

12

**EUROPEAN PATENT APPLICATION**

21 Application number: 89120118.8

51 Int. Cl.<sup>5</sup> **G10H 1/00**

22 Date of filing: 30.10.89

30 Priority: 31.10.88 JP 275740/88

43 Date of publication of application:  
09.05.90 Bulletin 90/19

54 Designated Contracting States:  
**DE FR GB**

71 Applicant: **NEC HOME ELECTRONICS, LTD.**  
**5-24, Miyahara 3-chome Yodogawa-ku**  
**Osaka-Shi Osaka(JP)**

Applicant: **NEC CORPORATION**  
**33-1, Shiba 5-chome**  
**Minato-ku Tokyo(JP)**

72 Inventor: **Utsumi, Yoshinari**  
**c/o NEC Home Elec. Ltd. No. 5-24 Miyahara**  
**3-chome**  
**Yodogawa-ku Osaka-shi Osaka(JP)**  
Inventor: **Tsuruta, Shichiro**  
**c/o NEC Home Elec. Ltd. No. 5-24 Miyahara**

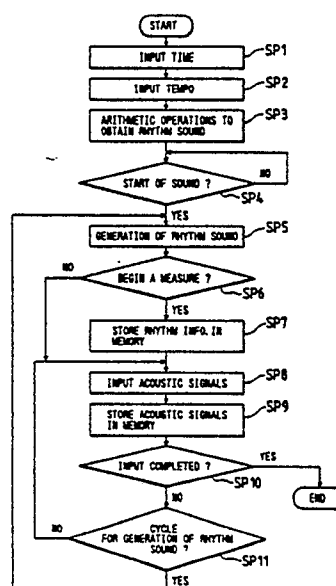
**3-chome**  
**Yodogawa-ku Osaka-shi Osaka(JP)**  
Inventor: **Fujii, Hiromi**  
**c/o NEC Corp. No. 33-1, Shiba 5-chome**  
**Minato-ku Tokyo(JP)**  
Inventor: **Fujimoto, Masaki c/o NEC Scientific**  
**Information**  
**System Develop., Ltd. No. 7-15, Shiba**  
**5-chome**  
**Minato-ku Tokyo(JP)**  
Inventor: **Mizuno, Masanori c/o NEC Scientific**  
**Information**  
**System Develop., Ltd. No. 7-15, Shiba**  
**5-chome**  
**Minato-ku Tokyo(JP)**

74 Representative: **DIEHL GLAESER HILTL &**  
**PARTNER**  
**Patentanwälte Flüggenstrasse 13**  
**D-8000 München 19(DE)**

54 **Automatic music transcription method and system.**

57 An arrangement (apparatus and method) for automatically transcribing music. Means are provided for capturing acoustic signals and storing them in a memory. Auxiliary rhythms are generated and displayed to a user in audio/video form. Once stored in memory along with the rhythms, pitch and power information are extracted from the acoustic signals. The acoustic signals are segmented based on the pitch and/or power information. This segmenting, in essence, divides the acoustic signals into sections each of which can be regarded to form a single level in musical interval. The segments are identified with a level on the axis of absolute musical interval on the basis of the pitch information. Then the results of the foregoing processing are displayed and/or printed.

**FIG. 1**



## AUTOMATIC MUSIC TRANSCRIPTION METHOD AND SYSTEM

This invention relates in general to an automatic music transcription method and system. The invention is in the field of automatic music transcription and refers to an arrangement (method and apparatus) for preparing musical score data from acoustic signals. These acoustic signals may include vocal sounds, humming voices, and musical instrument sounds.

An automatic music transcription system transforms such acoustic signals as those in vocals, hummed voices, and musical instrument sounds, into musical score data. It is necessary for such a system to be able to detect from the acoustic signals basic items of information, such as, for examples, sound lengths, musical intervals, keys, times, and tempos.

Acoustic signals comprise repetitions of fundamental waveforms in continuum. It is not possible to obtain directly from the acoustic signals the basic items of information needed to establish the musical score data.

According to a conventional method of automatic music transcription, the individual items of information are obtained by the following sequence of steps:

a) information is obtained regarding the repetitions of fundamental waveforms representing the levels of acoustic signals in pitch (which will hereafter be referred to as "the pitch information") and the power information of such signals for each of the analyzed circles,

b) then, the acoustic signals are divided into those sections (i.e. segments) which can be considered to form one and the same level in musical interval (this process being called "segmentation") on the basis of the pitch information and/or the power information so extracted, and

c) subsequently, the musical interval of each segment is determined with reference to the axis of absolute musical interval on the basis of the pitch information on the particular segment, and the key of the acoustic signal is determined on the basis of the information on the musical interval so determined, and

d) thereafter, the time and tempo of the acoustic signal are determined on the basis of the segment.

Now that also the time and the tempo are determined in the application of the existing automatic music transcription method, the user will eventually sing or play a desired song, while keeping time and tempo for himself. However, for a user who is not accustomed to performance or singing, such an act like this is difficult to do. Moreover, there are users who like to perform music or sing a

song, measuring the tempo with a metronome or the like.

Furthermore, the acoustic signals which are input from a music performance or a song by a user measuring the tempo or keeping the time for himself, above all those acoustic signals which occur in songs, contain fluctuations in power and pitch, and, because of this feature, it has been found difficult to perform segmentation even with the utilization of the power information and the pitch information. Segmentation is an element important for the compilation of musical score data, and a lower degree of accuracy in segmentation results in a considerably low degree of accuracy in the musical score data to be obtained ultimately. Therefore, it is to be desired that the accuracy of segmentation is improved. This object is solved by the method of independent claim 1 and the system of independent claim 7. Further advantageous features of the invention are evident from the dependent claims.

The present invention provides an automatic music transcription arrangement (apparatus and method) which is easier to use than known systems. Furthermore, the system according to the present invention provides more accurate segmentation than can be obtained from known systems.

According to a first aspect of the invention, there is provided an arrangement for capturing acoustic signals and storing them in the memory while reporting the information on the input auxiliary rhythms including at least information on tempo by an auditory sense process or a visual sense process, the system being incorporated in an automatic music transcription system which converts such acoustic signals into musical score data by a set of processes including at least the process for capturing such acoustic signals and storing them in the memory by means of an acoustic signal input means and thereafter extracting the pitch information, which represents the repetitive cycles of their waveforms and their sound pitch, and the power information of such acoustic signals out of the acoustic signals so stored in the memory, the process for segmentation, which consists in dividing the acoustic signals into sections each of which can be regarded to represent a single level in musical interval, by performing such segmentation on the basis of the pitch information and/or the power information, and the musical interval identifying process, which identifies each of the segments derived by such division with a level on the axis of absolute musical interval on the basis of the pitch information.

The system has been designed to give the

users the input auxiliary information by an acoustic sense process and/or a visual sense process, so that they may have ease and simplicity in generating the acoustic signals when they capture acoustic signals and take such signals into the system for storing them in the memory for the purpose of performing the process for music transcription.

According to a second aspect of the present invention, there is provided an automatic music transcription system that is somewhat similar to the first aspect described above, but wherein the system stores input auxiliary rhythm information as well in the memory on the same time axis at the time when it performs the capturing and storing in memory of the acoustic signals and wherein the segmentation process is divided among the first process for dividing the acoustic signals into those sections which can be regarded as representing the same level of musical interval as determined on the basis of the input auxiliary rhythm information stored in the memory, the second process for dividing the acoustic signals into those segments which can be regarded as representing the same level in musical interval as determined on the basis of the pitch information and/or the power information, and the third process for making adjustments to the sections as divided by the first process and the second process.

The system is arranged so as to utilize the input auxiliary rhythm information, so that the accuracy of the segmentation process may be improved. In other words, the system stores in its memory also the input auxiliary rhythm information at the same time as the acoustic signals are captured and stored in the memory. Then, the system performs its segmentation process on the basis of this input auxiliary rhythm information, performs its segmentation process also on the basis of the pitch information and the power information, and then makes adjustments to the results of such segmentation processes.

According to a third aspect of the present invention, there is provided a system including an input auxiliary rhythm reporting means whereby the input auxiliary rhythm information including at least the tempo information is reported by an auditory sense process and/or a visual sense process at the time when the acoustic signals are captured and stored in the memory, the system being incorporated in an automatic music transcription system for converting the acoustic signals into musical score data, the system being provided at least in some with the means of capturing and taking the acoustic signals into the system, the means of storing in the memory the acoustic signals so taken into the system, the pitch and power extracting means, which extracts the pitch information representing the repetitive cycle of the waveforms in the

acoustic signals stored in the memory and representing the level in pitch, and the power information from the acoustic signals, the segmentation means for dividing the acoustic signals into those sections which can be regarded as representing the same level in musical interval as determined on the basis of the pitch information and the power information, and the musical interval identifying means, which determines the musical interval of the acoustic signals, with respect to the sections so divided, with reference to the axis of absolute musical interval.

The system is so designed that its input auxiliary rhythm reporting means reports the input auxiliary rhythm information by an auditory sense process and/or a visual sense process at the time when the acoustic signals are captured and stored in the memory. As the result of this feature, it has been made possible for the user to perform the input operations on the basis of the input auxiliary rhythm information and consequently to enjoy greater ease in the input of signals.

According to a fourth aspect of the present invention, there is provided a system having a memory means designed to store also the input auxiliary information on the input auxiliary rhythm information in memory on the same time axis at the time when acoustic signal is captured and processed for storage in the memory and provided also with a segmenting means including a first segmenting section for segmenting the acoustic signals into those sections each of which can be regarded as forming one and the same level of musical interval, as determined on the basis of the input auxiliary rhythm information stored in the memory, a second segmenting section for segmenting the acoustic signals into those sections each of which can be regarded as forming one and the same level of musical interval, as determined on the basis of the pitch information and the power information, and a third segmenting section for making adjustments to those sections as divided into segments by the first segmenting section and the second segmenting section.

The memory means, which stores the acoustic signals in its memory, keeps also the input auxiliary rhythm information in memory on the same time axis as reported from the input auxiliary rhythm information when the captured acoustic signals are stored in the memory, and the system is so designed that the first segmenting section performs its segmentation process on the basis of this input reporting rhythm information and the third segmenting section makes adjustments to the results of this segmentation process and the results of the segmentation performed by the second segmenting section on the basis of the pitch information and the power information. As the result of this

feature, the accuracy of segmentation can be improved.

Preferred embodiment(s) of the invention will be described in detail with reference to the drawings wherein like reference numerals denote like or corresponding parts throughout.

