Dept. for Speech, Music and Hearing
Quarterly Progress and Status Report

Automatic notation of played music (status report)

Askenfelt, A.

journal: STL-QPSR
volume: 17
number: 1
year: 1976
pages: 001-011

KTH Computer Science and Communication

http://www.speech.kth.se/qpsr

I. MUSICAL ACOUSTICS

A. AUTOMATIC NOTATION OF PLAYED MUSIC (STATUS REPORT)

A. Askenfelt

Abstract

A computer program automatically converting sounding music into written music is presented. A hardware system is used for the identification of rests and for pitch detection. A statistical treatment of the pitch periods serves as the basis for the identification of the actual scale used in each individual melody. An analogous treatment of the durations defines the note values. The transcription obtained is presented in ordinary notation. The method used seems promising but further development is needed, especially as regards the pitch detection.

1. Introduction

In Sweden, folk music has been collected on tape systematically for the last 25 years, resulting in a large collection of tape recordings. The work is carried out by The Swedish Center for Folksong and Folk Music Research. Today this collection contains more than 25.000 melodies and the number increases with about 3000 melodies/year.

Folk music is interesting for various reasons.

1) Some melodies are beautiful pieces of music. They are still played and listened to for the same reasons as the so-called "classical" works.
2) Folk tunes are often very old. Therefore they may tell us about the conditions under which man lived in the past. In this way a collection of folk tunes is a contribution to the history of culture, comparable with collections of paintings, books, clothes, etc.
3) A large collection of folk music is a rich research object for music theorists.

Only rarely is the folk music available in written form. The tunes are taken over from one player to another directly without the aid of any notation.

However, the accessibility of the items in a collection of this magnitude requires that the melodies are available in ordinary notation. Other systems possible for notation of music such as numerical codes, graphical representation as obtained from a "melody writer" etc, have advantages when used for special purposes; but in most cases the conventional notation is preferable not least because the users of this collection including musicologists are familiar with and even tend to think in terms of ordinary notation. The material must also be stored in such a way that it is possible to make rapid searches through the melodies, using even rather complex search criteria.

The requirements mentioned above seem to make the task of notating and storing the melodies suitable for a computer. Attempts with automatic notation of played music have been made earlier[1]. This paper reports the development of computer program and hardware equipment for this purpose during 1975.

2. Hardware equipment

The hardware equipment used in the project is seen in Fig. I-A-1. The tape signal passes first a Rest Detector and Oscillator (RDO) which

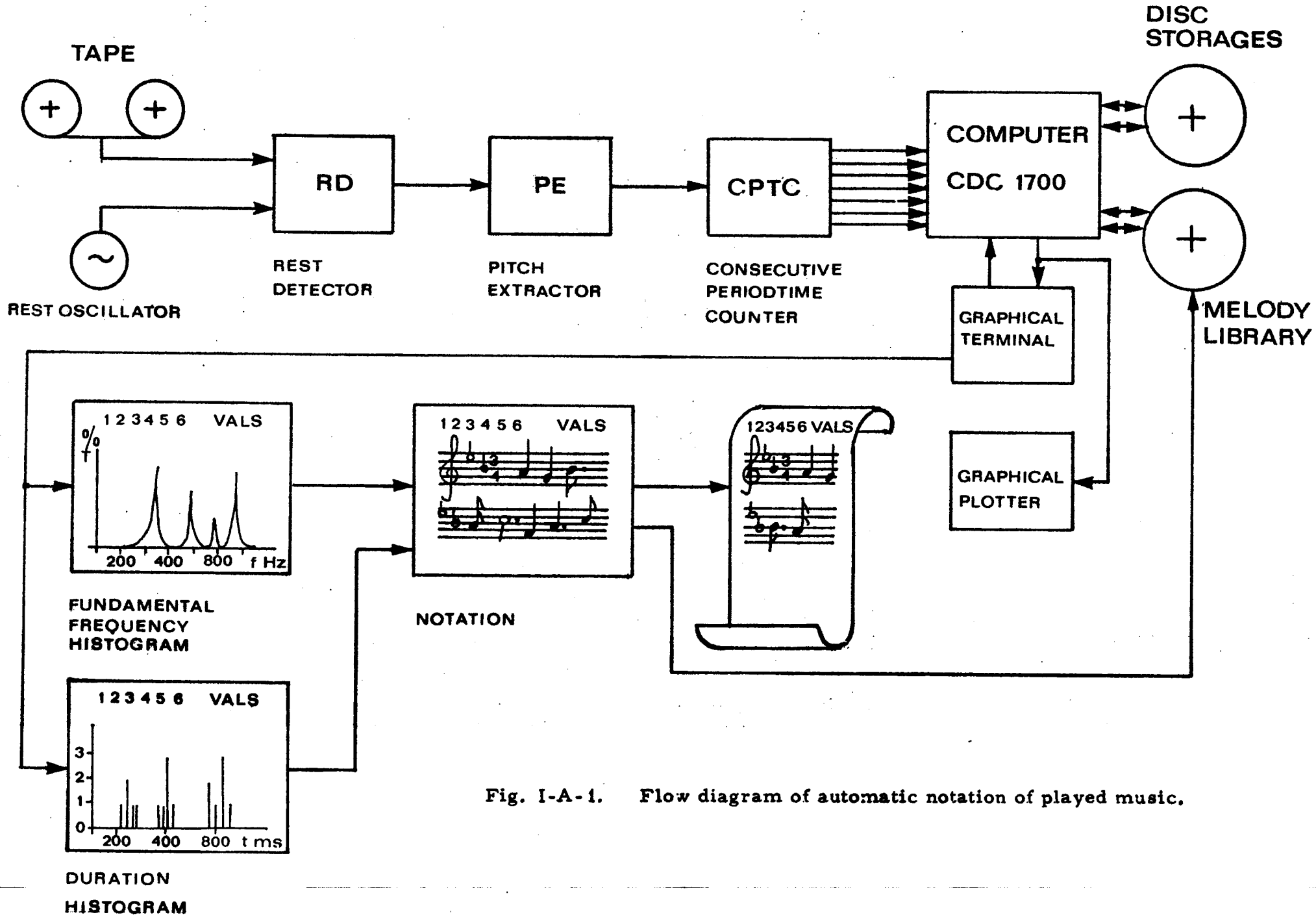


Fig. I-A-1. Flow diagram of automatic notation of played music.

attempts to identify pauses. The signal from the RDO is fed into a Pitch Extractor (PE) which extracts the fundamental frequency out of the complex signal and converts it into a pulse train. Pitch detection is a notorious problem in music and speech research. Although sophisticated and reliable methods for pitch detection by means of computer now are avialable, they are not perfectly suited in this case, because they are still rather time consuming. It is important that the computer can provide a notation of a melody at least as rapid as a skilled professional. Thus, it is not acceptable to wait 5 or 10 times real time only for pitch detection processing. Therefore two hardware systems for real time pitch detection are being tested at the moment. One enhances the pitch periodicity in the time domain while the other accomplishes spectrum flattening in the frequency domain.

2.1 Pitch detectors

a) TRIO

This method is based on the well known trigonometrical identity $\sin^2 x + \cos^2 x = 1$. Thus the name TRIO ("The Trigonometrical 1"). Consider one period of a bandpass filtered periodic signal where only one formant remains, i.e. the output of a pulse excited resonance circuit. The time function u(t) in Fig. I-A-2 is then expressed by

$$u(t) = Ae^{-\beta t}\cos\omega_n t \qquad (1)$$

where $\omega_n = 2\pi F_n$ $\quad F_n$ = formant frequency

Processing this signal in the network in Fig. I-A-3 gives:
Phase shifting Eq. (1) 90°

$$v(t) = Ae^{-\beta t}\sin\omega_n t \qquad (2)$$

Squaring (1) and (2) and adding forms:

$$z(t) = A^2 e^{-2\beta t}(\sin^2\omega_n t + \cos^2\omega_n t) = A^2 e^{-2\beta t} \qquad (3)$$

Thus, this processing enhances the pitch periodicity of the signal and ideally only a stressed envelope remains. In reality the assumption of a pulse excited resonator is not correct. Furthermore it is sometimes difficult to isolate one formant in music or speech with simple means. However, despite these difficulties the method seems promising. Typical examples of performance are shown in Fig. I-A-4.

The TRIO is followed by an ordinary peak detector, the highest peaks triggering a monostable flip-flop, which produces the pulse train.

