

19



Europäisches Patentamt
European Patent Office
Office européen des brevets



11 Publication number:

0 331 107 B1

12

EUROPEAN PATENT SPECIFICATION

45 Date of publication of patent specification: **21.07.93** 51 Int. Cl.⁵: **G10G 3/04**

21 Application number: **89103498.5**

22 Date of filing: **28.02.89**

54 **Method for transcribing music and apparatus therefore.**

30 Priority: **29.02.88 JP 46111/88**
29.02.88 JP 46112/88
29.02.88 JP 46118/88
29.02.88 JP 46125/88
29.02.88 JP 46126/88
29.02.88 JP 46127/88
29.02.88 JP 46128/88

43 Date of publication of application:
06.09.89 Bulletin 89/36

45 Publication of the grant of the patent:
21.07.93 Bulletin 93/29

84 Designated Contracting States:
DE FR GB

56 References cited:
EP-A- 0 113 257
GB-A- 2 139 405
US-A- 4 392 409
US-A- 4 479 416

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

Proprietor: **NEC CORPORATION**
7-1, Shiba 5-chome Minato-ku

Tokyo 108-01(JP)

72 Inventor: **Tsuruta, Shichirou**
c/o NEC Home Elec. Ltd. No. 5-24 Miyahara
3-chome
Osaka-shi Osaka(JP)
Inventor: **Takashima, Yosuke**
c/o NEC Corporation No. 33-1, Shiba
5-chome
Minato-ku Tokyo(JP)
Inventor: **Fujimoto, Masaki**
c/o NEC Scientific Inf.System Dev.Ltd.
No.7-15
Shiba 5-chome Minato-ku Tokyo(JP)
Inventor: **Mizuno, Masanori**
c/o NEC Scientific Inf.System Dev.Ltd. No.
7-15
Shiba 5-chome Minato-ku Tokyo(JP)

74 Representative: **Diehl, Hermann Dr. et al**
Diehl & Glaeser, Hiltl & Partner Flüggen-
strasse 13
W-8000 München 19 (DE)

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid (Art. 99(1) European patent convention).

EP 0 331 107 B1

Description

The present invention relates to a method of automatically transcribing music and an apparatus therefore for preparing musical score transcription data from vocal sounds of songs, humming voices, and musical instrument sounds.

One such transcription system as is known in the art is disclosed in UK patent application GB-A-2 139 405A, in which musical pitch and time value are determined using a fast Fourier transform operation and a power spectrum calculation, and wherein resulting values are displayed.

For an automatic music transcription system for transforming acoustic signals, such as those of vocal sounds of songs, hummed voices, and musical instrument sounds into musical score data, it is necessary to detect sound lengths, musical intervals, keys, times, and tempos, which are basic items of information for musical scores, out of the acoustic signals.

Generally, since acoustic signals are the kind of signals which contain repetitions of fundamental waveforms in continuum, it is not possible immediately to obtain the above-mentioned items of information.

Therefore, the present applicants have already proposed an automatic music transcription system as disclosed, for example, in Unexamined Japanese Patent Application No. 62-178409.

This automatic music transcription system shown in Fig. 1 is provided with a autocorrelation analyzing means 14 for converting hummed vocal sound signals 11 into digital signals by means of an analog/digital (A/D) converter 12 and thereby developing vocal sound data 13 and for extracting pitch information and sound power information 15 from the vocal sound data 13, a segmenting means 16 for dividing the input song or hummed sounds into a plural number of segments on the basis of the sound power information extracted by the afore-mentioned autocorrelation analyzing means, a musical interval identifying means 17 for identifying the musical interval on the basis of the afore-mentioned pitch data with respect to each of the segments as established by the afore-mentioned segmenting means, a key determining means 18 for determining the key of the input song or hummed vocal sounds on the basis of the musical interval as identified by the afore-mentioned musical interval identifying means, a tempo and time determining means 19 for determining the tempo and time of the input song or hummed vocal sounds on the basis of the segments established by division by the afore-mentioned segmenting means, a musical score data compiling means 110 for preparing musical score data on the basis of the results made available by the afore-mentioned segmenting means, musical interval identifying means, and key determining means, and tempo and time determining means, and a musical score data outputting means 111 for generating as output the musical score data 112 prepared by the afore-mentioned musical score compiling means.

It is to be noted in this regard that such acoustic signals as those of vocal sounds in songs, hummed voices, and musical instrument sounds consist of repetitions of fundamental waveforms. In an automatic music transcription system for transforming such acoustic signals into musical score data, it is necessary first to extract for each analytical cycle the repetitive frequency of the fundamental waveform in the acoustic signal. This frequency is hereinafter referred to as "the pitch frequency," and the cycle corresponding to this is called "the pitch cycle," and the concept representing the combination of these is to be known as "pitch". In order accurately to determine various kinds of information on such items as musical interval and sound length in acoustic signals.

Among the available extracting methods are frequency analysis and autocorrelation analysis, which have attained their development in the fields of vocal sound synthesis and vocal sound recognition. Yet, autocorrelation analysis has hitherto been employed because it can extract pitch without being affected by noises in the environment and additionally permits easy processing.

In the automatic musical score transcription system mentioned above, the system finds the autocorrelation function after it converts acoustic signals into digital signals. Therefore, an autocorrelation function can be found only for each sampling cycle.

Accordingly, pitch can be extracted only by the resolution determined by this sampling cycle. If the resolution of a pitch so extracted is low, then the musical interval and sound length determined by the processes described later will have a low degree of accuracy.

Then, it is conceivable to use a higher frequency for sampling, but such an approach is liable to result in the inability of the system to perform real-time processing, as well as a larger-sized construction of the apparatus for the automatic music transcription system and consequently a more expensive price for it, in consequence of an increase in the amount of data to be processed for the arithmetic operations, such as those for the calculation of the autocorrelation function.

Acoustic signals have the characteristic feature that their power is augmented immediately after a change in sound, and this feature is utilized in the segmentation of a stream of sounds on the basis of power information.

However, acoustic signals, particularly those appearing in songs sung by a man, do not necessarily take any specific pattern in the change of their power information, but have fluctuations in relation to the pattern of change. In addition, such signals also contain abrupt sounds, such as outside noises. In these circumstances, a simple segmentation of sound with attention paid to the change in the power information has not necessarily led to any good division of individual sounds.

In this regard, it is noted that acoustic signals generated by a man are not stable in sound length, either. That is, such signals have much fluctuations in pitch. This has caused an obstacle to the performance of good segmentation based on pitch information.

Thus, in view of the fluctuations existing in pitch information, the conventional systems are so designed as to treat two or more sounds as a single segment in some cases.

Moreover, even those sounds generated by musical instruments have not in some cases lent itself readily to segmentation based on pitch information on account of ambient noises intruding into the pitch information after they are captured by the acoustic signal input apparatus for converting acoustic signals into electrical signals.

Now that musical intervals, times, tempos, etc. are to be determined on the basis of sound segments (sound length), the process of segmentation is a very important factor particularly for the preparation of musical score data, and, as low accuracy of segmentation causes a considerable decline in the accuracy of the ultimately developed musical score data, it is to be desired that the accuracy of the segmentation process itself based on the power information will be improved both for the case in which the final segmentation is to be performed on the basis of both the results of the segmentation based on the pitch information and the results of the segmentation based on the power information and the case in which the final segmentation is performed on the basis of the power information.