FIGURE 1 is a flow chart of the input process for the acoustic signals to be processed by one of the embodiments of the present invention;

FIGURE 2 is a block diagram of an automatic music transcription system incorporating the present invention;

FIGURE 3 is a flow chart of the automatic music transcription process;

FIGURE 4 is a flow chart of the segmentation process based on the measure information and the power information generated by the system;

FIGURE 5 is a flow chart showing greater details of the segmentation process based on the measure information and the power information;

FIGURE 6 is a characteristic curve chart representing one example of the segmentation; and

FIGURE 7 is a block diagram of other embodiments of the automatic music transcription system.

Preferred embodiment(s) of the invention will be described in detail with reference to the drawings wherein like reference numerals denote like or corresponding parts throughout.

#### Automatic Music Transcription System

FIGURE 2 is a block diagram of an automatic music transcription system incorporating the present invention. A Central Processing Unit (CPU) 1 performs overall control for the entire system. CPU 1 executes an acoustic signal input program shown in the FIGURE 1 flow chart and a music transcription processing program shown in the FIGURE 3 flow chart. The acoustic signal input and music transcription processing programs are stored in a main storage device 3 connected to CPU 1 via a bus 2. Also coupled to bus 2 are a keyboard 4, which serves as an input device, a display unit 5, which serves as an output device, an auxiliary memory device 6 for use as working memory, and an analog/digital (A/D) converter 7. An acoustic signal input device 8, which may comprise a microphone, etc. provides input to A/D converter 7. Acoustic signal input device 8 captures the acoustic signals in vocal songs or humming voices or like sound signals generated by musical instruments and then transforms the signals into electrical signals, thereafter outputting the electrical signals to the A/D converter 7.

Also connected to bus 2 is a speaker driving section 9 for driving a speaker 10. Speaker 10

generates, when necessary, scattered input auxiliary rhythm sounds representing the predetermined time and tempo under control of CPU 1.

CPU 1 operates in accordance with the acoustic signal input program flow charted in FIGURE 1 to input acoustic signals into the system. These signals are stored in main storage device 3. When there has been received a command to input the acoustic signals, together with a command to operate with the specified time and tempo, as entered on the keyboard 4, the input acoustic signals are stored in an orderly sequence into the auxiliary storage device 6. The system also temporarily stores input auxiliary rhythm information in auxiliary memory device 6.

Upon completion of the input of acoustic signals into the system, CPU 1 executes the music transcription processing program (flow charted in FIGURE 3) stored in the main storage device 3 thereby converting the input acoustic signals into musical score data and outputting such data to display unit 5 as required.

#### Input of Acoustic Signals

FIGURE 1 is a flow chart of the process for inputting acoustic signals. When the CPU 1 receives a command by way of keyboard 4 to operate in its input mode, the CPU 1 starts executing the program flow charted in FIGURE 1. It first displays on the display unit 5 a prompt for the use to input timing information. It then receives timing information from the user in response to the prompt via keyboard 4. Display unit 5 then displays a prompt to the user to input tempo information. The tempo information is received from the user in response to that prompt (Steps SP 1 and SP 2). Thereafter, the CPU 1 carries out arithmetic operations to determine the cycle and intensity of the input auxiliary rhythm information on the basis of the timing information and the tempo information. CPU 1 then stands by for the input of an input start command from keyboard 4 (Steps SP 3 and SP 4).

When an input start command is given by the user, the CPU 1 causes an input auxiliary rhythm sound to be generated from the speaker 10. It thereafter determines whether or not the input auxiliary rhythm sound so generated indicates the beginning of any measure.

If the sound indicates the beginning of a measure, then the CPU 1 stores the sound in the auxiliary storage device 6 and thereafter receives into the system the acoustic signals composed of digital data as processed through the acoustic signal input device 8 and the A/D converter 7. However, if the sound does not indicate any beginning of a measure, then the CPU 1 immediately inputs

of the acoustic signals (Steps SP 5 through SP 8). Thereafter, the CPU 1 stores the acoustic signals so input into the system in the auxiliary storage device 6 (Step SP 9).

When the one set of data on the acoustic signal are thus stored in the auxiliary storage device 6, the CPU 1 determines whether or not any command to finish of the input operation has been given by way of the keyboard 4. When a finish command has been given, the CPU 1 stops its series of operations. However, if there has not been any finish combined given, the CPU 1 further determines whether or not the system is in a timing for the generation of any input auxiliary rhythm sound (Steps SP 10 and SP 11). If it is not in any timing for the generation of such a sound, the CPU 1 returns to step SP 8 and proceeds to the step at which it takes acoustic signals into the system. If it is found that the operation of the system is in the timing for generating the input auxiliary rhythm sound, the CPU 1 returns to step SP 5, and moves on to the step for the generation of the next input auxiliary rhythm sound.

Thus, the system takes in acoustic signals generated by a user while generating the input auxiliary rhythm sound, and stores the signals one after another in orderly sequence, together with marks indicating the beginning of a measure, in the auxiliary storage device 6.

The feature of the system related to generating the input auxiliary rhythms sound makes it easy for a user to input of the acoustic signals.

### Music Score Transcription Process

FIGURE 3 is a flow chart of the automatic music transcription process. This process does not occur until after the input of acoustic signals.

First, the CPU 1 extracts the pitch information for the acoustic signals for each analytical circle using autocorrelation analysis of the acoustic signals. It also extracts power information for each analytical cycle by processing the acoustic signals to find the square sum. Then, CPU 1 performs various pre-treatment processes, such as, for examples pre-treatments for noise elimination and smoothing (Steps SP 21 and SP 22).

Thereafter, the CPU 1 segments the input acoustic signals into predetermined sections on the basis of the marks placed at the beginning of each measure as stored in the auxiliary storage device 6. It then reviews such sections on the basis of the changes in power, thereby separating such sections to establish the segments which can be regarded as representing the same sound (Steps SP 23 and SP 24).

Next, the CPU 1 performs a tuning process

(Step SP 25). CPU 1 calculates the amount of deviation of the musical interval axis that the acoustic signal has in relation to the axis of the absolute musical interval on the basis of the state of distribution of the pitch information and effecting a shift of the obtained pitch information in accordance with the amount of deviation thereof. In other words, the CPU 1 modifies the pitch information in such a way that there will remain a smaller difference between the axis of musical interval for the singer or the musical instrument that has generated the acoustic signal and the axis of the absolute musical axis.

The CPU 1 thus identifies the musical interval of the particular segment with that on the axis of the absolute musical interval to which the relevant pitch information is considered to be closest as seen on the basis of the pitch information of the segment obtained by the above-mentioned segmentation process and further executes the segmentation process again on the basis of whether or not the musical interval of the identified segments in continuum are identical (Steps SP 26 and SP 27).

After that, the CPU 1 finds the product sum of the frequency of occurrence of the musical interval as obtained by working out the classified total of the pitch information after the tuning thereof and the certain prescribed weighing coefficient as determined in correspondence to the key, and, on the basis of the maximum value information of this product sum, the CPU 1 determines the key, for example, the C-major key or the A-minor key, for the piece of music in the input acoustic signals, thereafter ascertaining and correcting the musical interval by reviewing the same musical interval in greater detail with respect to the pitch information regarding the prescribed musical interval on the musical scale for the determined key (Steps SP 28 and SP 29).

Subsequently, the CPU 1 carries out a final segmentation by reviewing the segmentation results on the basis of whether or not the finally determined musical interval contains identical segments in continuum or whether or not there is any change in power among the segments in continuum (Step SP 30).

After the musical interval and the segments (i.e. the sound length) have been determined in this manner, the CPU 1 produces the finalized musical score data through adjustment of the information including the timing information and the tempo information which were input at the time when the input of the acoustic signals was started (Step SP 31).

### Segmentation Based on Measure and Power

FIGURE 4 is a flow chart of the segmentation process based on the measure information and the power information generated by the system and FIGURE 5 is a flow chart showing greater details of the segmentation process based on the measure information and the power information. The following is a detailed explanation of the flow charts in FIGURE 4 and FIGURE 5 related to the segmentation process (Steps SP 23 and SP 24 in FIGURE 3) based on the measure information and the power information on the acoustic signals. In this regard, FIGURE 4 is a flow chart illustrating such a process at the functional level while FIGURE 5 is a flow chart illustrating the greater details of what is shown in FIGURE 4.

The acoustic signals are brought to their squares with respect to the individual sampling points within the analytical cycle, and the sum total of those square values is used for the power information on the acoustic signals in the analytical cycle.

First, an outline of such a segmentation process is described with reference to FIGURE 4. For the purposes of illustration only, we now assume that the quadruple time has been selected for the measure of the signals. Of course, the invention is not limited to this assumption. The assumption is made only to facilitate explanation of the invention.

The CPU 1 takes out the mark for the beginning of a measure as stored in the auxiliary storage device 6, divides each measure into four equal portions, and puts a mark indicating the beginning of a beat at the initial part of each of the equally divided portions (Step SP 40). In the case of quadruple measure not being selected, but rather triple measure having been selected, the measure is to be divided into three equal portions. Next, the CPU 1 makes a further division of each of the obtained beats into four equal portions, and puts a mark for the beginning of a semiquarter note at the initial part of each of the equally divided portions (Step SP 41). In this manner, the acoustic signals are divided into 16 portions of each measure on the basis of the measure information. In those cases where not the quadruple measure but the triple measure has been selected, one measure will be divided into twelve equal portions. Thereafter, the CPU 1 reviews these divided portions on the basis of the power information.

The system has been so arranged as to reflect the power information on the segmentation process because users may produce changes accompanying some intensification of power when they change the pitch in the sounds, i.e. when they make a transition to the next sound.

CPU 1 then extracts the point of a rise in the power information, putting a mark indicating a rising point at the appropriate place and thereafter

taking the mark which indicates the beginning of a semiquarter note and is located at a point closest to each of the rising point and putting a mark indicating the beginning of a semiquarter note at the rising point (Steps SP 42 and SP 43). The reason why these steps are carried out is that it is practically difficult for the user to make a change in sound in complete agreement with the timing of the input of acoustic signals even if such signals are input with the generation of the input auxiliary rhythm sound, so that it is provided to ensure that it can be judged with certainty whether or not the following section is a rest section with the help of the division of a sound set at a point of change in the acoustic signal.