Fig. I-A-2. Idealized TRIO signal waveforms. u[t] = bandpass filtered signal, z[t] = processed signal

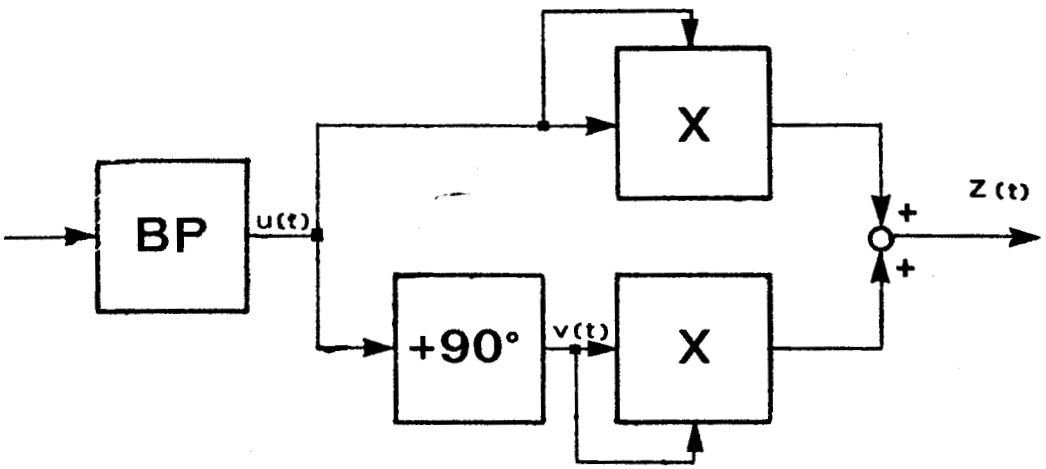


Fig. I-A-3. Block diagram of the TRIO. BP = bandpass filter, + 90° = phase shifter, x = multiplier.

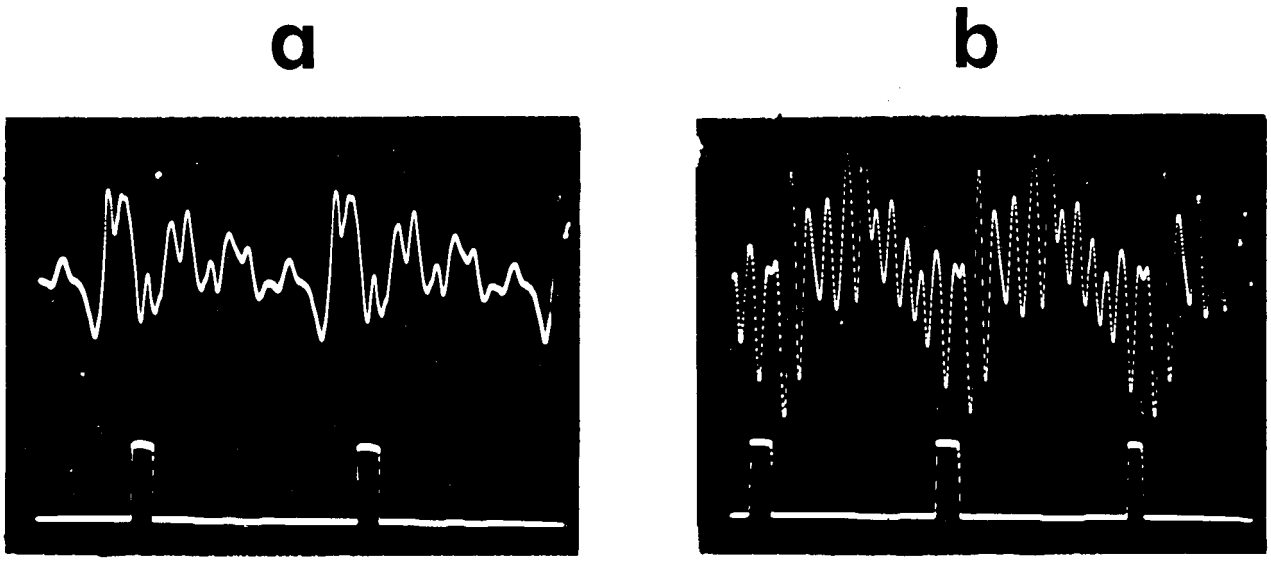


Fig. I-A-4. Typical performance of the TRIO.
Upper trace: Unprocessed signal.
Lower trace: Signal processed by TRIO and peak detector.

a: Male voice [ø], F_o = 125 Hz. b: Violin, pitch D_4, F_o = 294 Hz.

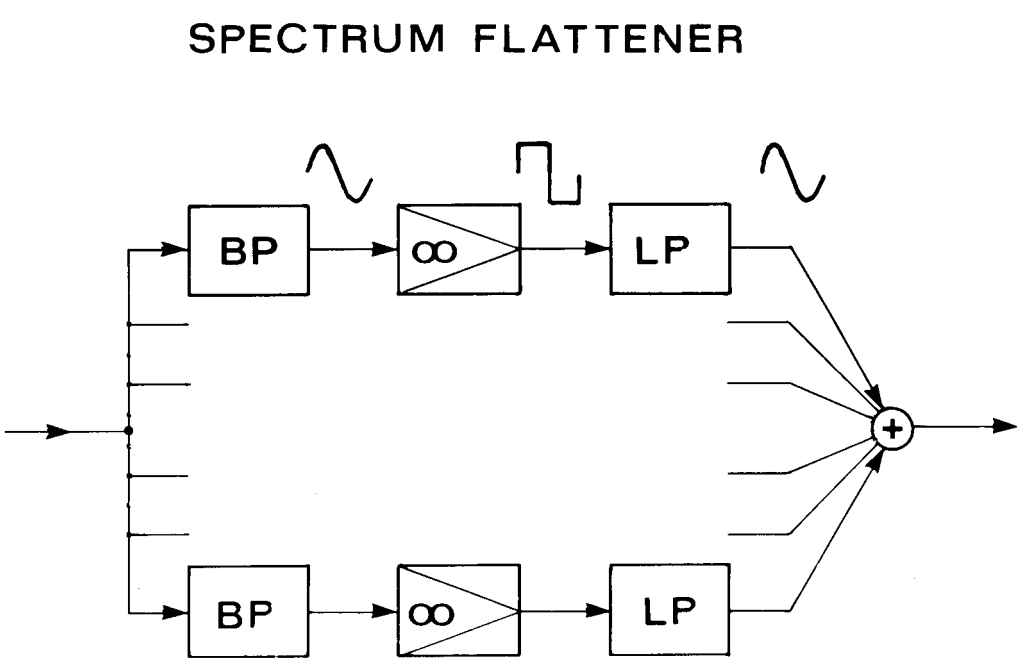


Fig. I-A-5. Block diagram of the spectrum flattener.
Each of the six channels consists of bandpass filter, clipper, and lowpass filter. See text.

b) Spectrum flattening[2]

Spectra of musical sounds very often show strong partials relative to the fundamental. If the harmonics of the fundamental frequency could be made equal in amplitude, the pitch periodicity would be more apparent. Moreover, if the partials also could be put into phase synchrony with each other, the result would be a train of highly peaked pulses. A block diagram of a circuit which dynamically adapts to a varying spectrum and in every moment produces a flat spectrum is shown in Fig. I-A-5.

The signal is filtered through a bank of six bandpass filters adjusted to cover the frequency range of interest. The BP-filters are one-third octave-band filters of six-pole Butterworth design. The signal at the output of each filter is infinitely clipped and refiltered through lowpass filters to recover sinusoidals. These LP-filters are fourth order Butterworth filters. The output of the six channels are thus the original spectral components, normalized to unit amplitude. Summing the channels gives the desired flat spectrum signal, which presumably shows a train of rather peaked pulses, pitch markers[3]. The spectrum flattening circuit is followed by the peak detector mentioned earlier.

Typical performance of the spectrum flattener is shown in Fig. I-A-6a. Less convincing performance, such as in Fig. I-A-6b, is due to partials which are present in two channels at the same time. This occurs if a partial happens to fall on the boundary between two bandpass filters.

Presently the TRIO seems to be more promising than the spectrum flattening method. The spectrum flattener must be provided with narrower and steeper BP-filters in order to improve the performance. It may even prove desirable to complete the spectrum flattener with phase synchronization.