Now, an effort to identify segments consisting of acoustic signals with reference to a musical interval on the axis of an absolute musical interval would lead to the finding that acoustic signals, particularly those acoustic signals uttered by a man, are not stable in their musical interval and have considerable fluctuations in pitch even when the same pitch (one tone) is intended. This has made it very difficult to perform the identification of a musical interval of such signals.

Above all, when a transition occurs from one sound to another, it often happens that a smooth transition cannot be made to the pitch of the following sound, with fluctuations in pitch before and after it. Consequently, such a part was often taken as a section of another sound in the course of a segmentation process with the result that it was identified as belonging to a different pitch level in the identification of a musical interval.

In order to explain this in specific terms, methods permitting simplicity in arithmetic operation, such as a method of identifying a given sound with a pitch closest on the absolute axis to the average value of the pitch information within the segment or with the pitch closest on the absolute axis to the medium value of the pitch information of the segment, are considered for the automatic music transcription system mentioned above. With a method like this, it is possible to identify the musical interval well, even if the acoustic signal has a fluctuation, in case the interval difference between two sounds adjacent to each other on a musical scale is a whole tone, for example, do and re on the C-major scale, but, if the difference in the interval between two adjacent sounds is a semitone, for example, the case of mi and fa on the C-major scale, there may sometimes be a lack of accuracy in the identification of the musical interval because of fluctuations in the pitch of the acoustic signals. For example, there were some cases in which a sound intended for mi on the C-major scale was identified as fa.

Now that the musical interval is a fundamental element, together with sound length, it is necessary to identify this accurately, and, if it cannot be identified accurately, the accuracy of the resulting musical score data will be low.

On the other hand, the key of an acoustic signal is not merely an element of musical score data, but also gives an important clue to the determination of a musical interval since a key has a certain kind of relationship with a musical interval and above all with the frequency of occurrence of a musical interval. Accordingly, for improving the accuracy of a musical interval, it is desirable to determine the key and to review the identified musical interval, and it is to be desired that the key of acoustic signals is determined well.

Furthermore, as mentioned above, the musical intervals of acoustic signals, particularly those of the voices uttered by a man, deviate from the absolute musical interval, and, the greater such a deviation is, the more inaccurate the musical interval identified on the musical interval axis is, which has resulted in the lower accuracy of music transcription data prepared ultimately.

The automatic music transcription system and apparatus disclosed in the published patent application No. 62-178409 by the present applicants as mentioned hereinabove could achieve low accuracy in the

musical score data obtained, so that it was found difficult to put them to practical use.

The present invention, has been made in consideration of the problems mentioned hereinabove. Therefore, a primary object of the invention is to provide a practically usable automatic music transcription system and apparatus which can improve the accuracy of the final musical score data.

Another object of the present invention is to provide an automatic music transcription method and apparatus which can further improve the accuracy of the final musical score data through their good performance of segmentation based on power information or pitch information without being influenced by fluctuations in acoustic signals or the abrupt intrusion of outside sounds.

Still another object of the present invention is to make a proposal for a novel method of identifying musical intervals which can identify musical scales with accuracy and to provide an automatic music transcription system and apparatus which are capable of making a further improvement on the accuracy of the final musical score data.

Still another object of the present invention is to provide an automatic music transcription method and apparatus which can make further improvements in accuracy of the final musical score data by virtue of their ability to obtain more accurate information on the musical interval through correction of the pitch of a segment identified with a musical interval different from that intended by the singer or the like on account of fluctuations occurring in the musical interval at the time of transition to the next sound in an acoustic signal, making such correction with reference to the musical interval information on the preceding segment and the following segment.

Still another object of the present invention is to provide an automatic music transcription method and apparatus which are capable of accurately determining the key of acoustic signals and making further improvements on the accuracy of the final musical score data.

Still another object of the present invention is to provide an automatic music transcription method and apparatus which are designed to be capable of detecting the amount of deviation of the musical interval axis of an acoustic signal from the axis of the absolute musical interval, making a correction of the pitch information in proportion to such a deviation, and thereby making it possible to compile musical score data better in the subsequent process.

Still another object of the present invention is to provide a pitch extracting method and pitch extracting apparatus which are capable of extracting the pitch of an acoustic signal with high accuracy without employing any higher sampling frequency.

In order to attain these and other objects, the invention provides a method for transcribing music as disclosed in claim 1 and an apparatus as described in claim 22. Further advantageous features of this method are evident from the dependent claims. The automatic music transcription system according to the present invention consists in extracting the pitch information and the power information from the input acoustic signal, correcting the pitch information in proportion to the amount of deviation of the musical interval axis for the afore-said acoustic signal from the absolute musical interval axis, dividing the acoustic signal into single sound segments on the basis of the corrected pitch information while also dividing the acoustic signal into single-sound segments on the basis of the changes in the power information, making more detailed divisions of the acoustic signal on the basis of the segment information obtained from both of these, identifying the musical intervals of the acoustic signals in the individual segments along the axis of the absolute musical interval with reference to the pitch information, and moreover dividing the acoustic signal again into single-sound segments on the basis of the point whether or not the identified musical intervals of the segments in continuum are identical, determining the key of the acoustic signal on the basis of the extracted pitch information, correcting the prescribed musical interval on the musical scale for the determined key on the basis of the pitch information, determining the time and tempo of the acoustic signal on the basis of the segment information, and finally compiling musical score data from the information on the determined musical interval, sound length, key, time, and tempo.

Furthermore, in order to achieve the objects mentioned hereinabove, the automatic music transcription system according to the present invention is provided with a means of extracting from the input acoustic signal the pitch information and the power information thereof, a means of correcting the pitch information in accordance with the amount of deviation of the musical interval for the acoustic signal in relation to the axis of the absolute musical interval, a means of dividing the acoustic signal into single-sound segments on the basis of the corrected pitch information, a means of dividing the acoustic signal into single-sound segments on the basis of the changes in the power information, a means of making further divisions of the acoustic signal into segments on the basis of both of these sets of segment information thus made available, a means of identifying the musical intervals for the acoustic signals in the individual segments along the axis of the absolute musical interval, a means of dividing the acoustic signal again into single-sound segments on the basis of the point whether or not the musical intervals of the identified segments in continuum are

identical, a means of determining the key for the acoustic signal on the basis of the extracted pitch information, a means of correcting the prescribed musical interval on the determined key on the basis of the pitch information, a means of determining the time and tempo of the acoustic signal on the basis of the segment information, and a means of finally compiling musical score data from the information on the musical interval, sound length, key, time and tempo so determined.

Furthermore, in order to achieve the above-mentioned objects, the automatic music transcription system according to the present invention is characterized by comprising a means of inputting acoustic signals, a means of amplifying the acoustic signals thus input, a means of converting the amplified analog signals into digital signals, a means of extracting the pitch information by performing autocorrelation analysis of the digital acoustic signals and extracting the power information by performing the operations for finding the square sum, a storage means for keeping in memory the prescribed music-transcribing procedure, a controlling means for executing the music-transcribing procedure kept in memory in the storage means, a means of starting the processing by the control means, and a means of generating as required the output of the musical score data obtained by the processing, with the input means for acoustic signals, the amplifying means, the analog/digital converting means, and the means of extracting the pitch information and the power information being constructed in hardware.