Subsequently, the CPU 1 counts the number of pieces of the pitch information in each semiquarter note section and puts a mark indicating the beginning of a rest at the initial point of each section where the number of pieces of such information is smaller than the threshold value (Step SP 44). Finally, the CPU 1 places a mark indicating the beginning of a segment at those points bearing a mark for the beginning of a measure, a rising point, or the beginning of a rest (Step SP 45). A mark is made indicating the beginning of a segment also at the point where a measure begins because one sound may extend over two measures, in which case it is the practice to show musical notes in the respective measures indicated on the score.

In this manner, the system obtains a plural number of segments obtained by the division based on the measure information and the power information. Even if some of the segments obtained by this segmentation process should turn out to be inadequate ones, such segments will be rectified to be proper segments by the effect of the segmentation to be executed at subsequent steps (Steps SP 27 and SP 30 given in FIGURE 3) as mentioned above.

Next, this process is explained in greater detail with reference to the flow chart in FIGURE 5. The CPU 1 first clears to zero the parameter *i* indicating each analytical circle (such an analytical circle like this is hereafter called an analytical point in view of the fact that it has a very short analytical cycle), and then, ascertaining that the analytical point data (which include pitch information and power information) to be processed has not yet been completed, the CPU judges whether or not any mark indicating the beginning of a measure is placed on that analytical point (Steps SP 50 through SP 52). In case no such mark is placed, the CPU increment the parameter *i* for the analytical point and returns to the above-mentioned Step SP 51, but, in case where such a mark is placed, the CPU 1 proceeds to perform the processes at the Step SP 54 and the subsequent steps (Step SP 54). In this manner,

the CPU 1 finds the mark indicating the beginning of the first measure.

Having detected a mark indicating the beginning of a measure, the CPU 1 sets  $i + 1$  in the parameter  $j$ , and, ascertaining that the analytical point data to be processed have not been completed, the CPU 1 judges whether any mark indicating the beginning of a measure is placed on the particular analytical point (Steps SP 54 through SP 56). In case no such mark is placed, the CPU 1 increments the parameter  $j$  and returns to the Step SP 55 mentioned above, but, in case such a mark is placed, the CPU 1 proceeds to the processing of the Step SP 58 and the subsequent steps (Step SP 57).

Here, at the timing which has generated an affirmative result at Step SP 56, the parameter  $i$  indicates the analytical point positioned at the former mark out of the two consecutive marks which indicate the beginning of a measure while the parameter  $j$  indicates the analytical point positioned at the latter of the two consecutive marks which indicate the beginning of a measure. Thus, the CPU 1 divides the section from the analytical point  $i$  to the analytical point  $j-1$  into four equal portions (or into three equal portions in the case of such a section with the triple beat) and puts a mark for the beginning of a beat on each of those portions, thereafter setting  $j$  in the parameter  $i$ , which indicates the analytical point positioned in the former of the marks indicating the beginning of a measure, and then returning to the above-mentioned Step SP 54 to proceed to the searching of the analytical point bearing the mark indicating the beginning of a measure and positioned in the latter of the analytical points (Steps SP 58 and SP 59).

By the repeated execution of this loop operation process including Steps SP 54 through SP 59, the marks indicating the beginning of each beat are placed one by one in orderly sequence in the individual measure sections until the data on the final analytical point are taken out to produce an affirmative result at the Step SP 55. At such a time, the CPU 1 places a mark indicating the beginning of a beat at the analytical point for the parameter  $i$  at the particular point in time, therewith completing a series of processes for putting the mark indicating the beginning of a beat and thereafter proceeds to Step 61 and the subsequent steps for putting the mark indicating the beginning of each semiquarter note (Step SP 60).

If CPU 1 obtains an affirmative result at the Step SP 51 as it comes to the final data without finding any mark indicating the beginning of the initial measure, the CPU proceeds, without placing any mark on such sections, to the processes for putting the marks indicating the beginning of the semiquarter notes. The portion of the process in-

cluding Steps SP 50 through SP 60 correspond to Step 40 in FIGURE 4.

The details of the processes corresponding to the Step 41 in FIGURE 4, which are to be performed for putting the marks indicating the beginning of the semiquarter notes by finding the two marks one preceding the other and indicating the beginning of the beat and dividing the sections with such marks into four equal portions, are almost identical to the processes of Steps SP 50 through SP 60. Marks are placed indicating the beginning of the beats by finding the marks indicating the beginning of the respectively preceding and following sections and dividing those sections into four equal portions. Therefore, a detailed discussion of that process is omitted (Steps SP 61 through SP 71).

Upon completion of the processes of placing marks indicating the beginning of the semiquarter notes, the CPU 1 clears to zero the parameter  $i$  for the analytical point and thereafter performs arithmetic operations to determine the function  $d(i)$  for extracting the rise in the power information with respect to that analytical point, ascertaining the point that the analytical point data to be processed have not yet been brought to a finish (Steps SP 72 through SP 74).

The rise extraction function  $d(i)$  for the power information, power  $(i)$ , with respect to the analytical point  $i$  is determined by arithmetic operations by the following equation and applied as such:

$$d(t) = \frac{\{\text{power}(i+t) - \text{power}(i)\}}{\{\text{power}(i+t) + \text{power}(i)\}} \quad (1)$$

Where  $t$  represents a natural number indicating an amount of time adequate for capturing the fluctuations in the rise of the power information.

Thereafter, the CPU 1 judges whether or not the value of the rise extraction function  $d(i)$  so obtained is any smaller than the threshold value  $\theta d$ , and, if it is smaller, the CPU 1 increments the parameter  $i$  for the analytical point and returns to the Step SP 73 (Steps SP 75 and SP 76). On the other hand, in case the rise extraction function  $d(t)$  is found to be in excess of the threshold value  $\theta d$ , the CPU 1 places the mark indicating the beginning of a rise point to that analytical point (Step SP 77). Thereafter, the CPU 1 ascertains that the processing has not yet been completed on the data with respect to all the analytical points and then, performing arithmetic operations to determine the rise extraction function  $d(i)$ , judges whether or not the rise extraction function  $d(i)$  smaller than the threshold value  $\theta d$  (Steps SP 78 through SP 80). In case the rise extraction function  $d(i)$  is smaller than the threshold value, the CPU 1 increments the parameter  $i$  and returns to the above-mentioned Step SP 78 (Step SP 81).

The process of Steps from SP 78 through SP

81 is a process for finding the analytical point at which the rise extraction function  $d(i)$  becomes smaller than the threshold value  $\theta d$  after the rise extraction function once grows larger than the threshold value. Now that there is an analytical point where the rise extraction function rises again after the analytical point thus obtained, the CPU 1 returns to the above-mentioned step SP 73 and resumes the process for extracting the rise point if it has found an analytical point where the rise extraction function becomes smaller than the threshold value, i.e. if it obtains an affirmative result at the above-mentioned step SP 80.

By repeating the processing procedure mentioned above, the CPU 1 soon detects that the processing has been completed of all the analytical points at the Steps SP 73 or SP 78, and the CPU 1 proceeds to a review of the rise points on the basis of the length between the adjacent rise points at the Step SP 82 and the subsequent steps.

In such a process, the CPU 1 clears to zero the parameter  $i$  for the analytical point, and then, ascertaining that the data on the analytical point have not yet been brought to a finish, the CPU 1 judges whether or not a mark indicating a rise point is placed on the analytical point (Steps SP 82 through SP 84). When the point is not a rise point, the CPU 1 increments the parameter  $i$  for the analytical point and then returns to the Step SP 83 (Step SP 85). Upon the detection of a rise point through the repeated performance of this process, the CPU 1 sets the length parameter  $L$  at the initial value "1" in order to measure the length from the rise point to the next rise point (Step SP 86).

Thereafter, the CPU 1 increments the analytical point parameter  $i$ , and then, ascertaining that the analytical point data has not yet been completed, further judges whether or not any mark indicating the beginning of a rise point is placed on the particular analytical point (Steps SP 87 through SP 89). If the CPU 1 finds as the result that the analytical point is not any rise point, the CPU 1 increments the length parameter  $L$  and also increments the analytical point parameter  $i$ , thereafter returning to the above-mentioned step, SP 88 (Steps SP 90 and SP 91).

By repeating the process of the steps SP 88 through SP 91, the CPU 1 will soon come to an analytical point where the next mark indicating the beginning of a segment is placed, obtaining an affirmative result at the step SP 89. The length parameter  $L$  found at this time corresponds to the distance between the marked analytical point being taken up for processing and the immediately preceding marked analytical point, i.e. to the length between the respectively preceding and following rise points. If an affirmative result is obtained at the step SP 89, the CPU 1 judges whether or not this

parameter  $L$  is shorter than the threshold value  $\theta L$ , and, when it is found to be above the threshold value  $\theta L$ , the CPU 1 returns to the above-mentioned Step, SP 83, without eliminating the mark indicating a rise point, but, when it is smaller than the threshold value  $\theta L$ , the CPU 1 removes the former mark indicating the rise point, and then returns to the above-mentioned step SP 83 (Steps SP 92 and SP 93).

Moreover, in case the CPU 1 has returned to the step SP 83 from the step SP 92 or SP 93, the CPU 1 will immediately obtain an affirmative result at the step SP 84, unless the analytical point data has been completed, and the CPU 1 will proceed to the processing at the subsequent steps beginning with the step SP 86 and will move on to the operation for searching for another mark next to the mark just found.

By repeating this sequence of steps, the CPU 1 will complete the review of the lengths between the rise points with respect to all the rise points, and when it soon obtains an affirmative result at the Step SP 83 or the Step SP 88, the CPU 1 will complete the series of processes for the extraction of the rise points in the power information. The process of Steps, SP 72 through SP 93 corresponds to the process of Step SP 42 shown in FIGURE 4.

The reason why this system has been arranged to review the rise points with reference to the distance between the respectively preceding and following rise points after the rise points have been extracted with the rise extraction function  $d(i)$  is the necessity of preventing the occurrence of a plural number of rise points in a section shorter than the length of a single sound in consequence of the fact that the power in acoustic signals may yet undergo fluctuations even though there are intended to be a single sound and the fact that acoustic signals may contain an intrusive outside noises.

When the CPU 1 completes the process for thus extracting the rise points in the power information by repeating this processing procedure, the CPU 1 first clears to zero the parameter  $i$  for the analytical point and then, ascertaining that the data to be processed are not yet finished, the CPU 1 judges whether or not any mark indicating a rise point in the power information is placed with respect to that analytical point (the Steps, SP 94 through SP 96). In case no such mark is placed, the CPU 1 increments the parameter  $i$  and then returns to the step, SP 95, mentioned above (Step SP 97). When the CPU 1 finds one rise point in this manner, the CPU judges whether or not any mark indicating a semiquarter note is placed on the analytical point  $i$  thereof (Step SP 98).