2.2 Consecutive Period Time Counter (CPTC)

The information on the fundamental frequency as provided by the pulse train from the pitch extractor is converted to binary numbers in the CPTC. Thus, the CPTC can be regarded as an A/D converter. At the end of each pitch period the periodtime is presented at the digital output of the CPTC together with an interrupt signal to the computer. The circuit diagram for the CPTC is shown in Fig. I-A-7. The clock frequency is 2 MHz offering a time resolution of 0.5 μs. The frequency resolution is 0.5 Hz at 1000 Hz (0.5 o/oo) rising to 0.005 Hz at 100 Hz (0.05 o/oo).

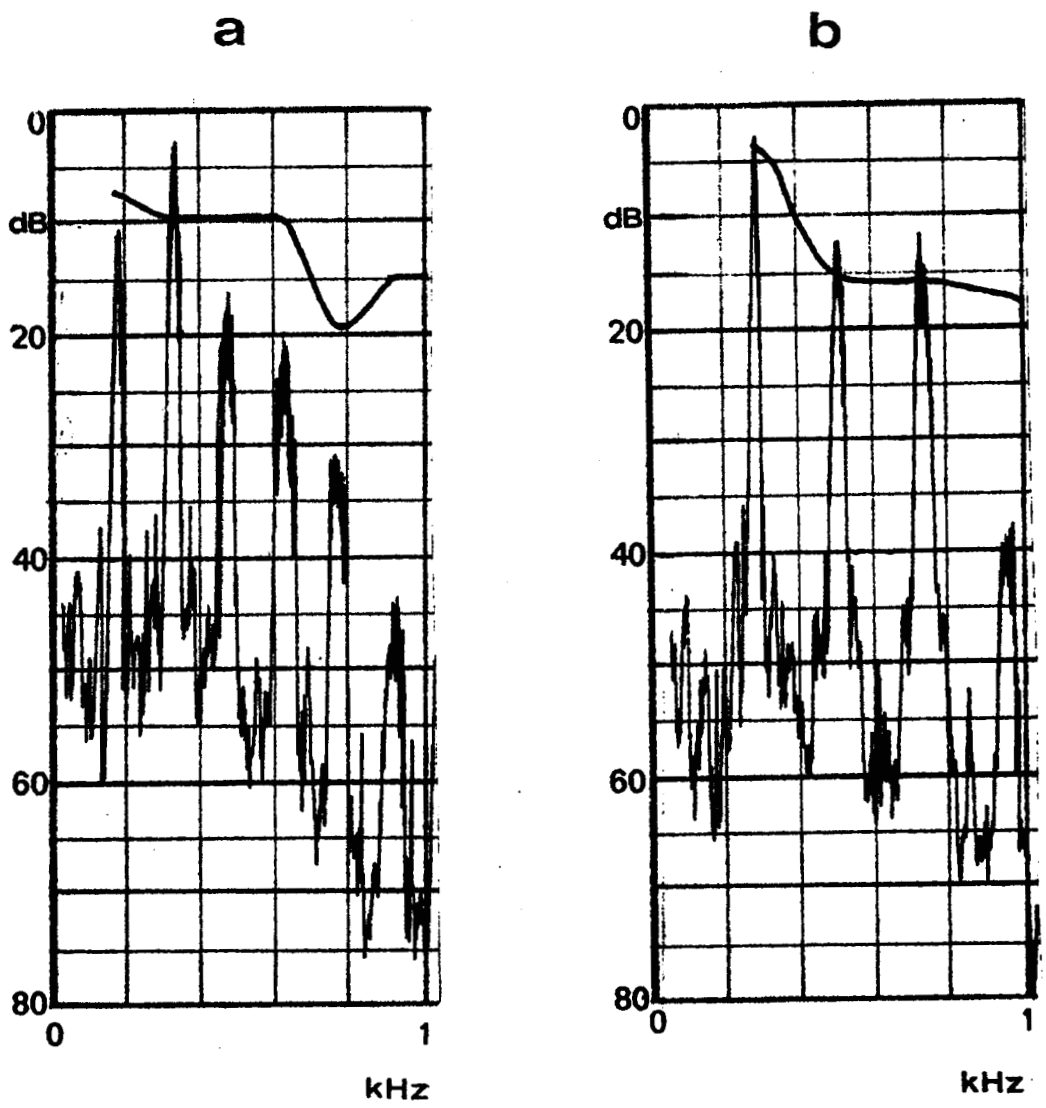
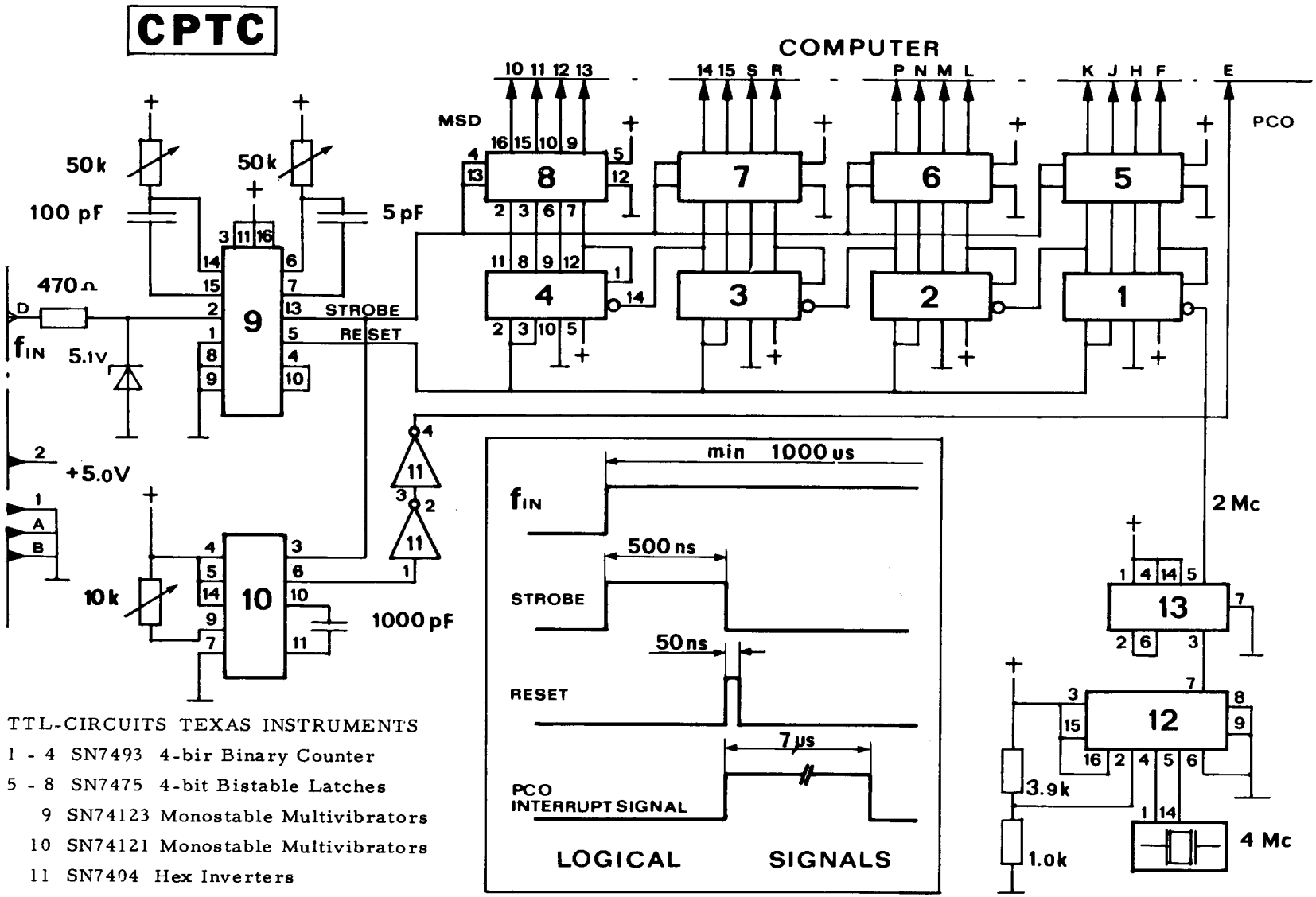


Fig. I-A-6. Spectra of violin tones. The solid line is the spectrum envelope after the spectrum flattening.

a: Pitch D_3, F_o = 147 Hz (obtained from D_4 by halving the tape speed)

b: Pitch B_3, F_o = 247 Hz



TTL-CIRCUITS TEXAS INSTRUMENTS

1 - 4 SN7493 4-bir Binary Counter
5 - 8 SN7475 4-bit Bistable Latches
9 SN74123 Monostable Multivibrators
10 SN74121 Monostable Multivibrators
11 SN7404 Hex Inverters
12 SN74124 Voltage Controlled Osc.
13 SN7474 Dual D-type Flip-Flops

Fig. I-A-7. Circuit diagram and logical signals of the CPTC.