The present invention has made it possible to provide an automatic music transcription system with sufficient capabilities for its practical-object application owing to the extremely significant improvement in its accuracy in generating the final musical score data since the system according to the present invention can accurately extract pitch information and power information from such acoustic signals as vocal sounds in songs, humming voices, and musical instrument sounds, divide the acoustic signals accurately into single-sound segments on the basis of such information, thereby identifying the musical interval and the key with high accuracy, these performance features therefore proving effective in reducing the influence of the noise components and power fluctuations in the acoustic signals in processing the input acoustic signals.

Fig. 1 is a block diagram illustrating the automatic music transcription system at a step leading to the present invention.

Fig. 2 is a block diagram illustrating the first embodiment of the construction for the automatic music transcription system according to the present invention.

Fig. 3 is a flow chart showing the procedure for the automatic music transcription process in the system for the first embodiment of the present invention.

Fig. 4 is a summary flow chart illustrating the segmentation process based on the power information pertinent to the present invention.

Fig. 5 is a flow chart illustrating an example of the segmentation process in greater detail.

Fig. 6 is a characteristic curve chart illustrating one example of segmentation by such a process.

Fig. 7 is a summary flow chart illustrating another example of the segmentation process based on the power information to be provided by the present invention.

Fig. 8 is a flow chart illustrating the segmentation process in greater detail.

Fig. 9 is a flow chart illustrating an example of the segmentation process based on the power information to be provided by the present invention.

Fig. 10 is a characteristic curve chart presenting the chronological change of the power information together with the results of the segmentation.

Fig. 11 is a flow chart illustrating an example of the segmentation process based on the power information to be provided by the present invention.

Fig. 12 is a characteristic curve chart presenting the chronological changes of the power information and those of the rise extracting functions, together with the results of the segmentation.

Fig. 13 and Fig. 14 are flow charts each illustrating an example of the segmentation process based on the power information to be provided by the present invention.

Fig. 15 is a characteristic curve chart presenting the chronological changes of the power information and the rise extracting functions, together with the results of the segmentation.

Fig. 16 and Fig. 17 are flow charts each illustrating an example of the segmentation process based on the pitch information to be provided by the present invention.

Fig. 18 is a schematic drawing provided for providing an explanation of the length of the series.

Fig. 19 is a flow chart illustrating the reviewing process for the segmentation pertinent to the present invention.

Fig. 20 is a schematic drawing provided for an explanation of the reviewing process.

Fig. 21 is a flow chart illustrating the musical interval identifying process according to the present invention.

Fig. 22 is a schematic drawing provided for an explanation of the distance of the pitch information to the axis of the absolute musical interval in each segment.

Fig. 23 is a flow chart illustrating an example of the musical interval identifying process according to the present invention.

5 Fig. 24 is a schematic drawing illustrating one example by such a musical interval identifying process.

Fig. 25 is a flow chart illustrating an example of the musical interval identifying process according to the present invention.

Fig. 26 is a schematic drawing illustrating one example by such a musical interval identifying process.

10 Fig. 27 is a flow chart illustrating one example of the musical interval identifying process according to the present invention.

Fig. 28 is a schematic drawing showing one example by such a musical interval identifying process.

Fig. 29 is a flow chart illustrating an example of the process for correcting the identified musical interval according to the present invention.

15 Fig. 30 is a schematic drawing illustrating one example of the correction of such an identified musical interval.

Fig. 31 is a flow chart illustrating an example of the musical interval identifying process according to the present invention.

Fig. 32 is a schematic drawing illustrating one example by such a musical interval identifying process.

20 Fig. 33 is a flow chart illustrating an example of the musical interval identifying process according to the present invention.

Fig. 34 is a chart for explaining the length of the series applicable to the present invention.

Fig. 35 is a schematic drawing illustrating one example by such a musical interval identifying process.

Fig. 36 is a flow chart illustrating an example of the process for correcting the identified musical interval according to the present invention.

25 Fig. 37 is a schematic drawing provided for an explanation of such a correcting process for the identified musical interval.

Fig. 38 is a flow chart illustrating an example of the key determining process according to the present invention.

30 Fig. 39 is a table presenting some examples of the weighing coefficients for each musical scale established in accordance with each key.

Fig. 40 is a flow chart illustrating an example of the key determining process according to the present invention.

Fig. 41 is a flow chart illustrating an example of the tuning process according to the present invention.

Fig. 42 is a histogram showing the state of distribution of the pitch information.

35 Fig. 43 is a flow chart showing an example of the pitch extracting process according to the present invention.

Fig. 44 is a schematic drawing presenting the autocorrelation function curves to be used for the pitch extracting process.

40 Fig. 45 is a flow chart illustrating an example of the pitch extracting process according to the present invention.

Fig. 46 is a schematic drawing showing the autocorrelation function curves to be used for the pitch extracting process.

Fig. 47 is a block diagram illustrating the second embodiment of the construction of the automatic music transcription system.

45 In the following part, a detailed description is made of the various embodiments of the present invention with reference to accompanying drawings.

Fig. 2 is a block diagram illustrating the construction of the automatic music transcription system to which the first embodiment according to the present invention is applied, and Fig. 3 is a flow chart illustrating the processing procedure for the system.

50 In Fig. 2, the Central Processing Unit (CPU) 1 performs overall control for the entire system and executes the music score processing program which is shown in Fig. 3 and stored in the main storage device 3 connected to the CPU through the bus 2, to which the keyboard 4, as an input device, the display unit 5, as an output device, the auxiliary memory device 6 for use as working memory, and the analog/digital converter 7 are connected in addition to the CPU 1 and the main storage device 3.

55 To the analog/digital converter 7 is connected, for example, the acoustic signal input device 8, which is composed of a microphone. This acoustic signal input device 8 captures the acoustic signals in vocal songs uttered by the user and then transforms the signals into electrical signals and outputs the electrical signals to the analog/digital converter 7.

The CPU 1 begins the music transcription process when it receives a command to that effect as entered on the keyboard input device 4, and executes the program stored in the main storage device 3, temporarily storing the acoustic signals as converted into digital signals by the analog/digital converter 7 in the auxiliary memory device 6 and thereafter converting these acoustic signals into musical score data by executing the above- mentioned program, so that the musical score data may be output as required.

Next, the processing for musical score transcription after the CPU 1 has taken up the acoustic signals for its program execution is described in detail with reference to the flow chart shown in terms of functional levels in Fig. 3.

First, the CPU 1 extracts the pitch information for the acoustic signals for each analytical cycle through its autocorrelation analysis of the acoustic signals and also extracts the power information for each analytical cycle by processing the acoustic signals to find the square sum, and then performs such post-treatments as the elimination of noises and an interpolation operation (Steps SP 1 and SP 2). Thereafter, the CPU 1 calculates, with respect to the pitch information, the amount of deviation of the musical interval axis of the acoustic signal in relation to the axis of the absolute musical interval on the basis of the state of distribution around the musical interval axis and then performs the tuning process (Step SP 3), which consists in causing the obtained pitch information to shift in proportion to the amount of deviation of the musical interval axis. In other words, the CPU makes a correction of the pitch information in such a way that the difference between the musical interval axis recorded for the acoustic signals generated by the singer or the musical instrument and the axis of the absolute musical interval will be smaller.

