.19%	< 1.19%	> 1.19%	> 1.19%
	and > 2.7%	and> 2.7%		
--	> 4.3%	< 8.8%	> 4.3%	< 8.8%
D7	< 15.5%	> 9%	< 15.5%	> 9%
Maj+7	< 3.6%	< 7.5%	> 3.6%	> 3.6%
	and > 7.5%			
Min+7	> 0.35%	< 1.23%	> 0.35%	< 0.35%
				and > 1.23%

Min-7	**< 6%**	**< 3.6%**	**> 3.6%**	**> 3.6%**
		and > 6%		
Dim7	**< 1.2%**	**> 1.2%**	**> 3.5%**	**> 1.2%**
	and > 3.5%			
Dm7	**< 3%**	**< 0.95%**	**< 0.95%**	**> 0.95%**
		and > 3%		
Aug+7	**0%**	**> 0%**	**> 0%**	**> 0.69%**
	and > 0.69%			

Table 8: Conclusions (findings)

combinations	**sukcedent:**	**sukcedent:**	**sukcedent:**	**sukcedent:**
composer	**Mozart**	**Schubert**	**Brahms**	*****
major composit.	**yes**	**yes**	**no**	**yes**
even metrum	**yes**	*****	**no**	**no**
order of opuses	**lower**	**lower**	**higher**	**higher**
allegro	**yes**	**yes**	*****	**no**

evaluation of 4 hypotheses:

type of combination	**number of compositions**			
	Mozart	**Schubert**	**Brahms**	
signification:	**0.000001**	**0.00006**	**0.00003**	**0.00005**
A) antecedent				

and sukcedent:	**4 comp.**	**4 comp.**	**4 comp.**	**5 comp.**
B) non antecedent				
and non sukcedent:	**83 comp.**	**84 comp.**	**84 comp.**	**86 comp.**
C) non antecedent				
and sukcedent:	**5 comp.**	**3 comp.**	**4 comp.**	**1 comp.**
D) antecedent				
and non sukcedent:	**0 comp.**	**1 comp.**	**0 comp.**	**0 comp.**

From D) it is evident, that the effectiveness of the method is quite precise in sense of validity of logic implication "if antecedent then succedent". The only one composition by Schubert is not possible to determine with these conditions and conclusions.

Thank you for your attention.