2.3 Computer equipment

The computer used in the project is a CDC 1700.

Basic specifications:

Memory cycle	1.1 μs
Word length	16 bits
Core memory	16 k
Disc memory	2 x 1.5 M

The programs are written in assembler language.

3. Computer program

The program for automatic notation is most easily described by following the processing of a specific melody from sound to notes printed on a staff. The melody shown in Fig. I-A-8 as played on a recorder will serve as an example. The overall strategy is based on a statistical processing of the events in the music, first concerning the frequency domain and then, utilizing the results thereby obtained, the time domain. The statistical results serve as a template for the quantization of the musical events in the melody.

3.1 The Fundamental Frequency Histogram

The tape with the melody is played on a tape recorder. All pitch period times provided by the hardware equipment earlier described are recorded on the computer's disc storage. Nearly one million words are saved on the disc storage for this purpose, which corresponds to a maximum recording time of 750 sec = 12 min at 1000 Hz. The computer converts the period times into frequency values and assorts the pitch periods according to their frequencies in a fundamental frequency histogram (FFH). The FFH is presented to the operator on the screen of the graphic terminal. The FFH of the melody in the example is shown in Fig. I-A-9. Frequencies occurring frequently in the melody, i.e. the scale tone frequencies, appear as more or less pronounced peaks along the logarithmical frequency axis. The y-coordinate corresponds to time. The number of recorded periods as well as the number of periods falling outside the allowed frequency range (70-1000 Hz) are also given in the FFH. Any part of this frequency range can be expanded to fill up the screen. The curve can be smoothed with an averaging algorithm in order to remove ripple in the contour. If so, each histogram

FISKESKÄRSVISAN

Fig. I-A-8. "Fiskeskärsvisan" in conventional notation.

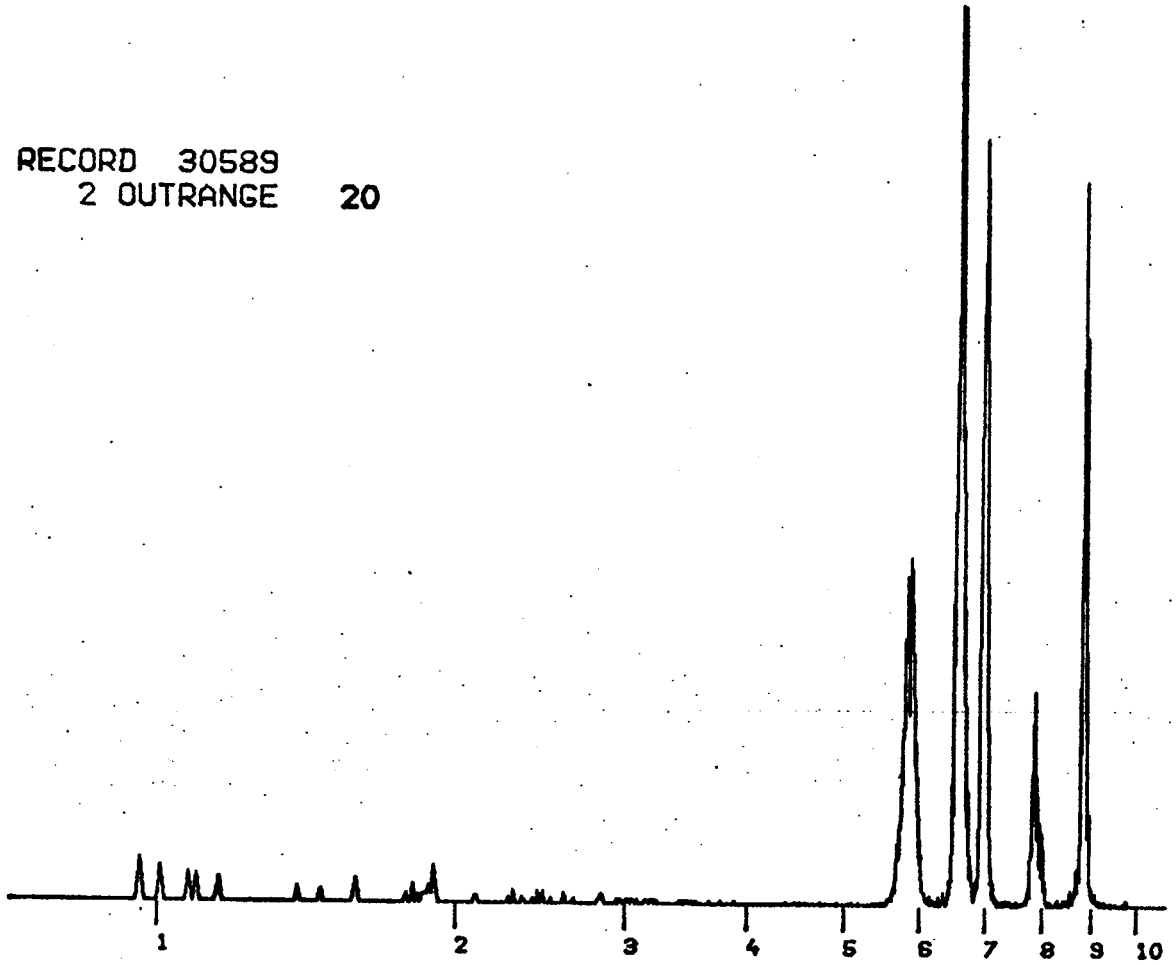


Fig. I-A-9. Unprocessed FFH of "Fiskeskärsvisan". Frequency scale in Hz x 100.
RECORD = number of recorded periods.
OUTRANGE = number of periods falling outside the lower and the upper limit respective of the allowed frequency range 70-1000 Hz.

value f_n is modified by

$$f_n = \frac{f_{n-1} + 2f_n + f_{n+1}}{4}$$ c.f.Fig. I-A-10a and 10b.

It is now possible to define a scale which is perfectly adapted to the actual melody played. The operator points at the scale tones, i.e. the peaks in the FFH, with a movable cursor line, thereby defining the center frequencies of the scale tones. The tonic and possible accidentals are marked specially. A pattern corresponding to the scale tone frequencies of an equally tempered major- or minor scale is available in order to assist in the identification of the peaks, Fig. I-A-11.

When this has been accomplished by the operator, a processed FFH is presented, Fig. I-A-12. The frequency axis is now divided into partitions, each corresponding to a scale tone. All frequencies within such a partition is considered as representing a specific scale tone. Whenever the fundamental frequency of the melody falls within one such tone partition, the melody is assumed to dwell on the corresponding scale tone. The limits between these partitions are located at the geometrical mean of two adjacent center frequencies. In most cases these limits agree with the absolute minima between two neighboring peaks.

The FFH also gives the intervals of the scale tones relative to the tonic in cents. The extension index β tells how great a per cent of the total recorded time the melody spent in that particular scale tone partition, i.e. to what extent the melody used that particular scale tone. The intonation index α is a measure of the distribution of a peak within a tone partition. This index is obtained by transforming the peak into a symmetrical triangle peak having the same height and area as the original peak. The base of this triangle, expressed in cents, is called intonation index α. It informs about the distribution of the peak, e.g. the acuity of the intonation. This information mentioned above is often useful in folk music research. Also, the sounding part of the recorded time as well as the frequency of the tonic and the type of scale pattern used are given in the FFH.