Then, the CPU 1 executes the segmentation process, which divides the acoustic signals into single-sound segments, with a continuous duration of pitch information in which the obtained pitch information can be regarded as indicating one musical interval, and executes the segmentation process again on the basis of the changes in the obtained power information (Steps SP 4 and SP 5). On the basis of these sets of segment information, the CPU 1 calculates the standard lengths corresponding respectively to the time lengths of a half note and an eighth note and so forth and execute the segmentation process in further detail on the basis of such standard lengths (Step SP 6).

The CPU 1 thus identifies the musical interval of a given segment with the musical interval on the absolute musical interval axis to which the relevant pitch information is considered to be closest as judged on the basis of the pitch information of the segment obtained by such segmentation 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 7 and SP 8).

After that, the CPU 1 finds the product sum of the frequency of occurrence of the musical interval obtained by working out the classified total of the pitch information around the musical interval axis after tuning and the certain prescribed weighing coefficient determined in correspondence to the key, and, on the basis of the maximum information of this product sum, 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 9 and SP 10). Next, CPU 1 executes a review of the segmentation results on the basis of whether or not the finally determined musical interval contain identical segments in continuum or whether or not there is any change in power and performs the final segmentation process (Step SP 11).

When the musical interval and the segments are determined in this manner, the CPU 1 extracts the measures from the viewpoint that a measure begins with the first beat, that the last tone in a phrase does not extend to the next measure, that there is a division for each measure, and so forth, determines the time on the basis of this measure information and the segmentation information, and determines the tempo on the basis of this determined time information and the length of a measure (Steps SP 12 and SP 13).

Then, the CPU 1 compiles musical score data finally by putting in order the determined musical interval, sound length, key, time, and tempo information (Step SP 14).

Segmentation Based on Power Information

Next, a detailed explanation is given in specific terms, with reference to the flow charts in Fig. 5 and Fig. 4, in respect of the segmentation process (Step SP 5 in Fig. 3) based on the power information on those acoustic signals applicable to an automatic music transcription system like this. In this regard, please note that Fig. 4 gives a flow chart illustrating such a process at the functional level while Fig. 5 presents a flow chart illustrating greater details of what is shown in Fig. 4.

Moreover, for the power information on the acoustic signals, 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 to represent the power information on that analytical cycle.

The CPU 1 compares the power information at each analytical point with the threshold value divides the acoustic signal between a section larger than the threshold value and a section smaller than the value, treating the section larger than the threshold value as the segment for the effective section and the section smaller than the threshold value as the segment of the invalid section and placing a mark for the beginning of an effective segment to the initial part of the effective section and placing a mark for the beginning of an invalid segment to the initial part of the invalid section (Steps SP 15 and SP 16). This feature has been incorporated in the system in view of the fact that a failure often occurs in the identification of a musical interval because of a lack of stability often appearing in the musical interval of acoustic signals in the range where the power information is small and also that this feature serves the object of detecting rest sections.

Then, the CPU 1 performs arithmetic operations to find a function for the variation of the power information within the effective segment derived by the division mentioned above and extracts the point of change in the rising of the power information on the basis of this function of variation, and then the CPU divides the effective segment into smaller parts at the point of change in the rise as extracted, placing a mark for the beginning of an effective segment at the point so determined (Steps SP 17 and SP 18). This feature has been introduced because the above-mentioned process alone is liable to generate a segment containing two or more sounds since there may be a transition from a sound to the next sound while the power is maintained at a somewhat high level, so that such a segment may be divided further, taking advantage of the notable fact that such a segment shows an increase of power at the start of the next sound.

Thereafter, the CPU 1 measures the lengths of the individual segments, regardless of the point whether they are effective segments or invalid ones, connecting any segment with a length shorter than the prescribed length to the immediately preceding segment to form one segment (Steps SP 19 and SP 20). This feature has been adopted in view of the fact that signals may sometimes be divided into minute fragmentary segments as the result of the presence of noises or the like, so that such a fragmentary segment may be connected to the other segment. Also, this feature is used for the object of connecting a plural number of segments resulting from the further division of segments on the basis of the point of change in the rise as mentioned above.

Next, this process is explained in greater detail with reference to the flow chart in Fig. 5.

The CPU 1 first clears the parameter t for the analytical point to zero, and then, ascertaining that the analytical point data to be processed has not yet been completed, the CPU judges whether or not the power information (Power (t)) of the acoustic signal at the analytical point is smaller than the threshold value power (Steps SP 21 - SP 23).

In case the power information, Power (t), is any smaller than the threshold value p , the CPU 1 increments the parameter t for the analytical point again and, returning again to the Step SP22, passes judgment on the power information at the next analytical point (Step SP 24).

On the other hand, the CPU 1 places a mark for the beginning point of an effective segment at that analytical point in case it finds at the Step SP 23 that the value of the power information, Power (t) is above the threshold value p , and moves on to the processing of the subsequent steps beginning with the next Step SP 26 (Step SP 25).

At this time, the CPU 1 ascertains that the processing has not yet been completed on all the analytical points and judges again whether or not the value of the power information is smaller than the threshold value p , and returns to the Step SP 26, incrementing the parameter t for the analytical point if the value of the power information, Power (t), is above the threshold value power (Steps SP 26 -SP 28). On the other hand, in case the value of the power information, Power (t), is smaller than the threshold value p , the CPU 1 places a mark for the beginning point of an invalid segment at the analytical point and then returns to the Step SP 22 mentioned above (Step SP 29).

The CPU 1 performs the above-mentioned process until it detects the completion of the process at all of the analytical points at the Steps, SP 22 or SP 24, and it shifts to its processing of the subsequent steps beginning with the Step 30 after it has established the division of the segments between the effective segments above the threshold value p and the invalid segments below the threshold value p through its comparison of the power information, Power (t), and the threshold value p at all the analytical points.

In the process subsequent to this, the CPU 1 clears the parameter t for the analytical point to zero and begins the subsequent process as from the initial analytical point (Step SP 30). The CPU 1 judges whether the analytical point is one marked as the beginning of an effective segment (Steps SP 31 and SP 32) after it ascertains that the analytical point data requiring its processing has not yet been completed. In case the analytical point is not one in which an effective segment begins, the CPU 1 increments the parameter t for the analytical point and then returns to the Step SP 29 mentioned above (Step SP 33).

On the other hand, in case the CPU 1 has detected any analytical point where an effective segment begins, it ascertains again that there is no analytical point remaining to be processed and further judges whether the analytical point is one in which an invalid segment begins (Steps SP 34 and SP 35). In case the analytical point is not one in which an invalid segment begins, which means that it is an analytical point within an effective segment, the CPU 1 finds the function for the variation $d(t)$ of the power information, Power (t), (which is to be called a rise extraction function in the following part since it is to be used for the extraction of a rise in the power information in the subsequent process) by performing arithmetic operations according to the equation (1) (Step SP 36)

$$d(t) = \{\text{power}(t+k) - \text{power}(t)\} / \{\text{power}(t+k) + \text{power}(t)\} \quad (1)$$

where k represents a natural number appropriate for capturing the fluctuations in power.

