COMPUTER ANALYSIS OF MUSICAL SOUND AND PROBLEMS OF RESULT NOTATION

Alexander Kharuto

Abstract: Contemporary computer techniques of sound pitch analysis provide results with accuracy near to human’s resolution of hearing. Further precise examination of traditional music performance requires fixation of sound pitch variations. Usually, European musicologists apply for this purpose musical notation (score). This kind of fixation means quantization on sound pitch, i.e. rounding of exact values to ‘allowed’ pitch levels, and also quantization of sound durations. Melody notation in academic music uses 12-halftone equal-tempered system. There exist also systems with more stages in octave. For durations, a row of time units like 1/32, or 1/16, or 1/8… will be used — with certain main tempo. In the paper, quantization error values will be analyzed in dependence of number of pitch row stages in octave, used time units and different character of melody line.

Key words: musical sound, sound pitch, sound duration, notation, sign system, computer analysis, measurement accuracy, quantization, error

1. Computer sound pitch analysis

In musicological investigation of traditional music performances (folkmusic, religious rites etc.), one of the main tasks is the recovery of entire melody line, which means precise sound pitch extraction and estimation of duration and intensity (relative loudness) of sound. Pitch analysis can be fulfilled with help of various methods [1], based on auto-correlation function of sound oscillations [2], or calculation of weighted quadratic difference with further numerical minimization of it upon the duration of supposed period, as in YIN algorithm [3], or cepstrum calculation method, which proposes direct and reverse Fourier transforms of sound oscillations with logarithmic transformation of spectrum after the first step. Human’s perceiving of sound pitch provides accuracy about 4..5 musical cents (100 cents = 1 halftone, 1200 cents = 1 octave). By using of well-known methods listed above, the required accuracy will be achieved under sufficient signal-to-noise ratio (about 10 dB or more) and suitable form and size of ‘analysis window’ in time domain used for ‘sliding’ Fourier transform [4]. An example of calculated melody line (Russian folk song) positioned upon the 12-halftone system is shown on fig. 1.
2. Sound pitch notation

Graphical melody representation (named 'melo-
gram') like shown on fig. 1 is incomprehensible for
many musicians. For fixation of investigation re-
results, usual musical notation will be used. The most
habitual for academic musicians is the 12-halftone
equal-tempered notation system. (But also 24-lev-
el, 48-level and other notation systems with larger
number of stages in octave can be used.) These
notation systems have certain sets of ‘allowed’
sound pitch stages, which can be used for melody
fixation. In order to transform ‘source’ melogram
into a chain of tones belonging to such set, every
sound pitch will be rounded to nearest pitch stage
of notation system. An example of such melogram
conversion for the fragment on fig. 1 into tones of
12-halftone pitch row (thick horizontal lines) is
shown on fig. 2. During this operation, some de-
dtails of melody line will be missed, but differences
between true melody line and rounded sound pitch
values, i.e. quantization errors, will be neglected
by storing in musical score.

Error value for such case will not exceed \( \frac{1}{2} \) of
distance between two adjacent ‘allowed’ stages:

\[ e_{pt} \leq \frac{1200}{(2N)}, \quad (1) \]

where \( N \) is the number of stages in the pitch row
used for notation of melody; \( e_{pt} \) will be calculated
in musical cents.

These maximal error values are represented on
fig. 3 as upper solid line (‘worst case’). For keeping
quantization error under the ‘perception value’ 4..5
cents, \( N \) must be over 120..150; but notation system
with such large number of signs (or combination
of signs) used for denotation of sound pitch values
will not be suitable for practice.

In order to control the quantization process, a
real quantization error can be calculated for every
sound pitch value. Pitch quantization errors for
entire melody can be characterized through root-
mean-square deviation (RMSD) of source sound
pitch from rounded values:

\[ s_{pt} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (M_i - \Delta_m \times \left[ \frac{M_i}{\Delta_m} \right])^2}, \quad (2) \]

where \( n \) is the number of estimated sound pitch
points, \( M_i \) is the pitch value number \( 'i' \), \( i=1,2,3,...,n; \)
\([x]\) denotes the integer part of \( 'x' \), \( \Delta_m = 1200/N, \) \( N \)
is number of sound pitch stages in octave used in
given notation system.