The main purpose of the FFH is to establish a notation staff which is optionally adapted to the melody actually played, i.e. the frequency definition of the lines and spaces of the notation staff are determined in

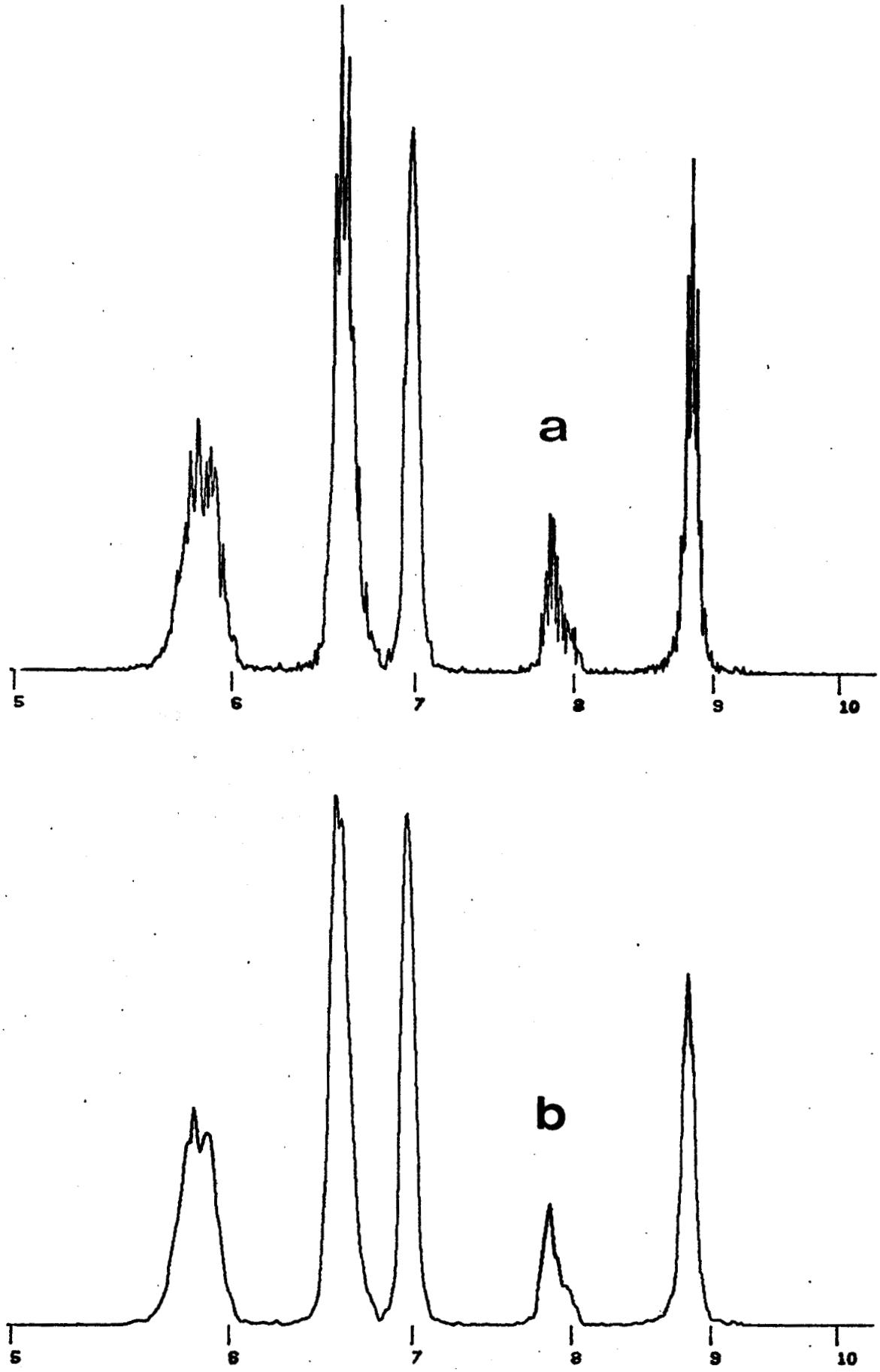


Fig. I-A-10. a: FFH not smoothed.

b: FFH smoothed once.
Frequency scale in Hz x 100.

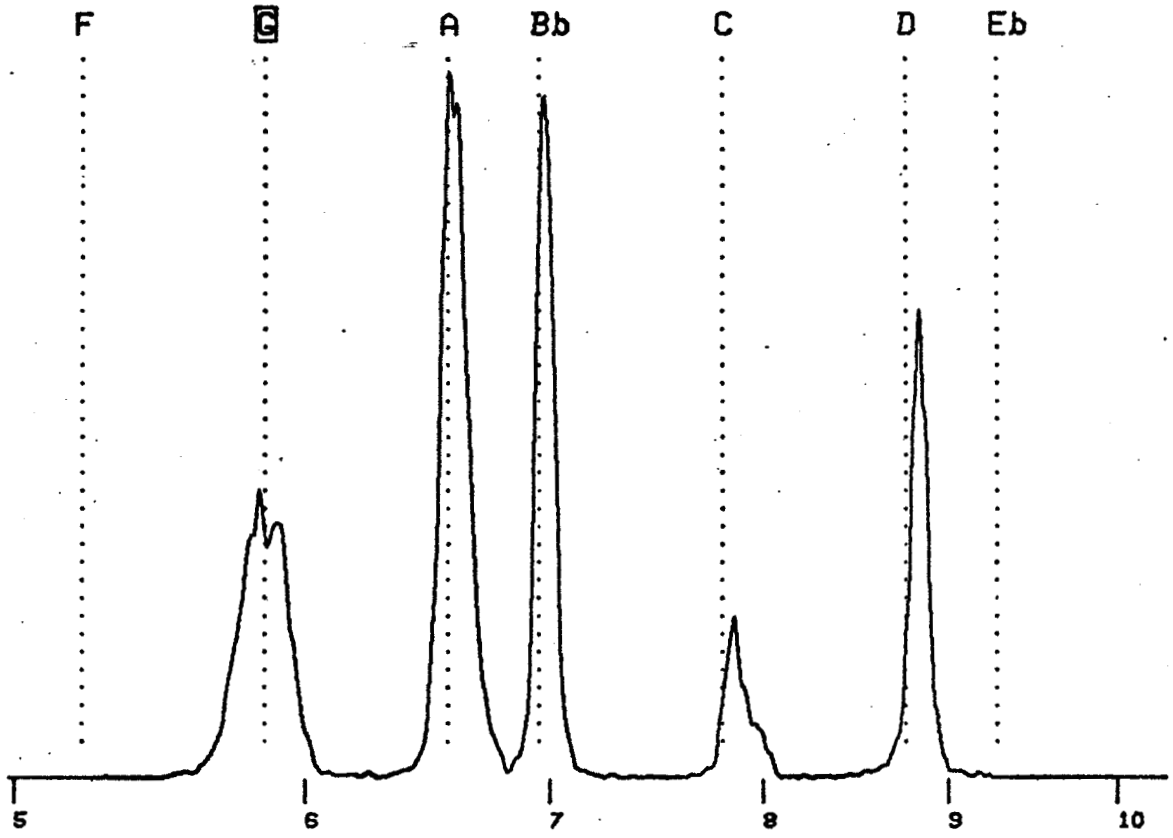


Fig. I-A-11. Pattern of an equally tempered G-minor scale matched with respect to the tonic (G) peak in the FFH.