Thereafter, the CPU 1 judges whether or not the value of the rise extraction function $d(t)$ so obtained is smaller than the threshold value d , and, if it is smaller, the CPU 1 increments the parameter t for the analytical point and returns to the Step SP 34 (Steps SP 37 and SP 38). On the other hand, in case the rise extraction function $d(t)$ is found to be in excess of the threshold value d , the CPU 1 places the mark for the beginning of a new effective segment to the analytical point (Step SP 39). With this, the effective segment has been divided into smaller parts.

Thereafter, the CPU 1 ascertains that the processing has not yet been completed on all the analytical points and then judges whether or not a mark for the beginning of an invalid segment is placed on the analytical point where the processing is being performed, and, in case any such mark is placed there, the CPU returns to the above-mentioned step, SP 31, and performs the detecting process for the beginning point of the next effective segment (Steps SP 40 and SP 41).

On the other hand, when the point is not an analytical point for the beginning of an invalid segment, the CPU 1 obtains the rise extraction function $d(t)$ by the equation (1) on the basis of the power information, Power (t) and judges whether or not the rise extraction function $d(t)$ is smaller than the threshold value d (Steps SP 42 and SP 43). If the function is any smaller, the CPU 1 returns to the above-mentioned step, SP 34, and proceeds to the processing of extraction of a point of change in the rise of the power information. In the meantime, if the rise extraction function $d(t)$ at the analytical point is continuously above the threshold value at the step SP 43, the CPU 1 returns to the step SP 40 to increment the parameter t for the analytical point and to judge whether or not the rise extraction function $d(t)$ in respect of the next analytical point has become smaller than the threshold value d .

When the CPU 1 has detected by repeating the above-mentioned process at Steps SP 31, SP 34 or SP 40 that the process has been completed on all the analytical points, the CPU 1 proceeds to the process for reviewing the segments on the basis of the segment length at the step SP 45 and the subsequent steps.

In this process, the CPU 1 clears the parameter t for the analytical point to zero and thereafter ascertains that the analytical point data has not yet been completed, and then judges whether or not any mark for the beginning of a segment is placed on the particular analytical point, regardless of its being an effective segment or an invalid segment (Steps SP 45 - 47). In case the point is not a beginning point of a segment, the CPU 1 returns to the step SP 46 in order to increment the parameter t for the analytical point and to move on to the data at the next analytical point (Step SP 48). In case the CPU 1 has detected any beginning point for a segment, the CPU 1 sets the segment length parameter L at the initial value "1" in order to calculate the length of the segment starting from this beginning point (Step SP 49).

Thereafter, the CPU 1 increments the analytical point parameter t and, ascertaining that the analytical point data has not yet been completed, further judges whether or not any mark for the beginning of a segment, regardless of an effective one or an invalid one, is placed on the particular analytical point (Steps SP 50 - SP 52). If the CPU 1 finds as the result that the analytical point is not a point where a segment begins, the CPU 1 increments the segment length parameter L and also increments the analytical point parameter t , thereafter returning to the above-mentioned step, SP 51 (Steps SP 53 and SP 54).

By repeating the process consisting of the steps SP 51 to SP 54, the CPU 1 will soon come to an analytical point where a mark for the beginning of a segment is placed, obtaining an affirmative result at the step SP 52. The segment length parameter found at this time corresponds to the distance between the marked analytical point for processing and the immediately preceding marked analytical point for processing, i.e. to the length of the segment. If an affirmative result is obtained at the step SP 52, the CPU 1 judges whether or not the parameter L (i.e. the segment length) is shorter than the threshold value m , and, when it is above the threshold value m , the CPU 1 returns to the above-mentioned step, SP 46 without eliminating the mark for the beginning of a segment, but, when it is smaller than the threshold value m , the CPU 1 removes the mark placed at the front side to indicate the beginning of a segment, thereby connecting this

segment to the preceding segment, and then returns to the above-mentioned step SP 46 (Steps SP 55 and SP 56).

Moreover, in case the CPU 1 has returned to the step SP 46 from the step SP 55 or SP 56, the CPU 1 will immediately obtain an affirmative result at the step SP 47, unless the analytical point data has been completed, and will proceed to the processing at the subsequent steps beginning with the step SP 49 and will move on to the operation for searching for another mark next to the mark just found, and the CPU finds the next mark in the same manner as described above, then carrying out the review of its segment length.

By repeating a processing operation like this, the CPU 1 will complete the review of all the segment lengths, and when it obtains an affirmative result at the step SP 46, the CPU 1 will complete the processing program.

Fig. 6 presents one example of segmentation by a process in the manner just described. In the case of this example, the repetition of the processes in the steps up to SP 29 will establish the distinction between the effective segments, S1 - S8, and the invalid segments, S11 - S18, on the basis of the power information, Power (t). Thereafter, by the repetition of the processes up to the step SP 44, the effective segment S4 will be further divided into smaller segments, S41 and S42, at the point of change in the rise of power on the basis of the rise extraction function d(t). Furthermore, the processing at the step SP 45 and the subsequent steps will thereafter be performed, and then a review will be made on the basis of the segment length. In this example, however, no connection of segments in particular will take place since there is no segment shorter than the prescribed length.

Therefore, with the embodiments described above, the system will be capable of performing a highly accurate segmentation process not liable to any faulty segmentation due to noises or power fluctuations for the reason that the power information divides the acoustic signals between the effective segments above the threshold value and the invalid segments below the value, and that the effective segments are further divided into smaller segments by the point of change in the rise of the power information, and that the segments so established are reviewed on the basis of the segment length.

In other words, this process can also eliminate the use of the unstable period with little vocal power in the subsequent processes such as the identification of the musical interval because the sections containing power information in excess of the threshold value are taken as effective segments. Moreover, as the system has been designed to divide a segment into smaller parts by extracting a point of change in the rise of power, it is possible to have the system perform segmentation well even in case where there occurs a transition to the next sound while the power is maintained above the prescribed level. Moreover, as the system is designed to conduct a review on the basis of the segment length, it is possible to avoid dividing one sound or a rest period into a plural number of segments.

In the example given above, moreover, the length of the effective sections mentioned above, including the further divided effective sections mentioned above, and that of the invalid sections mentioned above, have been extracted, this, however, is not necessarily required. In such a case, a beginning mark and an ending mark are to be placed respectively in the beginning and end of each section above the threshold value at the step SP 66 as shown in the block diagram representing the processing procedure given in Fig. 7. In specific terms, it is seen with reference to the flow chart in Fig. 8, which represents greater details of what is shown in Fig. 7, that the CPU 1 returns to the above-mentioned step, SP 22, after putting a mark of a segment ending point at the analytical point concerned in case the value of the power information, Power (t), becomes smaller than the threshold value power (Step SP 29'). With this embodiment, the system will finish the program when it detects the completion of the processing in respect of all the analytical points at the steps, SP 31, SP 34, or SP 40, by repeating the processes mentioned above. The segments processed at this time are the same as those shown in Fig. 6.