In case it is found that a mark indicating a



semiquarter note is placed on the point, the CPU 1 increments the parameter  $i$  and then returns to the Step SP 95 mentioned above, thereupon proceeding to the process for searching the next rise point because it is not necessary to perform any processing for the matching of that rise point and the beginning point of the semiquarter note (Step SP 99).

On the other hand, in case the rise point so found does not have any mark indicating the beginning of a semiquarter note, then the CPU 1 returns to the above-mentioned step SP 95 and proceeds therefrom to the process for searching the rise point positioned closest to this rise point.

First, the CPU 1 puts a mark indicating the beginning of a semiquarter note at the rise point, and then it sets the parameter  $j$  at its initial value "1" for finding the analytical point preceding the rise point and bearing a mark indicating the beginning of a semiquarter note (Steps SP 100 and SP 101).

Thereafter, ascertaining that the section  $i-j$  is not any less than 0 (which means that the analytical point  $i-j$  is an analytical point loaded with data), the CPU 1 judges whether or not any mark indicating the beginning of a semiquarter note is placed on the analytical point  $i-j$ . In case no such mark is placed there, the CPU 1 increments the parameter  $j$ , thereafter returning to the Step 102 (Steps SP 102 through SP 104). By repeating the process of the steps, SP 102 through SP 104, the CPU 1 finds the analytical point  $i-j$ , which is located in a position closest on the side preceding the rise point where a mark indicating a semiquarter note is placed, then obtaining an affirmative result at the Step SP 103.

In such a case, the CPU 1 sets the parameter  $k$ , which is a parameter for finding the analytical point bearing a mark indicating a semiquarter note at the side following the rise point, at the initial value "1" (Step SP 105). Thereafter, the CPU 1 ascertains that the analytical point  $i+k$  does not have any value larger than that of the final analytical point, which amounts to saying that the analytical point  $i+k$  is one where data are present, and then judges whether or not any mark indicating the beginning of a semiquarter note is placed on the analytical point  $i+k$ . If no such mark is placed there, the CPU 1 increments the parameter, then returning to the step SP 106 (Steps SP 106 through SP 108). By repeating the process of steps, SP 106 through SP 108, the CPU 1 finds the analytical point  $i+k$ , which is positioned closest to and following the rise point which bears the mark indicating the beginning of a semiquarter note, then obtaining an affirmative result at the step SP 107.

Thus finding the analytical points positioned closest to the rise point and respectively preceding

and following the point where a mark indicating the beginning of a semiquarter note is placed, the CPU 1 compares the two parameters  $j$  and  $k$  in terms of size and judges which of the two analytical points are closer to the rise point, and, in case the analytical point  $i-j$  positioned on the preceding side is closer to the rise point (including those cases where the analytical point is equally close to the rise point), the CPU 1 removes the mark indicating the beginning of a semiquarter note from the analytical point  $i-j$ , where it has been placed, and thereafter the CPU 1 increments the parameter  $i$  and proceeds to the process of searching the next rise point. On the other hand, if the analytical point  $i+k$  positioned on the following side is closer to the rise point, the CPU 1 removes the mark indicating the beginning of a semiquarter note from that analytical point  $i+k$  where it has been attached, and thereafter increments the parameter  $i$  and proceeds to the process of searching the next rise point (Steps SP 109 through SP 113).

By repeating this process, the CPU 1 places a mark indicating the beginning of a semiquarter note on every rise point while it removes the mark indicating the beginning of a semiquarter note from the point closest to that rise point. And, when this process is completed with respect to all the analytical points, the CPU 1 finishes the process for matching the series of rise points and the points marking the beginning of the semiquarter points by the step, SP 95. Moreover, the process of steps SP 94 through SP 113 corresponds to the step SP 43 of FIGURE 4.

Having thus completed the process of changing the rise points in the power information, the CPU 1 clears to zero the parameter  $i$  for the analytical point and then, ascertaining that the data to be processed with respect to the analytical point are not yet finished, the CPU 1 judges whether or not a mark indicating the beginning of a semiquarter note is placed on that analytical point (steps, SP 114 through SP 116). In case no such mark is placed, the CPU 1 increments the parameter  $i$  and returns to the above-mentioned step SP 115 (Step SP 117). When the first mark indicating the beginning of a semiquarter note has thus been located, the CPU 1 sets at  $i-1$  the parameter  $j$  applied to the next mark indicating the beginning of a semiquarter note, and then, ascertains that the data to be processed of the analytical data have not been finished yet, the CPU 1 judges whether or not a mark indicating the beginning of a semiquarter note is placed on that analytical point  $j$  (Steps SP 118 through SP 120). In case no such mark is placed, the CPU 1 then increments the parameter  $j$  and returns to the step SP 119 mentioned above (Step SP 121).

When the next mark indicating the beginning of

a semiquarter note is found, the CPU 1 clears to zero the number-of-pieces parameter  $n$  for a segment with a pitch and thereafter sets at 1 the finish parameter  $k$  for the processing of a segment with the presence of pitch (Steps SP 122 and SP 123). Next, after ascertaining that the parameter  $k$  is smaller in value than the parameter  $j$ , the CPU 1 judges whether or not there is any pitch information present at the analytical point  $k$ , i.e. whether or not the analytical point  $k$  contains a voiced sound (Steps SP 124 and SP 125).

If an affirmative result has been obtained from this process, the CPU 1 then increments the number-of-pieces parameter  $n$  and thereafter also increments parameter  $k$ , then returning to the step SP 124 mentioned above. On the other hand, when a negative result has been obtained, the CPU 1 immediately increments the parameter  $k$ , thereafter returning to the above-mentioned step SP 124 (Steps SP 125 and SP 126). The repetition of this process will soon results in obtaining an affirmative answer at the step SP 124. Here, the parameter  $k$  changes within the range from  $i$  to  $j-1$ , and, when an affirmative result is obtained at the step SP 124, the number-of-pieces parameter  $n$  indicates the number of pieces of the analytical points with the presence of the pitch information between the analytical point  $i$  and the analytical point  $j-1$ , i.e. the number of pieces of the analytical points where there is some pitch information between the preceding and the following marks each indicating the beginning of a semiquarter note.

The CPU 1 judges whether or not the value of the number-of-pieces parameter  $n$  is larger than the prescribed threshold value  $\theta n$ . If the value of the parameter is smaller than the threshold value  $\theta n$ , the CPU 1 puts a mark for the beginning of a rest at the analytical point  $i$ , which is the first analytical point in the count of the number of pieces of the analytical points, where a mark indicating the beginning of a semiquarter note is placed, and thereafter the CPU 1 sets the parameter  $i$  at  $j$  and returns to the step SP 118 mentioned above. On the other hand, if the value of the parameter is more than the threshold value  $\theta n$ , the CPU 1 immediately sets the parameter  $i$  at  $j$ , thereafter returning to the above-mentioned step SP 118 and proceeding to the process of searching the next analytical point where a mark indicating the beginning of a semiquarter note is placed (Steps SP 128 through SP 130). By repeating this process, a mark indicating the beginning of a rest is placed one by one in orderly sequence at the first analytical point that is positioned between the respectively preceding and following marks each indicating the beginning of a semiquarter note and having a fewer number of pieces of analytical points with the presence of the pitch information,

and soon an affirmative result is obtained at the steps SP 115 or SP 119, and the series of processes for placing a mark indicating the beginning of a rest will be brought to a finish. In this regard, the process of steps, SP 114 through SP 130, corresponds to the process at the step SP 44 of FIGURE 4.

Upon completion of the process of placing a mark indicating the beginning of a rest, the CPU 1 clears to zero the analytical point parameter  $i$ , and, ascertaining that the analytical point data to be processed have not yet been finished, the CPU 1 judges whether or not a mark indicating the beginning of a measure is placed on that analytical point (Steps SP 131 through SP 133). In case no mark indicating the beginning of a measure is placed, the CPU 1 further judges whether or not a mark indicating a rise point in the power information is placed there (Step SP 134). In case there is no mark placed for indicating a rise point, the CPU 1 further judges whether or not a mark indicating the beginning of a rest is placed there (Step SP 135). In case the mark indicating the beginning of a rest is not placed, the CPU 1 increments the parameter  $i$  and returns to the above-mentioned step, SP 132, then ascertaining the presence of a mark on the next analytical point (Step SP 136).

Meanwhile, if any mark is placed on the analytical point  $i$  for the indication of the beginning of a measure or the beginning of a rise point or a rest, the CPU 1 puts a mark on the analytical point thereby to indicate the beginning of a segment, and then increments the parameter  $i$ , thereafter returning to the above-mentioned step, SP 132, and ascertaining whether or not the prescribed mark is attached to the next analytical point (Steps SP 137 and SP 138).

In this manner, the CPU 1 places marks indicating the beginnings of segments one by one on those analytical points which bear a mark indicating the beginning of a measure, a rise point, or the beginning of a rest, and the process soon comes to the final data, and an affirmative result is obtained at the step SP 132. Thereupon the series of processes for placing the mark indicating the beginning of a segment is finished. The process of the steps SP 131 through SP 138 corresponds to the process of step SP 45 of FIGURE 4.

Thus, the CPU 1 finishes the process of segmentation on the basis of the measures and power information, thereafter proceeding to the tuning process as described above.

FIGURE 6 presents the changes in the pitch information, PIT, the power information, POW, and the rise extraction function  $d(i)$  with respect to the one-measure section. Here, the "dual circle" mark represents the beginning of a measure, and the "white star" mark represents a rise point, while the

"circle" mark indicates the beginning of a beat, and the "X" mark indicates the beginning of a semi-quarter note before the matching with a rise point is executed, and the "triangle" mark shows the beginning of a rest. Therefore, in the case of this example of a section corresponding to one measure, the mark indicating the beginning of a segment is placed as shown by the "black circle" mark shown in it as the result of the execution of the series of segmentation processes as described above.

According to the embodiment described above, the system is so designed as to generate input auxiliary rhythm sounds in order to help the users in their input of acoustic signals, thereby offering simplicity and ease of use with regard to the input of acoustic signals and enabling their input with accuracy in terms of rhythm, which results in greater facility in the segmenting of such signals and therefore in improvements upon the precision of the produced musical score data.