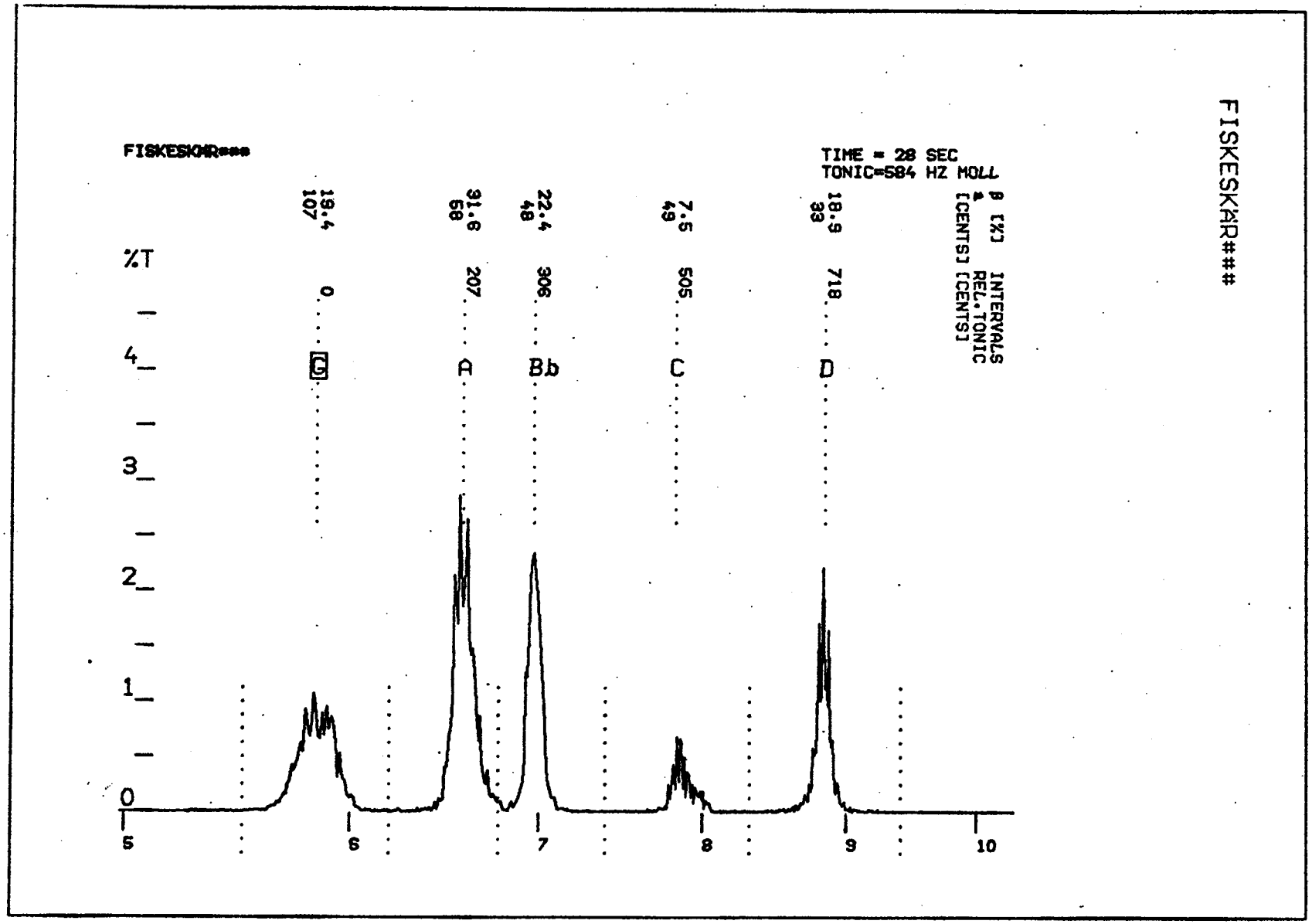


Fig. I-A-12. Processed FFH of "Fiskeskärsvisan". The numbers are explained in the text. Frequency scale in Hz x 100.

accordance with the distribution of fundamental frequencies. The intervals between the lines and spaces of this staff may thus differ more or less from the corresponding intervals of the equally tempered scale. However, it is essential that the distance between e.g. two lines or two spaces represents an interval at least resembling a third. This may necessitate insertion of scale tone frequencies in the FFH even in places where there are no peaks. This applies as soon as a scale tone represented on the staff is missing in the melody. In detecting such cases, the operator is efficiently assisted by the scale pattern mentioned above.

The FFH presents complete statistics over the frequencies of all pitch periods in the melody processed, and thus gives an average of the scale used during the melody and nothing more. If the player changes the scale during the melody, i.e. if a certain scale tone is not always represented by the same frequency value whenever it occurs, the peaks in the FFH will broaden and eventually merge. It may be impossible to define a useful set of scale tone frequencies on the basis of such an FFH. If, however, the scale changes abruptly or continuously while the intervals between the scale tone frequencies remain the same, it will be possible to track the scale and restore it to the starting pitch. If, on the other hand, even the intervals change considerably, e.g. depending on their position in the musical context, no scale can be defined on the basis of the FFH. However, in such cases, the idea of a musical scale must be said to be ambiguous, and one would question if it is meaningful to force such music into the frame work of conventional notation.

The remarks above have concerned the FFH as a basis for notation. However, as mentioned earlier, the FFH itself can serve as a powerful tool for measuring e.g. scales, intervals, tone distribution in a melody and other factors of relevance to the field of ethnomusicology.

3.2 The Duration Histogram

Given the information contained in the FFH on the frequency limits between adjacent scale tones, the recorded string of period times can be transformed into a string of pitch quantified melody tones. The duration of these tones are obtained by summing up all period times belonging to a given tone. Boundaries between consecutive melody tones is detected either as a change of the scale tone, or in the case of tone repetitions, by discontinuities in the period time values due to transition.

The durations of the melody tones can now be assorted into a duration histogram (DH), analogous to the FFH. Here the logarithmical x-axis represents time and the y-axis the number of melody tones. The DH for the melody in the example is shown in Fig. I-A-13. The melody tone durations are generally distributed in groups corresponding to half notes, quarter notes, eight notes, etc. The time resolution is 3 % and the allowed duration range 50-2500 ms.

The operator now has to define a typical time value for one group of durations. This is easily accomplished by means of the cursor line mentioned. Furthermore, the operator decides which type of note value this group represents. After this has been accomplished, the computer defines the center time values for all remaining note values and divides the time axis into duration partitions, each of which corresponds to a note value, Fig. I-A-14. This is done in accordance with the nominal time values, i. e. a half note is twice as long as a quarter note etc. Using this information the computer now quantifies the durations of the melody tones. All melody tones falling into a given duration partition will be represented by the note value assigned to this partition. The result is a string of pairs of quantified numbers, i. e. the played melody has been transformed into a succession of certain states with fixed pitches and durations, in other words, into a notation.

The main purpose of the DH is to establish a basis for the coming notation. The DH is a complete statistical treatment of the durations of the melody tones. However, this treatment is considerably more sensitive to errors than the FFH, because of the small size of the underlying statistical material, in this example about 30 notes. Also, the player often changes the tempo during the melody. In such cases the groups corresponding to the different note values will merge. At worst the DH will assume the shape of a splint fence. This can be avoided if only a part of the melody with essentially constant tempo is selected for processing, but this leads to an even smaller statistical material resulting in increased uncertainty. Thus, there are certain problems associated with the DH as the basis for the quantization of the durations of the melody tones. A method which combines the features of a statical DH over the whole melody with a momentary DH which dynamically adapts to tempo changes, seems more promising. It is also important