Furthermore, it is possible to perform the segmentation process also by the procedure illustrated in the flow chart in Fig. 9. In this case, the procedure from the beginning to the step SP 28 is identical to the same steps shown in Fig. 8. The CPU 1 will soon detect an analytical point having the power information, Power (t), smaller than the threshold value p by repeating the processing at the steps, SP 26 to SP 28, in the same way as what is shown in Fig. 8, and will obtain an affirmative result at the step SP 27. At this time, the CPU 1 places a mark for the ending of the segment at this analytical point and thereafter detects the length L of the segment on the basis of the beginning mark information for the above-mentioned segment and the ending mark information for the segment, and judges whether or not the length L is smaller than the threshold value m (Steps SP 68 - SP 70). Such a judging step is one designed not to regard too short a segment as an effective one, and the threshold value m has been decided in relationship to musical notes. When the CPU 1 increments the parameter t and returns to the above-mentioned step SP 22 after it eliminates the beginning and the ending marks for the segment if it obtains an affirmative result at this step SP 70. On the other hand, when it obtains a negative result because the length of the segment is sufficient,

it immediately increments the parameter t , without eliminating those marks, and returns to the above-mentioned step SP 21 (Steps SP 71 and SP 72).

By repeating this processing procedure, the CPU 1 completes its processing with respect to all the power information and, with an affirmative result obtained at the step SP 23 or SP 26, it completes the particular program.

Fig. 10 presents the chronological change of power information and an example of the results of segmentation corresponding to this chronological change. In the case of this example, the segments, S1, S2 ... SN, are obtained by the execution of the process given in Fig. 9. Moreover, in the period for the points in time, t_1 - t_2 , the power information is in excess of the threshold value p , but the period is short and its length is below the threshold value m , it is not extracted as a segment.

Furthermore, also the segmentation processing procedure as presented in the following can be applied. This procedure is explained with reference to the flow chart shown in Fig. 11.

The CPU 1 first clears the parameter t for the analytical point to zero and then, ascertaining that the data to be processed is not yet completed, (Steps SP 80 and SP 81), and performs arithmetic operations with respect to that analytical point t on the basis of the power information (t) for that analytical point t and the rise extraction function $d(t)$ (Step SP 82).

Here, k is to be set an appropriate time difference suitable for capturing the change in the power information.

Thereafter, the CPU 1 judges whether or not the rise extraction function $d(t)$ at the analytical point t is above the threshold value d and, if it obtains a negative result because the function is smaller than the threshold value d , it increments the parameter t and returns to the above-mentioned step SP 81 (Steps SP 83 and SP 84).

By repeating this processing procedure, the CPU 1 soon finds an analytical value immediately after its rise extraction function $d(t)$ has changed to a level above the threshold value d , and obtains an affirmative result at the step SP 83. At this time, the CPU 1 ascertains, after it places a segment beginning mark to that analytic point, that the data on the analytical point to be processed has not yet been completed, and then the CPU 1 performs arithmetic operations to find the rise extraction function $d(t)$ of the power information again with respect to that analytical point on the basis of the power information Power (t) on that analytical point and the power information Power ($t+k$) for the analytical point $t+k$, which is ahead of that analytical point by k -segments (Steps SP 85 and SP 87).

Thereafter, the CPU 1 judges whether or not the rise extraction function $d(t)$ at that analytical point t is smaller than the threshold value d , and, if it obtains a negative result because the function is above the threshold value d , it increments the parameter t and returns to the above-mentioned step SP 86 (steps SP 88 - SP 89). In contrast to this, if the CPU 1 obtains an affirmative result because the function is smaller than the threshold value d , it returns to the above-mentioned step SP 81 and then proceeds to its processing operation for extracting a point of change immediately following a change of the rise extraction function $d(t)$ to a level above the threshold value d .

By repeating a processing procedure in this manner, the CPU 1 places a segment beginning mark to every point of change of the rise in the power information, and will soon complete its processing of all the power information, obtaining an affirmative result at the step SP 81 or SP 86 and thereupon finishing the particular program.

Moreover, the system is designed to execute the segmentation process through its extraction of the rise in power information in this way in view of the fact, for example, that a singer will raise the power to the highest level at the point of the onset of a new sound when he or she changes the pitch of sounds, letting the voice have a gradual decrement in power thereafter. It also reflects the consideration of the fact that musical instrument sounds have such nature that an attack occurs in the beginning of a sound with a decay occurring thereafter.

Fig. 12 represents one example of the chronological change of the power information Power (t) and the chronological change of the rise extraction function $d(t)$, and, in the case of this example, the execution of the processing operation shown in Fig. 11 will result in the division of the signals into the segments, S1, S2

Furthermore, a segmentation review process as shown in Fig. 13 and Fig. 14 may be performed.

Another arrangement of the segmentation process on the basis of the power information may be employed, as described below.

Fig. 13 presents a flow chart illustrating this process at the functional level while Fig. 14 is a flow chart illustrating greater details of what is shown in Fig. 13. First, the CPU 1 performs arithmetic operations to find the function of variation for the power information with respect to each analytical point, extracts a rise in the power information on the basis of the function, and places a segment beginning mark at the analytical point for the rise (Steps SP 90 and SP 91).

Moreover, the system has been designed to perform segmentation by extracting a rise in the power information in view of the fact that acoustic signals are of such nature that they will attain the maximum power at the beginning point of a new sound, when their musical interval has been changed, with a gradual decrement of power occurring thereafter.

After that, the CPU 1 measures the length from the beginning point of a segment to that of the next segment, i.e. the segment length, and eliminate a segment having any insufficient segment length, connecting the section to another segment before or after it (Steps SP 92 and SP 93).

The system has been designed not to treat a segment as such in case its length is too short because acoustic signals may sometimes have fluctuations in their power information and may also have intrusive noises in them and additionally because it is necessary to prevent segmentation errors from their occurrence in consequence of a plural number of peaks which may sometimes occur in the change of power in vocal sound even when the singer intends to utter a single sound.

Thus, this system is capable of executing its segmentation process based on the information on a rise in the power information and additionally taking account of the segment length.

Next, this process is explained in further detail on the basis of Fig. 14.

In Fig. 14, the steps from SP 80 to SP 89 are the same as those given in Fig. 11, and their explanation is omitted here. That is, the step SP 110 and the subsequent steps are taken for a review of the segments.

For processing a review of segments, the CPU 1 first clears the parameter t to zero and then ascertains that the analytical point data to be processed has not yet been completed, and it judges whether or not any mark for the beginning of a segment is placed in respect of the analytical point (Steps SP 110 - SP 112). When the CPU 1 obtains a negative result as no such mark is placed, it increments the parameter t and returns to the above-mentioned step SP 111 (Step SP 113). By repeating this process, the CPU 1 soon finds an analytical point with such a mark placed on it and obtains an affirmative result at the step SP 112.

At this time, the CPU 1 increments the parameter t , setting 1 as the length parameter L , and then, ascertaining that the analytical point data to be processed has not yet been completed, it judges whether or not a segment beginning mark is placed on the analytical point t (Steps SP 114 - 117). When the CPU 1 obtains a negative result as no such mark is placed on the analytical point being processed, the CPU 1 increments both the length parameter L and the analytical point parameter t , and returns to the above-mentioned step SP 116 (steps SP 118 and SP 119).