The system is arranged in such a way that the information on the input auxiliary rhythm sounds generated at the time of the input are recorded on the same time axis as for the acoustic signals, so that such information may be used for segmenting such signals. This feature enhances accuracy of segmentation, which in turn leads to improvements on the precision of the musical score data produced.

#### Alternative Embodiments

The preferred embodiment, described above, employs the square sum of the acoustic signal as the power information, but another parameter may also be used. For example, the square root of the square sum may be used. Moreover, the rise extraction function has been obtained in the manner expressed in the equation (1), but also another parameter may be employed. It is acceptable to extract the rise in the power information by the application of a function representing only the numerator in the equation (1).

In the preferred embodiment the system takes away the mark of the rise point on the preceding side in case the distance between the preceding and following rise points is short, but it is acceptable to remove the mark of the rise point.

In the preferred embodiment, described above, the system generates the input auxiliary rhythm sounds to permit the users to input the acoustic sounds with ease. However, the rhythm information for assisting the user with the input procedure may be provided in the visual form. For example, it is feasible to display on display unit 5 an image of a baton which moves with the appropriate rhythm.

Also, it is acceptable to use a combination of audio and visual means for indicating rhythm to the user. In this regard, the sounds of a metronome or rhythmic accompanying sounds could be provided as the input auxiliary sounds.

In the preferred embodiment, described above, the system makes use of the information on the beginning of a measure, out of the input auxiliary rhythm information, for performing the segmentation process. However, the information indicating the beginning of a beat, out of the input auxiliary rhythm information, may well be used for performing the segmentation process.

The preferred embodiment uses display unit 5 to output of the musical score data, but a character printing device can be used in its place.

In the preferred embodiment CPU 1 executes all the processes in accordance with the programs stored in memory in the main storage device 3. Yet, some or all of the processes can be executed by a hardware system or sub-system. For example, as illustrated in FIGURE 7, where the identical reference numbers are given for the parts corresponding to those shown in FIGURE 2, the acoustic signals input from the acoustic signal input device 8 can be amplified while there are passed through the amplifying circuit 11 and thereafter channeled through a pre-filter 12 and then fed into the A/D converter 13, where they are converted into digital signals. The acoustic signals as thus converted into digital signals are then processed for autocorrelation analysis by the signal-processing processor 14, which thereby extracts the pitch information or may otherwise extract the power information by processing the signals to find their square sum, and the pitch information or the power information, as the case may be, can then be supplied to the CPU 1 for their processing with the software system. As a signal-processing processor 14 which can be utilized for such a hardware construction (11 through 14), it is possible to use a processor which is capable of performing the real-time processing of the signals and is also provided with the signals for establishing an interface with the host computer (for example,  $\mu$  PD 7720 made by Nippon Electric Corporation).

The preferred embodiment performs the initial segmentation process on the basis of the input auxiliary rhythm information and the power information, but the system can be designed to perform the process on the basis of the input auxiliary rhythm information and the pitch information, or can also be so designed as to perform the process on the basis of the input auxiliary rhythm information and the power information and the pitch information.

The system according to this invention is arranged so as to provide a user with input auxiliary

rhythm information and let the user input acoustic signals, thereby enabling the user to input acoustic signals with greater ease and simplicity, so that he can input the intended acoustic signals with accuracy in terms of rhythm, with the result that greater facility is attained in the performance of the segmentation process for such acoustic signals and that the precision of the musical score data so prepared can be improved positively.

Moreover, the system is designed also to record the input auxiliary rhythm information provided to the users on the same time axis as the acoustic signals, so that the information so recorded may be made available for the process of segmentation process. This feature makes it possible to perform accurate segmentation, thereby enhancing the precision of the musical score data generated by the system.

While this invention has been described in connection with what is presently considered to be the most practical and preferred embodiment, it is to be understood that the invention is not limited to the disclosed embodiment, but, on the contrary, is intended to cover various modifications and equivalent arrangements included within the spirit and scope of the appended claims.

## Claims

1. A method for automatically transcribing music, comprising the steps of:  
capturing acoustic signals;  
storing the acoustic signals in a memory;  
simultaneous with the step of capturing, reporting information on input auxiliary rhythms including at least information on tempo;  
extracting, from said acoustic signals stored in memory, pitch information, which represents the repetitive cycles of their waveforms and their sound pitch, and power information;  
segmenting the acoustic signals on the basis of the pitch information and/or the power information, the process of segmenting including dividing the acoustic signals into sections each of which can be regarded to form a single level in musical interval;  
identifying each of the segments with a level on the axis of absolute musical interval on the basis of the pitch information; and  
displaying/reporting the results of the foregoing steps.

2. An automatic music transcription method according to claim 1, wherein the reporting step comprises the step of reporting with an audio signal.

3. An automatic music transcription method according to claim 1, wherein the reporting step comprises the step of reporting with a video signal.

4. An automatic music transcription method according to claim 1, wherein the reporting step comprises the step of reporting with both audio and video signals.

5. An automatic music transcription method according to one of the preceding claims, further comprising the step of storing the auxiliary rhythms in the memory on the same time axis as that of the acoustic signals.

6. An automatic music transcription method according to one of the preceding claims wherein the segmenting step comprises the steps of:  
first segmenting, on the basis of the input auxiliary rhythm information stored in the memory, the acoustic signals into sections each of which can be regarded as forming one and the same level of musical interval,  
second segmenting, on the basis of the pitch information and the power information, the acoustic signals into sections each of which can be regarded as forming one and the same level of musical interval, and  
third making adjustments to those sections as divided into segments by the first and the second steps.

7. An automatic music transcription system, comprising:

means for receiving acoustic signals to be transcribed;

means for providing auxiliary rhythm information including tempo information, at the time when the acoustic signals are being received;

a memory;

means for processing and storing into the memory the acoustic signals and rhythm information;

pitch and power extraction means for extracting from the acoustic signals stored in memory pitch information, which represents a repetitive cycle of the waveforms of the acoustic signals and a level of musical interval of such signals, and the power information;

segmentation means for dividing the acoustic signals into sections each of which can be regarded as forming one level of musical interval as determined on the basis of the pitch information and/or the power information; and

musical interval identification means for identifying the musical interval of the above-mentioned acoustic signals with reference to an absolute axis of musical interval.

8. An automatic music transcription system according to claim 7, further comprising means for reporting the input auxiliary rhythm information in an audio form.

9. An automatic music transcription system according to claim 7, further comprising means for reporting the input auxiliary rhythm information in a visual form.

10. An automatic music transcription system according to claim 7, further comprising means for reporting the input auxiliary rhythm information in both audio and visual form.

11. An automatic music transcription system according to one of claims 7 to 10, wherein said means for processing and storing comprises means for storing the auxiliary rhythm information and the acoustic signals in memory on the same time axis at the time when the above-mentioned acoustic signals are captured and stored in the memory.

12. An automatic music transcription system according to claim 11, wherein the segmentation means comprises:  
first segmenting section for segmenting, on the basis of the input auxiliary rhythm information stored in the memory, the acoustic signals into sections each of which can be regarded as forming one and the same level of musical interval;  
a second segmenting section for segmenting, on the basis of the pitch information and the power information, the acoustic signals into sections each of which can be regarded as forming one and the same level of musical interval; and  
a third segmenting section for making adjustments to those sections as divided into segments by the first and second segmenting sections.