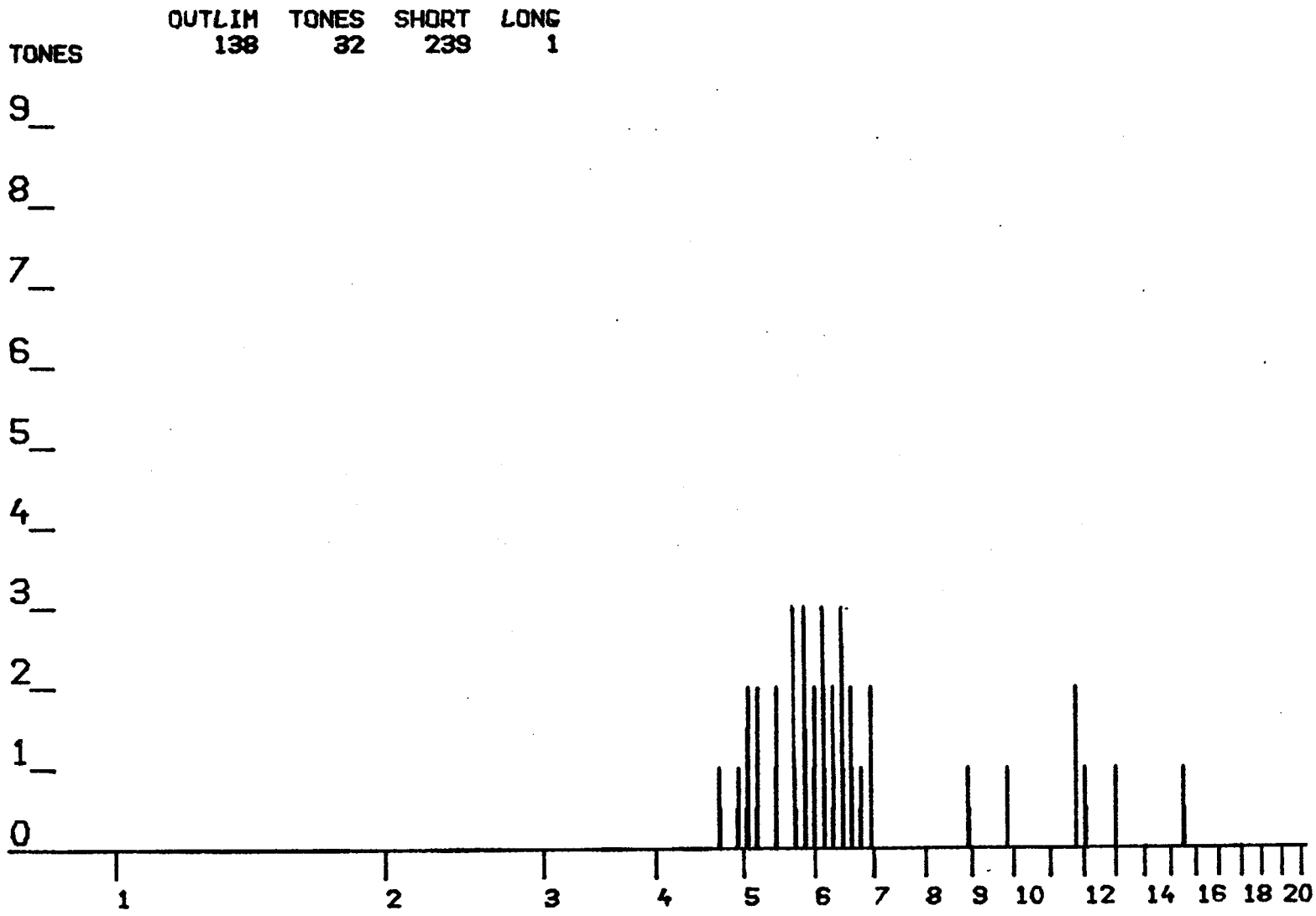


Fig. I-A-13. Unprocessed DH of "Fiskeskärsvisan".
Time scale in sec x 0.1.
OUTLIM = number of periods falling outside the extreme frequency limits in the FFH.
TONES = number of tones in the DH.
SHORT = number of tones with durations shorter than 50 ms.
LONG = number of tones with durations longer than 2500 ms.

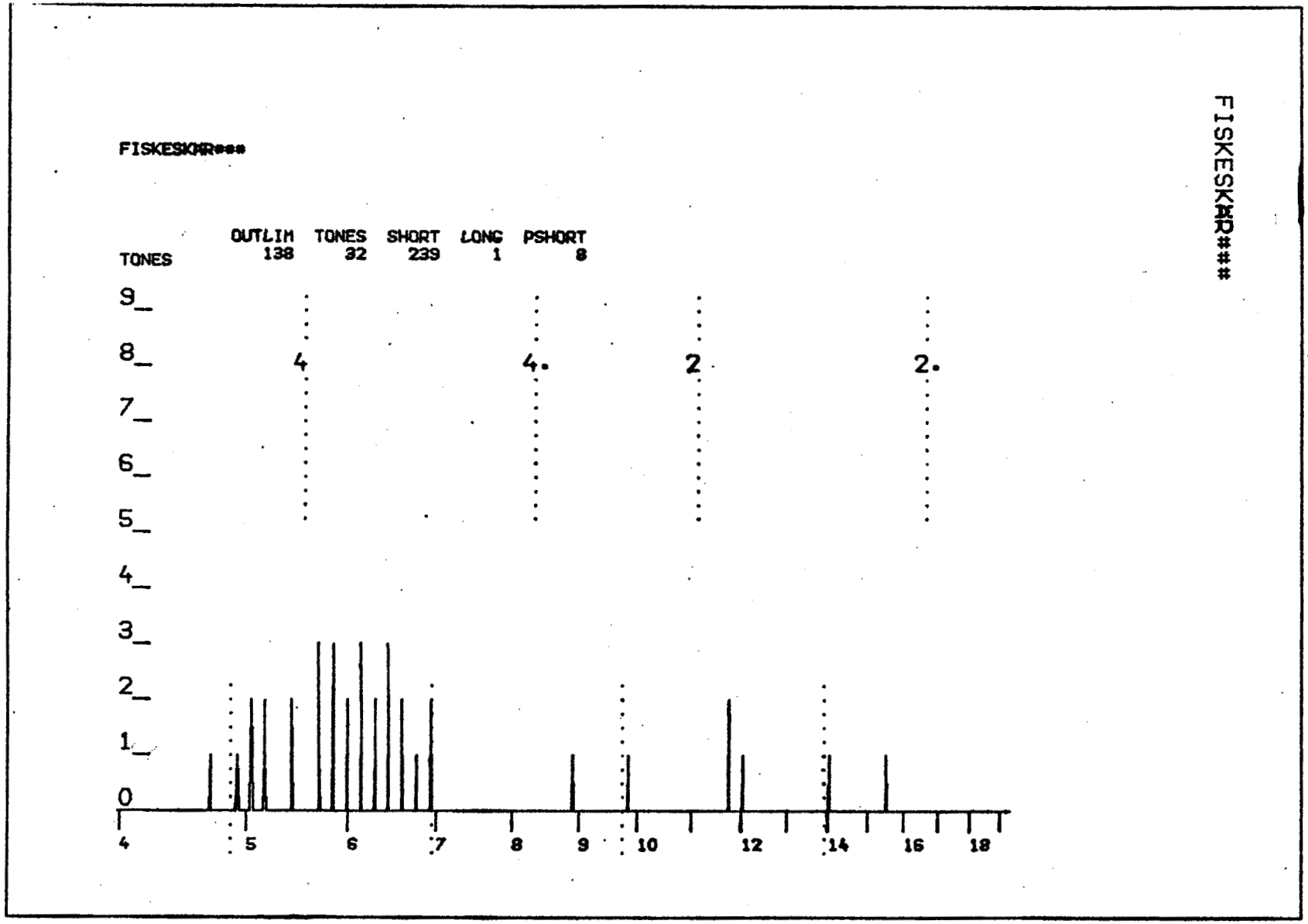


Fig. I-A-14. Processed DH of "Fiskeskärsvisan". Time scale in sec x 0.1.

to remember that there are other techniques possible for solving the problem with duration quantization, e. g. sequential analysis, some form of hypothesis testing, and adaptive variants of pattern recognition.

One aspect of the DH that must not be forgotten is that it shows the distribution of the durations of the musical events in a melody. This can be of interest in comparing different styles of playing and similar questions in ethnomusicology.

3.3 The notation

After the melody has been processed as described above, a string of pairs of quantified numbers is obtained. The transformation of that string into a printed notation is easily accomplished by the computer. The notation of the melody in the example is shown in Fig. I-A-15. The operator may decide any type of measure to be used in the notation. Presently, all melodies are notated in either G-major or G-minor. The discrepancies in the notation compared with the original notation in Fig. I-A-8 are few and not important to the identification of the melody. The discrepancies appear in bar 4 and 6 relative the original notation in terms of rests due to breathing and phrasing. The two final notes are prolonged because of a slight ritardando.

These discrepancies illustrate some of the difficulties associated with automatic computer notation of played music. The output reflects details in the actual performance, e. g. prolongings and shortenings of the tones, insertion of rests etc., which in conventional notation appear as diacritic signs or which are simply omitted and left to be realized by the performed musical understanding. Such discrepancies may reduce the legibility of an automatic computer notation.

The most serious shortcoming in the notation in Fig. I-A-15 is the absence of bar lines. Although this melody is of very simple structure, it is not quite trivial to read the notation without the support of such lines. Ideally it would be sufficient to inform the computer about the location of the first bar line as the remaining bar lines can be obtained simply by adding the note values according to the type of relevant measure. In practice, however, the final placing of the bar lines must be preceded by a routine which adjusts the actual note values of the melody tones to fit bars. A possible basis for such a routine would be a

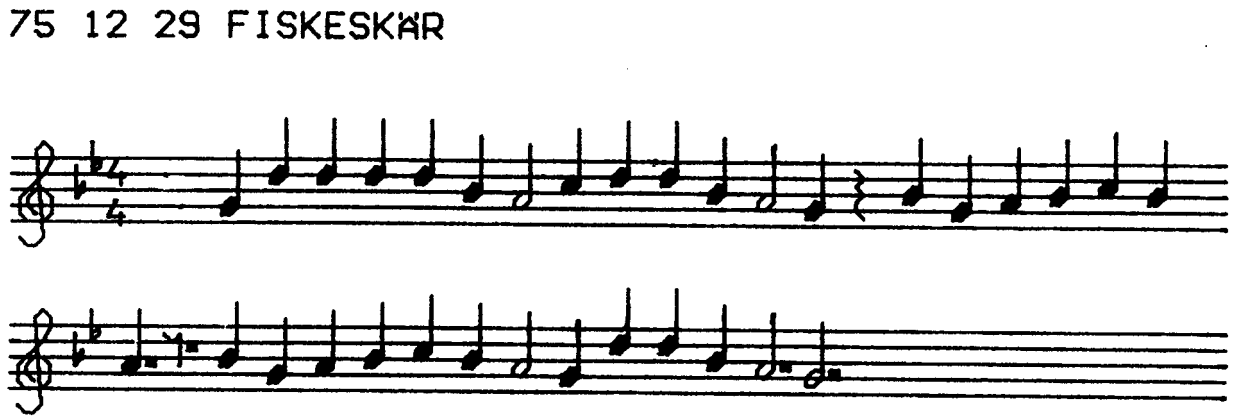


Fig. I-A-15. Automatically obtained notation of "Fiskeskärsvisan".

rule system defining legal and illegal rhythm sequences in the style of music concerned. It may be desirable to let this rule system be of a standard kind, but allowing the operator easy means of modifying the rules in accordance with the style of music. Remaining singularities in the notation must be corrected manually.