The achieved accuracy depends on \( N \) and also
on melody line character. Graphical representations
of some examples of traditional vocal music from
different areas of Russia are positioned between
bounds shown on fig. 3.

Another parameter which influences quantiza-
tion error is sound pitch shift of melogram position
against the set of pitch row stages. Estimations

Fig. 2
of error values for all the phonogram examples have been obtained for optimal determined sound pitch shift of the whole set of ‘allowed’ sound pitch values against source melogram. The appropriated optimal shift is dependent on N. Amplitude of shift variations is always smaller than ∆m. Typical variations of optimal shift for different values of N are represented on fig. 4 for an example of phonogram (Russian traditional song).

In order to keep in score some details like ‘unevenness’ of melody line (which may be characteristic for the given traditional culture), musicologists must choose the needed number of stages in octave on the base of calculating RMSD for every examined melogram with optimization of sound pitch shift (for every N). All these procedures — calculating of melogram, estimation of RMSD and searching of optimal sound pitch shift — will be provided by author’s program SPAX [5].

Let’s note, that for instruments with fixed tuning the function $s_{pt}(N)$ has a special character: it contains a row of minimums (for pianoforte — at the points $N=12\times i$, $i=1,2,…$), as shown on fig. 5. Other minimums on fig. 5 appear because of some detuning of the given instrument and estimation errors.

### 3. Notation of tone durations

For fixation of melody line with help of some notation system, also tone durations must be written down. The sound pitch rounding described above makes also some kind of segmentation of melody line: every detected tone (thick lines on fig. 2) has certain duration. But in order to make notation (musical score), these durations must be rounded to some ‘allowed’ values (duration quantization). The system of durations in European academic music contains a row of values based on proportion 1:2 (1:2, 1:4, 1:8,… 1:64… of ‘base’ duration which corresponds to ‘1’). The ‘base’ time unit will be given
by musical tempo. For an often used tempo of 120 beats per minute, the ‘base’ duration (1) is equal to 2 s, the ‘1:2’ continues 1 s, ‘1:4’ — 0.5 s, … ‘1:64’ continues 62.5 ms and so on. From the viewpoint of quantization accuracy, the shortest note duration (like 1:128) would be the best time unit for duration notation of computer analysis results.

Computer estimation of sound pitch uses time step about 5 ms: smaller steps are not suitable because widths limitation of ‘sliding’ analysis window in time domain — they must be about 20..40 ms [4]. With such time step, tone durations are estimated with some possible errors — limited from above by ±2.5 ms. The representation of tone with duration 1:64 may have a relative error until 4%. It is evident, that shorter note durations are not proper for fixation of tone durations in notation.

For many phonograms of traditional culture (folksongs from different areas of Russia), relative quantization errors are positioned approximately between 0.2 and 0.35 from the measured duration. (By these measurements, all durations lower than used time unit will be neglected.) Some examples of dependence of relative error of duration quantization are shown on fig. 6. The represented errors are remarkable large (more than 18%). This fact indicates that real sound durations in traditional performances build not any row of proportional values: rather, real durations contain considerable accidental parts.

4. Problems of notation accuracy and contradiction of sound pitch- and duration fixation

Notation of results of computer analysis for traditional musical performance requires quantization of both parameters — sound pitch and sound durations; the number of stages $N$ in octave must be chosen and also the time unit $T_0$ for notation of durations.

Growing of number of sound pitch stages $N$ leads to decreasing of quantization errors for sound pitch (see fig. 3). But it also provokes increasing of number of detected ‘tones’ and shorten of average tone duration. An example of such transformation for $N = 24$ is shown on fig. 7 (compare with fig. 2 where $N = 12$).

For more adequate representation of results, calculations of errors must be made for different stage numbers $N$ and different time units $T_0$. This procedure will provide objective controlling of quality of result fixation (which is traditionally substituted through ‘intuitive’ estimation of musicians based on hearing). The author’s program SPAX [5] contains all the needed tools for calculations.

Fig. 6
References


Information about the author
Alexander Vitalievitch Kharuto
Head of Department of Musical Informatics (since 1994) in Moscow P. I. Thaikovsky Conservatory.
Bolshaya Nikitskaya str., 13, Moscow, Russia, 125009.
e-mail: akharuto@yahoo.com
Web site: www.kharuto.ru