5

10

15

20

25

30

35

40

45

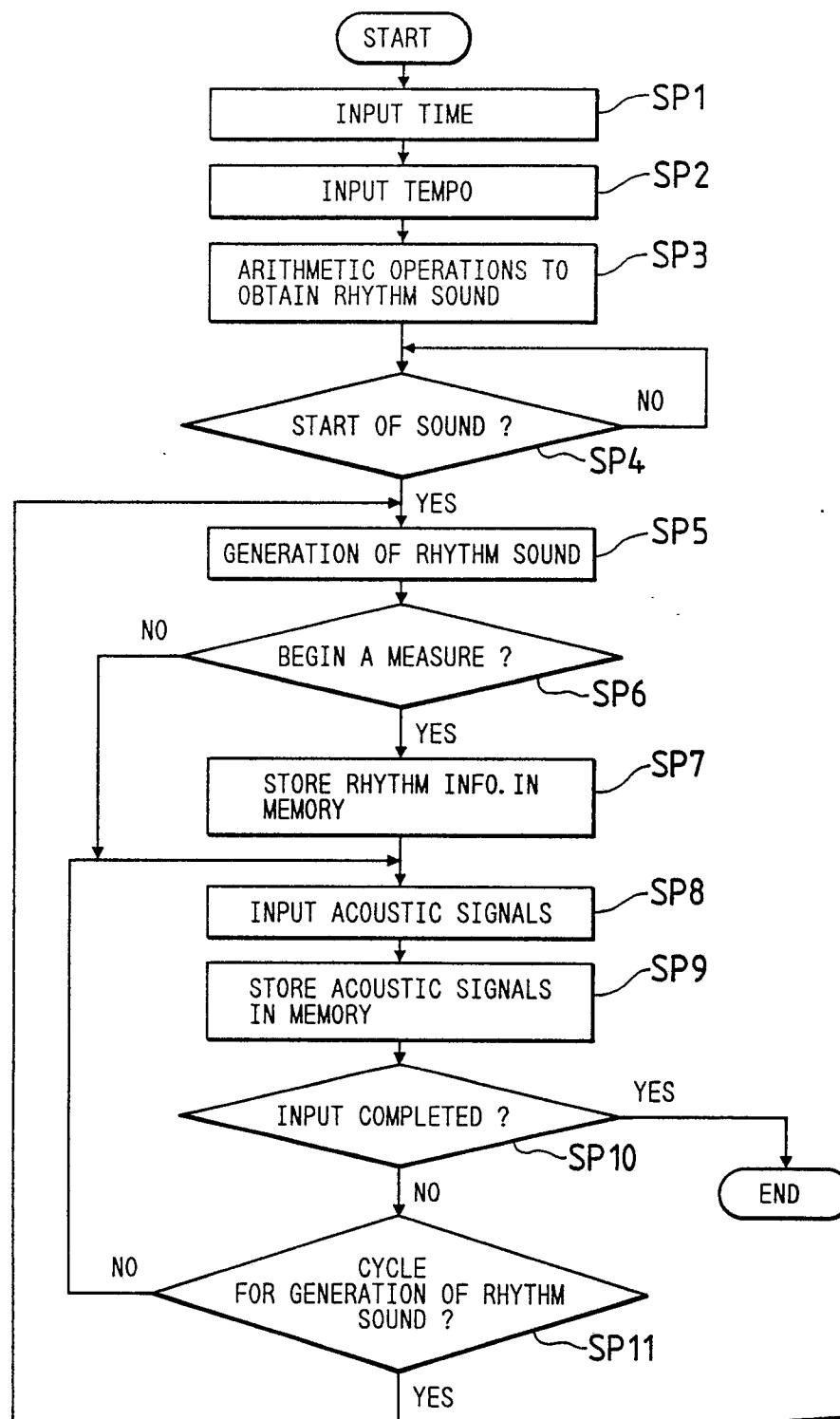
50

55

13

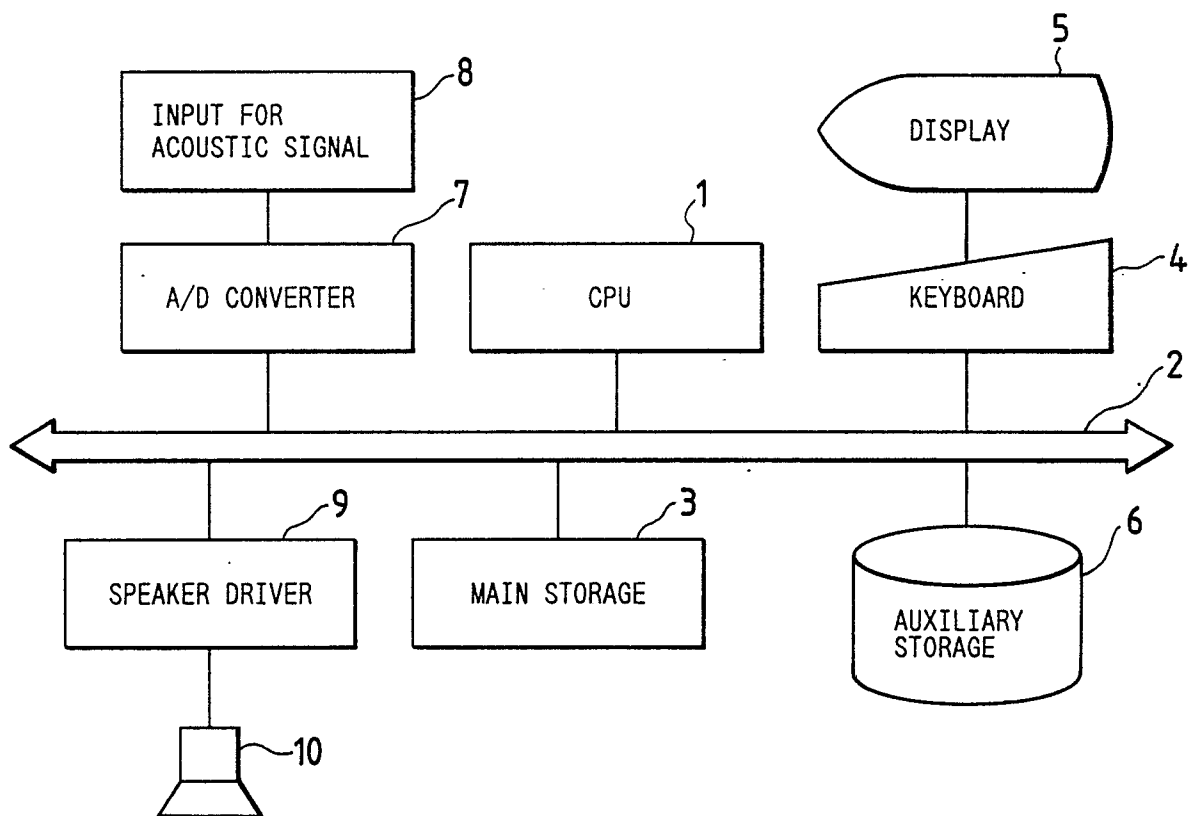


FIG. 1



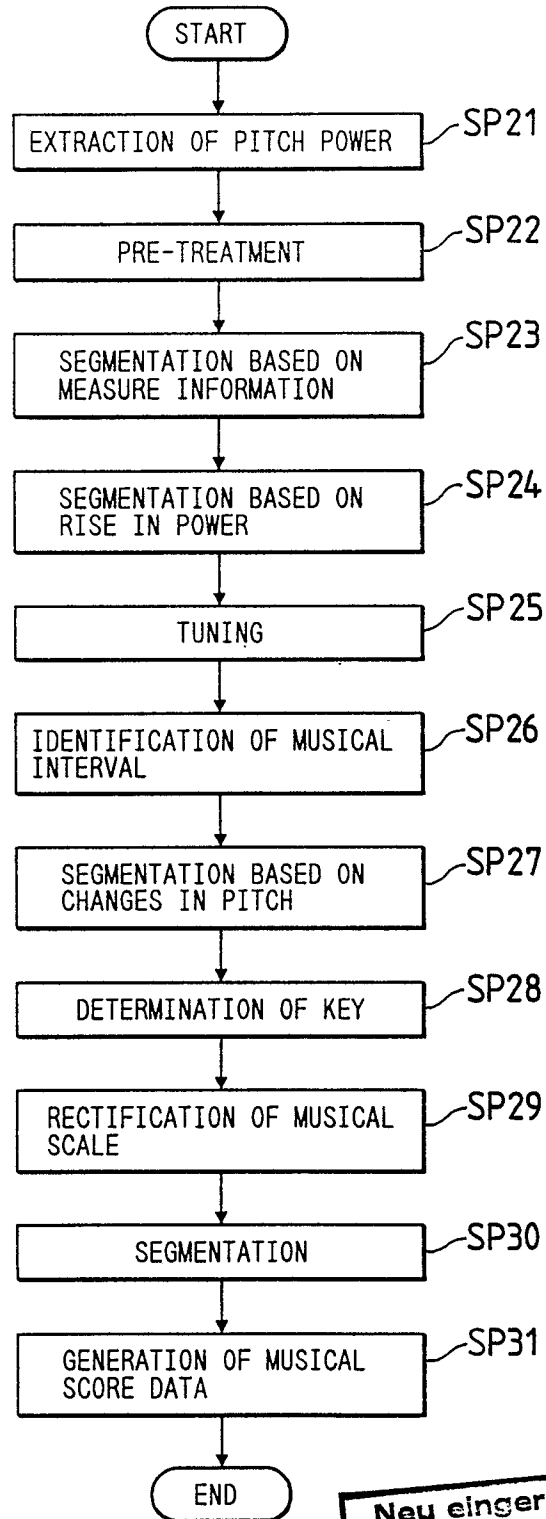
Neu eingereicht / Newly filed  
Nouvellement déposé