As has been pointed out several times above there are still many problems left to be solved before the program can be used without difficulties on a great variety of melodies. However, there are two different types of problems, which should not be confused. One type is due to the incapacity of the system to register the actual acoustic event and transform it into a formally correct transcription. Another type is due to the limitation inherent in the representation of the music in terms of series of discrete states. The answer of the underlying question, i.e. if a particular melody can at all be represented as a sequence of discrete states (a sequence of notes) cannot be obtained from the melody itself. That decision must be made a priori[4]. However, the statistical treatment of the melody may provide useful hints. For instance, if the fundamental frequency histogram (FFH) and the duration histogram (DH) of a melody show no discernable quantization states whatsoever, this would suggest that this melody is not well suited for conventional notation.

For cases where the DH lacks clear peaks another version of the program has been developed, which is not afflicated with the burden of associations of the conventional notation, c.f. Fig. I-A-16. The basis for this representation is only the FFH. Here the tones appear as heavy lines on an ordinary note staff, defined in accordance with the FFH. The duration of a tone is proportional to the length of the corresponding note line, i.e. the durations of the tones are not quantified. The curve below the note staff shows the momentary frequency deviation in cents from the center frequency of the current tone. Thus, the pitch lapse can also be studied in detail. The intensity level and a time scale are also presented. This representation of a melody contains essentially the same information as recordings obtained from a "melody writer" but it is much easier to read.

4. Conclusion and plans for the future

This project shows that it is possible to obtain an automatic notation of played melodies with simple methods. The transcription is in fair

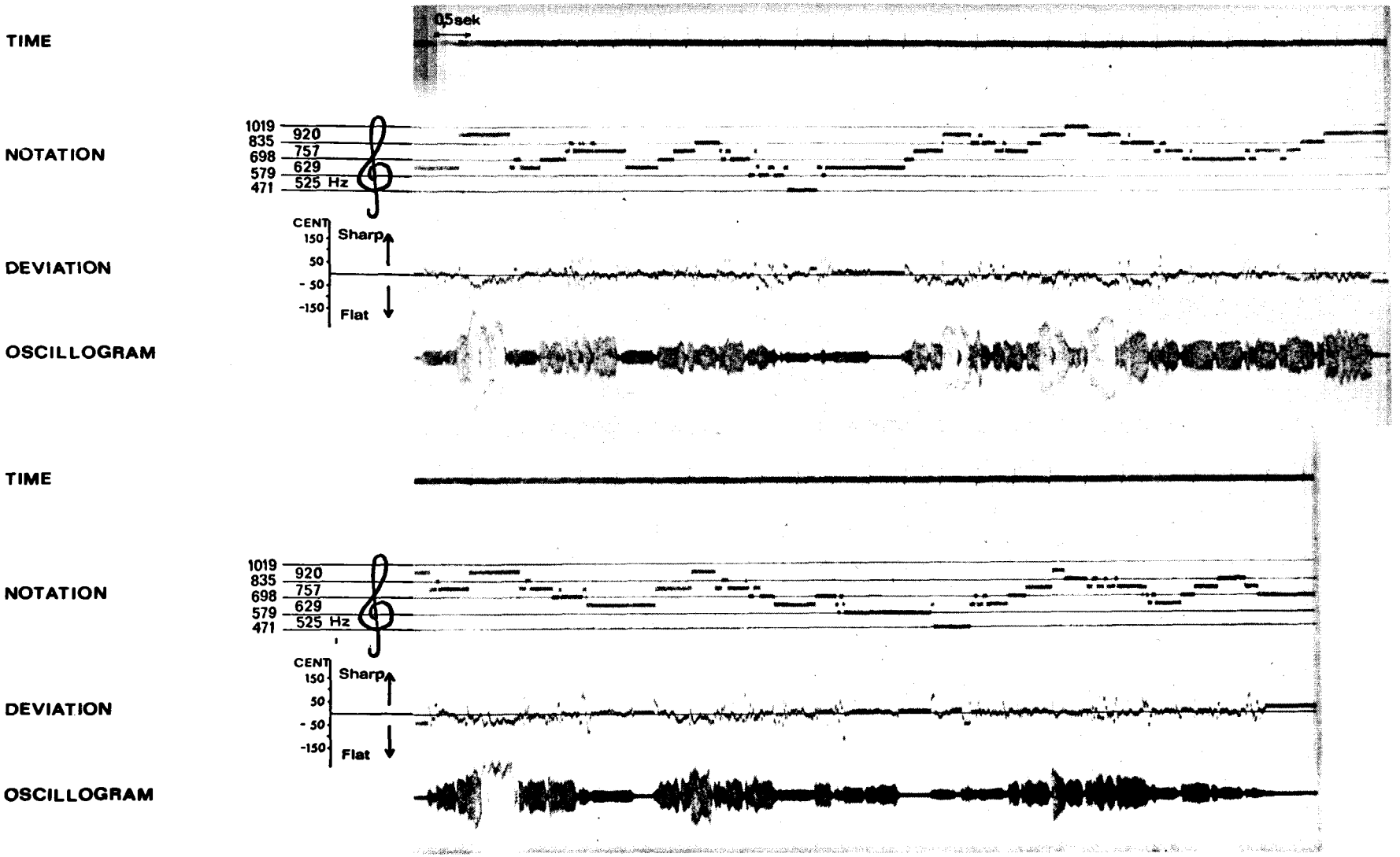


Fig. I-A-16. Notation without quantified durations, see text.

agreement with conventional notation. The melody used as an example in this paper is very simple and much refinement and enlargement of the computer program and hardware equipment is needed in order to facilitate successful processing of more complex melodies. Still the strategy used seems promising. A goal for the future is a computer system capable of processing large arrays of melodies in a uniform way.

The developmental work in the near future is planned as follows:

1) A closer evaluation of the performance of the pitch extractors under test.
2) Implementation of strategies for digital filtering of the recorded period time values.
3) Development of a routine for recording the intensity level values simultaneously with the period time values, in order to obtain a better recognition of pauses and tone repetitions.
4) Development of a transpose routine that allows the notation of a melody in optional key.
5) Development of the rule systems and manually routines for insertion of bar lines.
6) Connecting MUSSE (Music and Singing Synthesis Equipment) to the computer in order to make it possible to listen to the notated melody.

In the more distant future lies the processing of a number of melodies and the collecting of them in a melody library and also the development of search routines and associated programs.

Acknowledgments

This project is supported by the Swedish Humanistic Research Council, Contract No. 7914. The author is indebted to Johan Sundberg, Pekka Tjernlund, e specially as regards the idea of the TRIO, Kjell Elenius and Mats Blomberg for valuable discussions and hints.

References

(1) SUNDBERG, J. and TJERNLUND, P.: "A computer program for the notation of played music", STL-QPSR 2-3/1970, p. 46.

(2) SONDHI, M.M.: "New methods of pitch extraction", IEEE Trans. Audio Electroac., Vol. AU-16 (1968), pp. 262-269.

(3) HOLMSTRÖM, S.: "Grundtonsdetektering med spektrumutjämning", Thesis work, Dept. Elec. Eng., KTH; to be publ. June 1976.

(4) REMMEL, M., RUUTEL, I., SARV, J., and SULE, R.: "Automatic notation of one-voiced song", Acad. of Sciences of the Estonian SSR, Inst. of language and literature. Preprint KKT-4, Tallin 1975.