Repeating this process, the CPU 1 will soon find an analytical point to which a segment beginning mark is placed next to it and will obtain an affirmative result at the step SP 117. The length parameter L at this time corresponds to the distance between the analytical point which has a mark on it and is an object of processing and the marked analytical point immediately preceding it, i.e. the length of the segment. When an affirmative result is obtained at the step SP 117, the CPU 1 judges whether or not this parameter L (the segment length) is shorter than the threshold value m , and, in case the parameter is in excess of the threshold value m , the CPU 1 returns to the step SP 111 mentioned above without eliminating the segment beginning mark, but, if the parameter is smaller than the threshold value m , the CPU 1 eliminates the segment beginning mark at the front side, i.e. connects this segment to the segment at the front side, and returns to the above-mentioned step 111 (Steps SP 120 and SP 121).

Fig. 15 shows one example of the chronological change of the power information Power (t) and the chronological change of the rise extraction function $d(t)$, and, in this example, the acoustic signals are divided into the segments, $S_1, S_2 \dots S_N$ by their processing up to the step SP 89 shown in Fig. 14. However, by executing their processing as from the step SP 110, those segments short in length are excluded, with the result that the segment S_3 and the segment S_4 are combined into the single segment S_{34} .

In the above-mentioned embodiment, moreover, the function expressed in the equation (1) has been applied as the function for extracting the rise, but another function may be applied. For example, a differential function with a fixed denominator may be applied.

Furthermore, in the embodiment given above, a square sum of the acoustic signal is used as the power information, but another parameter may be used. For example, a square root for the square sum may be used.

Moreover, in the embodiment mentioned above, it is shown that a segment in an insufficient length is connected to the immediately preceding segment, but such a short segment may well be connected to the immediately following segment. Such a short segment may also be connected to the immediately preceding segment unless the immediately preceding segment is one other than a rest section, but to the immediately following segment if the immediately preceding segment is a rest section.

Segmentation Based on Pitch Information

Next, the segmentation process of the automatic music transcription system according to the present invention as based on the pitch information (Refer to the step SP 4 in Fig. 3) is explained in detail with reference to the flow charts presented in Fig. 16 and Fig. 17.

In this regard, Fig. 16 shows a flow chart illustrating such a process at the functional level, and Fig. 17 gives a flow chart showing greater details.

The CPU 1 calculates the length of a series with respect to all the sampling points in each analytical cycle on the basis of the obtained pitch information (Step SP 130). Here, the length of a series means a series of period RUN assuming the value of the pitch information in a prescribed narrow range R1 symmetrical in form centering around the pitch information on the observation point P1 as illustrated in Fig. 18. The acoustic signals generated by a singer or the like are generated with the intention of making such sounds as will assume a regular musical interval for each prescribed period, and, even though they may have fluctuations, it can be considered that, the changes in the pitch information for a period in which one and the same musical interval is intended should take place in a narrow range. Thus, the series length RUN will serve as a guide for capturing the period of the same sound.

Subsequently, the CPU 1 performs calculation to find a section in which sampling points with a series length in excess of the prescribed value appear in continuation (Step SP 131), thereby eliminating the influence due to the changes in the pitch information. After that, the CPU 1 extracts as a typical point a sampling point having the maximum series length in respect of each of the sections found by the calculation (Step SP 132).

Then, finally, when the difference in the pitch information (i.e. the difference of tonal height) at two adjacent typical points is in excess of the prescribed level, the CPU 1 finds the amount of the variation in the pitch information between the typical points with respect to the individual sampling points between them and segments the acoustic signals at the sampling point where the amount of such variation is in the maximum (Step SP 133).

In this manner, this system is capable of performing the segmentation process on the basis of the pitch information without being influenced by fluctuations in the acoustic signals or by sudden outside sounds.

Next, this process is explained in greater detail on the basis of Fig. 17.

First, the CPU 1 works out the length of the series run(t) by calculation with respect to all the sampling points t ($t = 0$ to N) in every analytical cycle (Step SP 140).

Next, after clearing to zero the parameter t indicating the sampling point to be processed, the CPU 1 ascertains that the processing has not yet been completed in respect of all the sampling points and judges whether or not the series length run(t) at the sampling point t, which is the object of the processing, is smaller than the threshold value r (Steps SP 141 to 143). If the CPU judges as the result of this operation that the length of the series is insufficient, it increments the parameter t and returns to the above-mentioned step SP 142 (Step SP 144).

By repeating this process, the CPU 1 will soon take up a sampling point with a series length run(t) longer than the threshold value r as the object of processing and obtains a negative result at the step SP 143. At this time, the CPU 1 stores that parameter t as the parameter s and marks it as the beginning point where the series length run(t) has exceeded the threshold value r, thereafter ascertaining that the processing has not yet been completed with respect to all the sampling points and judging whether or not the series length run(t) at the sampling point t taken as the object of the processing is smaller than the threshold value r (Steps SP 145 to SP 147). If the CPU 1 finds as the result of this operation that the series length run(t) is sufficient, it increments the parameter t and returns to the above-mentioned step SP 146 (Step SP 148).

By repeating this processing operation, the CPU 1 soon finds a sampling point where the series length run(t) is shorter than the threshold value r as the object of its processing and obtains an affirmative result at the step SP 147. Thus, the CPU 1 detects those sections in continuum where the series length run(t) is shorter than the threshold value r, i.e. the section from the marked point s to the sampling point t-1 at one point ahead, and the CPU 1 puts a mark as a typical point to the point which gives the maximum series length among these sampling points (Step SP 149). Moreover, upon completion of this process, the CPU 1 returns to the above-mentioned step SP 142 and performs the detecting process for the next continuous section where the series length run(t) is in excess of the threshold value r.

When the CPU 1 has completed the detection of the continuous section where the series length run(t) is in excess of the threshold value r and the marking of the typical points, with the processing of all the sampling points completed in this way, the CPU 1 clears the parameter t to zero again, thereafter ascertaining that the processing has not yet been completed in respect of all the sampling points and

judging whether or not the mark as a typical point is placed on the sampling point taken as the object of the processing (Steps SP 150 to SP 152). In case no such mark is placed, the CPU 1 increments the parameter t and returns to the above-mentioned step SP 151 (Step SP 153).

By repeating this process, a sampling point with a mark placed on it will be taken up as the object of processing, and the first typical point will be found. Then, the CPU 1 stores and marks this value t as the parameter s, and, further incrementing the parameter t and ascertaining that the processing has not yet been completed with respect to all the sampling points, the CPU 1 judges whether or not a mark as a typical point is placed on the sampling point taken as the object of the processing (Step SP 154 to 157). In case no such mark is placed there, the CPU 1 increments the parameter t and returns to the above-mentioned step SP 154 (Step SP 158).

As this process is repeated, a sampling point with a mark placed on it will soon be taken up as the object of the processing, and the next typical point t will be found. At this time, the CPU 1 judges whether or not the difference in pitch information between these mutually adjacent typical points s and t is smaller than the threshold value q, and, in case it is smaller, the CPU 1 returns to the above-mentioned step SP 154, proceeding to the process for finding the next pair of adjacent typical points, but, in case the difference is in excess of the threshold value q, the CPU 1 finds the amount of variation in the pitch information between the typical points in respect of the individual sampling points s to t between them and places a segment mark on the sampling point with the maximum amount of variation (Steps SP 159 to 161).