FIG. 2



Neu eingereicht / Newly filed  
Nouvellement déposé

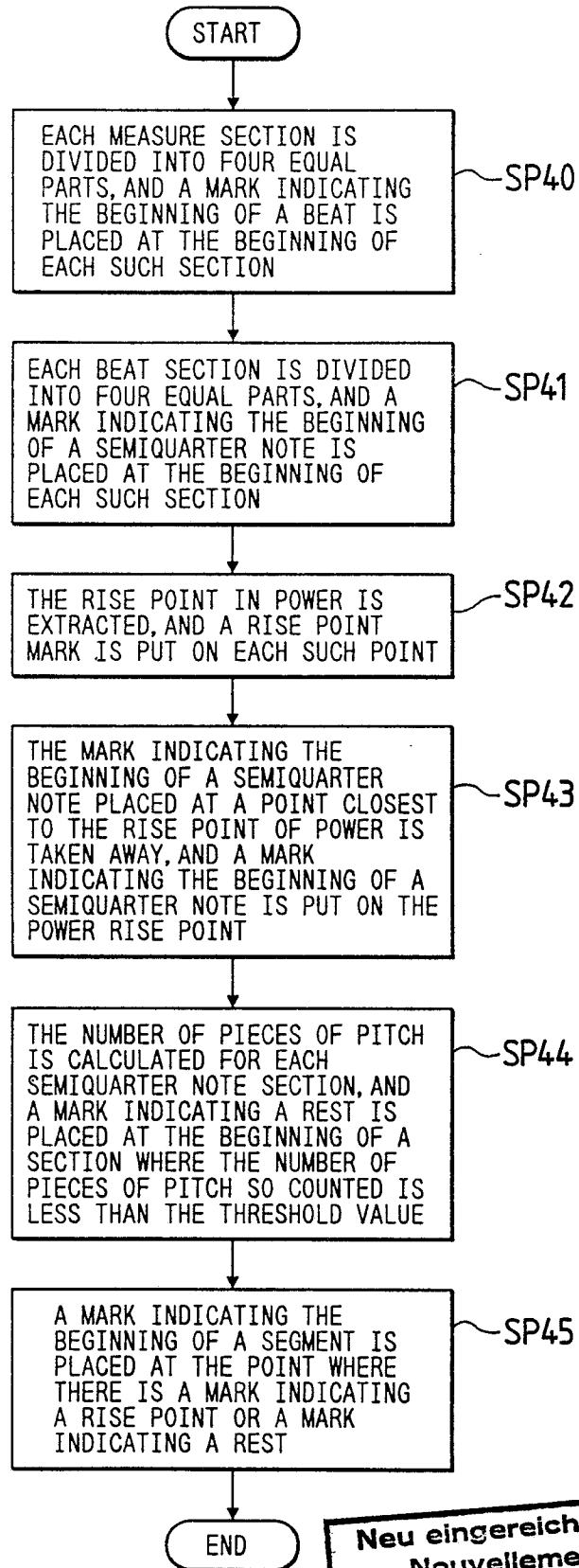
FIG. 3



Neu eingereicht / Newly filed  
Nouvellement déposé



FIG. 4



Neu eingereicht / Newly filed  
Nouvellement déposé

FIG. 5-1

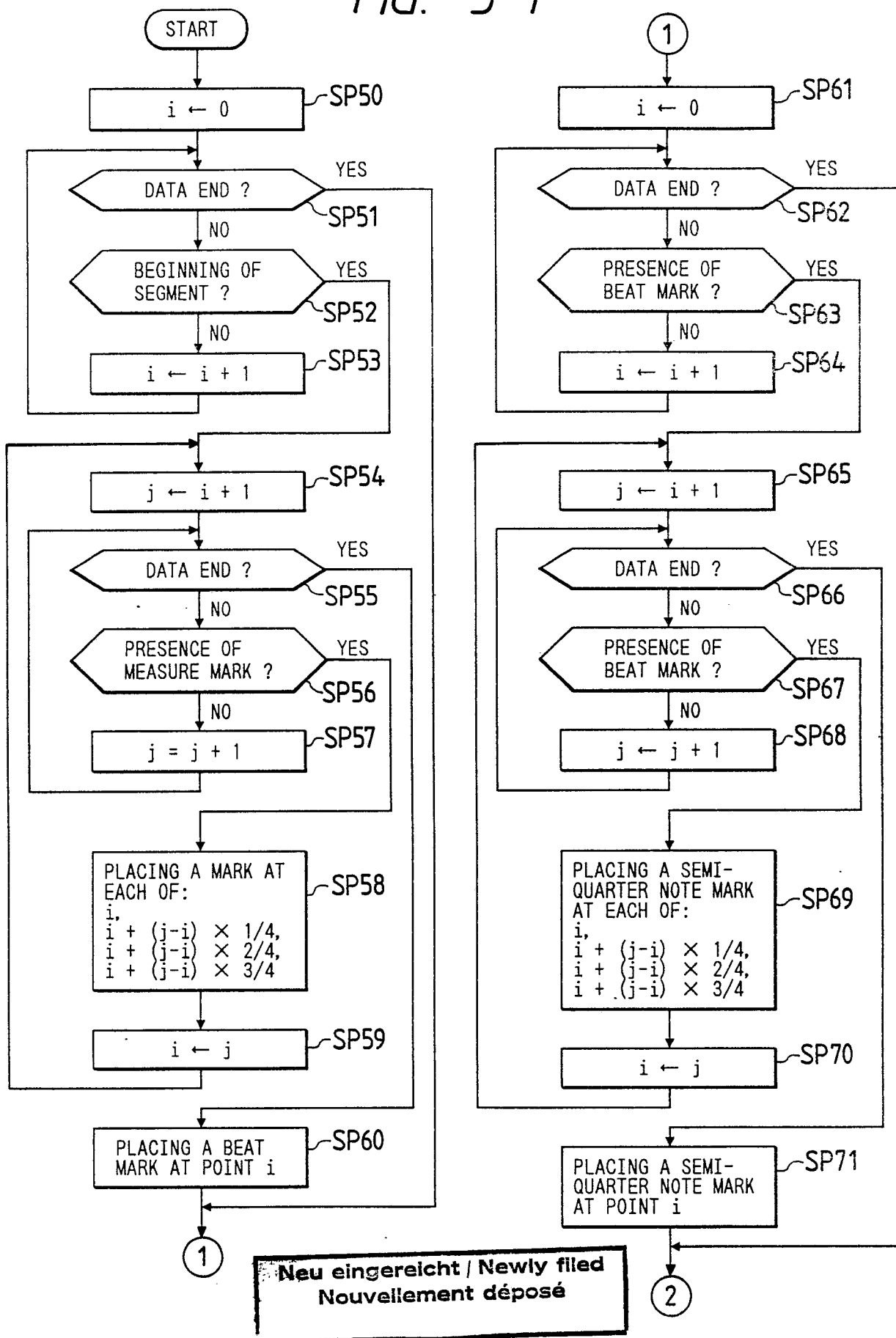


FIG. 5-2

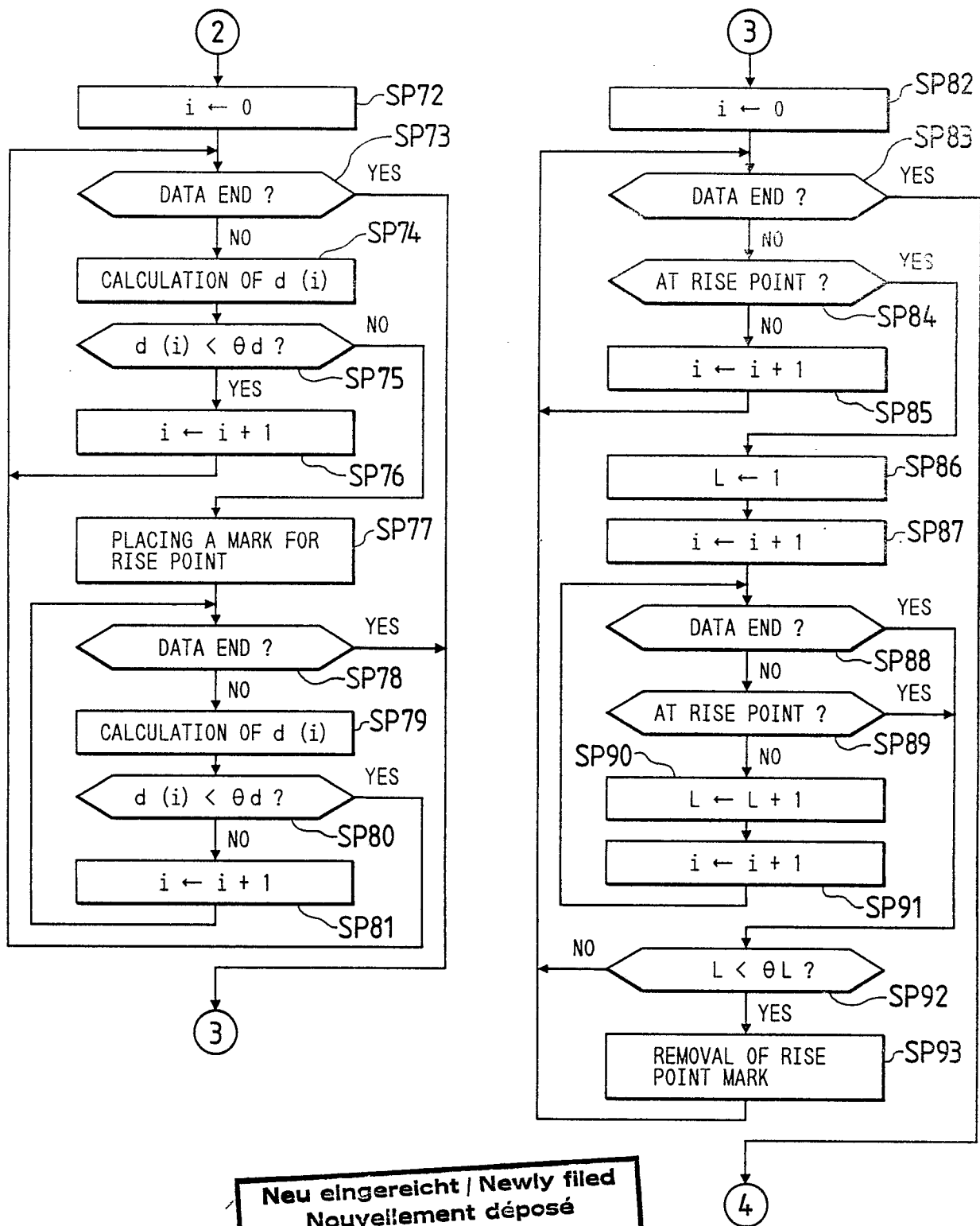
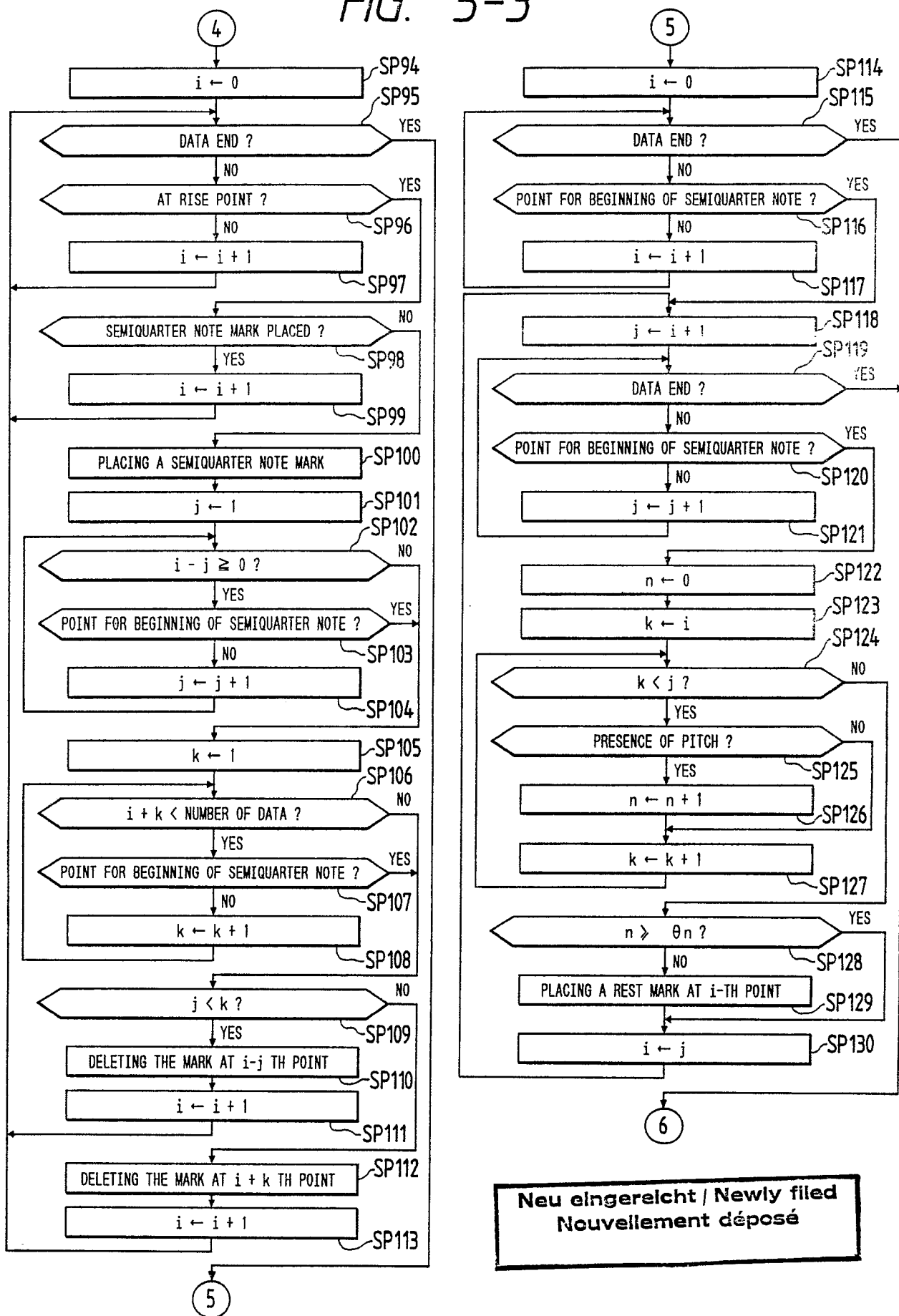
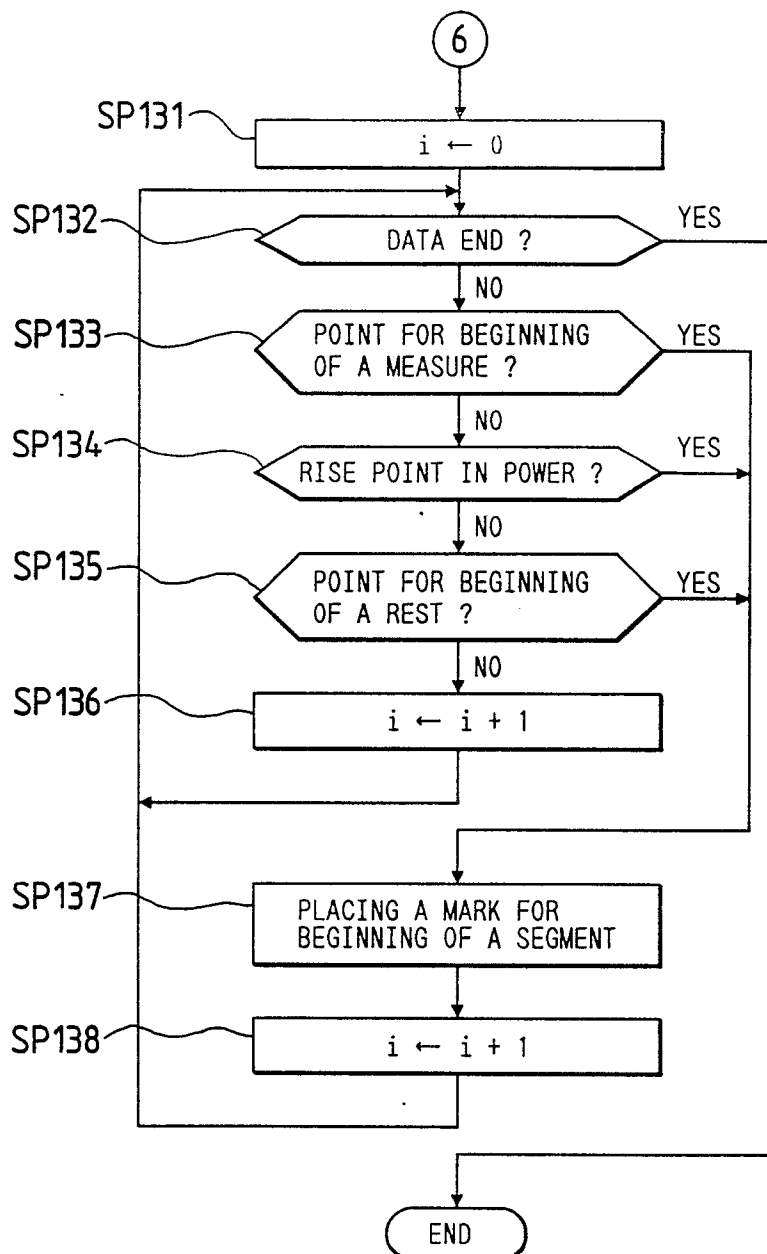


FIG. 5-3



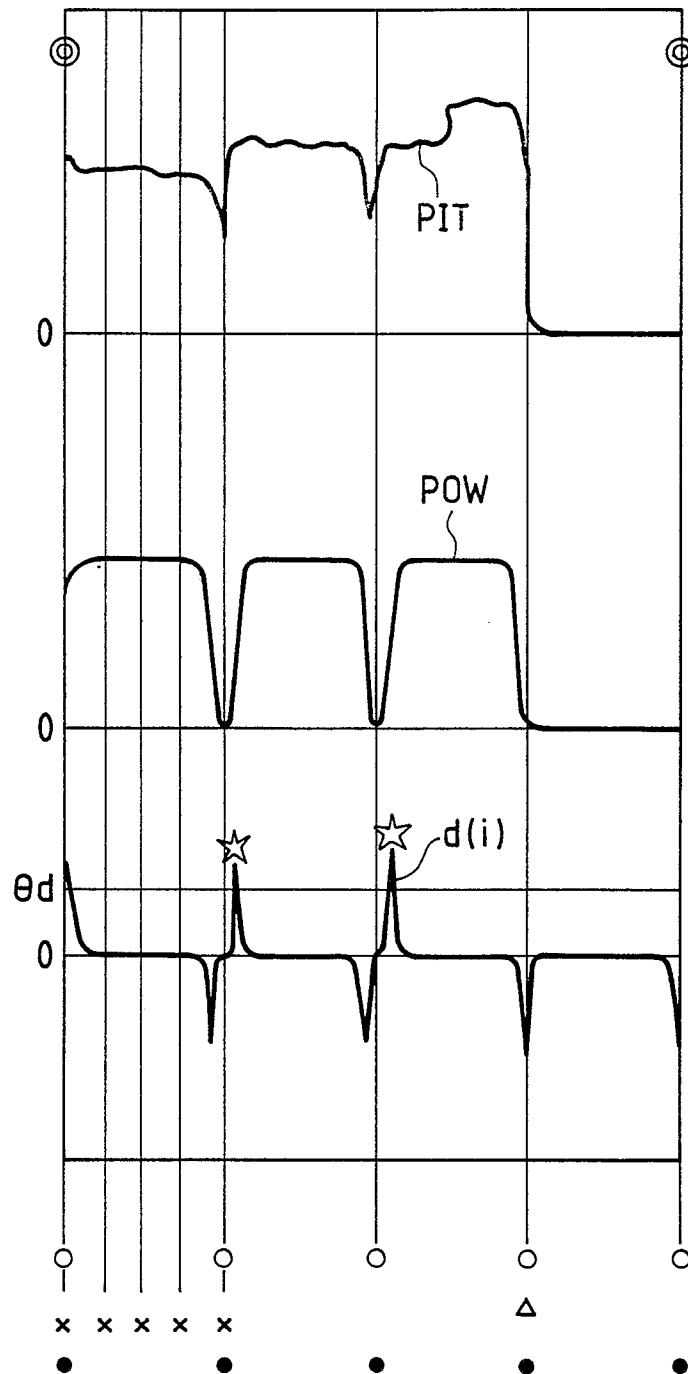
Neu eingereicht / Newly filed  
Nouvellement déposé

FIG. 5-4



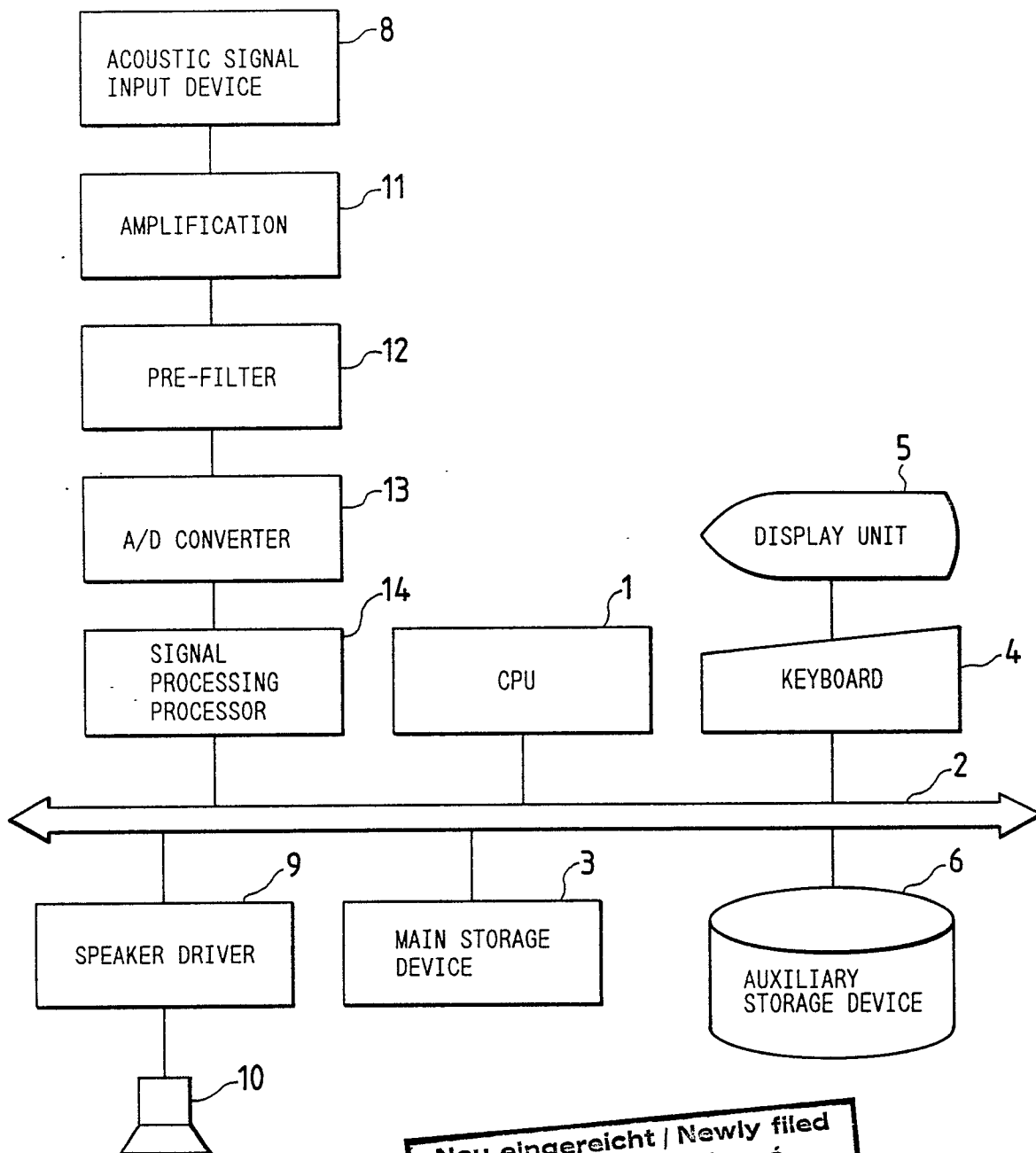
Neu eingereicht / Newly filed  
Nouvellement déposé

FIG. 6



Neu eingereicht / Newly filed  
Nouvellement déposé

FIG. 7



Neu eingereicht / Newly filed  
Nouvellement déposé