By the repetition of this process, segment marks are placed one after another between typical points, and an affirmative result is soon obtained at the step SP 156, the process being thereupon completed.

Accordingly, the above-mentioned embodiment is capable of performing the segmentation process well even if there are fluctuations in the acoustic signals or if sudden outside sounds are included in them since the system performs its segmentation process by the use of a series length representing a length in which the pitch information is present in a narrow range.

In the embodiment mentioned above, moreover, the system processes for segmentation the pitch information obtained by autocorrelation analysis. Yet, it goes without saying that the method of extracting the pitch information is not confined to this.

Processing for Review of Segmentation

Next, with reference to the flow chart in Fig. 19, a detailed description is presented with regard to the processing for the review of segmentation in the operation of the automatic music transcription system according to the present invention (Refer to the step SP 6 in Fig. 3).

Now, this reviewing process has been adopted in order to improve the accuracy of the musical interval identifying process through application of further segmentation of the segments prior to the process for identifying a musical interval and by executing the musical interval identifying process with those segments because the musical interval identified is highly likely to be erroneous, resulting in a decline in the accuracy of the generated musical score data, in case any segment has been established by mistake in such a manner as to consist of two or more sounds. In this case, it is conceivable that a single sound may be divided into two or more segments, this process will not present any problem because those segments which are considered to form a single sound on the basis of the identified musical scale and the power information are connected to each other by the segmentation processing at the step SP 11. In such a reviewing process for segmentation, the CPU 1 first ascertains that the segment to be taken up for processing is not the final segment and then execute the matching of the particular segment with the entire segmentation result (Steps SP 170 and SP 171).

Here, matching means a process which finds the grand total sum of the absolute values of the differences between the value of one part of the particular segment length as divided by its integral number or the value obtained by multiplying the segment length by its integral number and the length of the length of the other segment and the frequency of the disagreement between the value for one part of the length of the segment as divided by its integral number or the value obtained by multiplying it with its integral number and the value for the length of the other segment (i.e. the number of times of mismatches). Moreover, in the case of this embodiment, the other segment to be taken as the partner for the matching will be both of the segment obtained on the basis of the pitch information and the segment obtained on the basis of the power information.

For example, in case the first segment S1 is the object of the processing out of the ten segments which are as shown in Fig. 20 and have been established by the former-stage process of segmentation (Steps SP 4 and SP 5 in Fig. 3), this matching process generates "1 + 3 + 1 + 1 + 5 + 0 + 0 + 1 + 9 = 21" as the grand total sum information on the differences and seven times as the number of times of mismatching.

When the number of times of mismatching and the degree of such mismatching (i.e. the information on the grand total sum of the differences) have thus been obtained for the object of the processing, the CPU 1 stores the information in the auxiliary memory device 6 and then returns to the above-mentioned step, SP 170, taking up the next segment as the segment to be the object of the processing (Step SP 172).

5 The repetition of the processing loop composed of these steps SP 170 to SP 172 generates information on the number of times of mismatching and the degree of the mismatches with respect to all the segments, and soon an affirmative result is obtained at the step SP 170. At this time, the CPU 1 determines the standard length on the basis of the segment length which is liable to the minimum of these factors in light of the information stored on all the number of times of mismatching and the degree of such mismatches in
10 the auxiliary memory device (Step SP 173). Here, the standard length means the duration of time equivalent to a quarter note or the like.

In the case of the example in Fig. 20, "60" is extracted as the segment length with the minimum of the number of times of mismatching and the minimum of its degree, and "120," i.e. the value two times as large as this length "60," is selected as the standard length. In practice, the length which the time for a
15 quarter note can take corresponds to the value within the prescribed range, and, from this viewpoint, "120" instead of "60" is extracted as the standard length.

When the standard length is extracted, the CPU 1 further divides the segments generally longer than the standard length by a value roughly corresponding to one half of the standard length, completing the reviewing process for this segmentation Step(SP 174). In the case of the example given in Fig. 20, the fifth
20 segment S5 is further divided into "61" and "60"; the sixth segment S6 is further divided into "63" and "62"; the ninth segment S9 is further divided into "60" and "59"; the tenth segment S10 is further divided into "58", "58", "58," and "57".

Therefore, according to the embodiment given above, it is possible to make further division of segments even in case two or more sounds have been segmented as a single segment. Hence, it is possible for the
25 system accurately to execute such processes as the musical interval identifying process and the musical interval correcting process.

As regards this manner of further segmentation, it will not happen that any segments corresponding to a single sound erroneously divided into two or more sections ever remain as they are since the system provides for a post-treatment process for connecting to each other the segments considered to form a
30 single sound.

Moreover, the embodiment given above showed the extraction of the standard length on the basis of the number of times of mismatching and the degree of mismatching, but the extraction of the length may be done also on the basis of the frequency of occurrence of a segment length.

Furthermore, the embodiment given above showed a case in which a duration of time equivalent to a
35 quarter note is used as the standard length, but a duration of time equivalent to an eighth note may be employed as the standard length. In this case, further segmentation will be performed not by a length equivalent to one half of the standard length, but by the standard length itself.

Furthermore, the embodiment given above showed a case in which the present invention is applied to a processing system which has both the segmentation based on the pitch information and that based on the
40 power information, and yet the present invention may be applied to an automatic music transcription system which has at least the segmentation process based on the power information.

Identification of Musical Interval

45 Next, a detailed description is given with reference to the flow chart in Fig. 21 about the musical interval identifying process (step SP 7 in Fig. 3) for an automatic music transcription system like this.

The CPU 1 first ascertains that the processing of the final segment has not yet been completed, and then sets the pitch information (x_0) for the lowest interval that the acoustic signals are considered to take on the axis of an absolute musical interval as the musical interval parameter x_j ($j = 0$ to $m - 1$, where m
50 expresses the number of musical intervals which the acoustic signal is considered to take on the axis of the absolute musical interval in the high tone range) and finds by calculation and stores the distance e_j of the pitch information p_i ($i = 0$ to $n - 1$, where n expresses the number of items of the pitch information for this segment) in relation to that musical interval (Steps SP 180 to SP 182).

Here, the distance e_j is defined by the sum of the square of the difference $p_i - x_j$ (Refer to Fig. 22)
55 between each item of the pitch information p_i in the segment taken as the object of the calculation of the distance and the pitch information x_j for the musical interval on the axis of the absolute musical interval, as expressed in the following equation:

$$\epsilon_j = \sum_i (p_i - x_j)^2 \quad \dots \quad (2)$$

Thereafter, the CPU 1 judges whether or not the musical interval parameter x_j has become the pitch information x_{m-1} for the musical interval on the axis of the highest absolute musical interval that the acoustic signal is considered to be able to take, and, if it obtains a negative result, it renews the musical interval x_j to develop the pitch information x_{j+1} for the musical interval higher by a half step on the axis of the absolute musical interval than the musical interval used for the processing until the present time, then returning to the above-mentioned distance-calculating step, SP 182 (Steps SP 183 and SP 184).

By the repetition of the processing loop consisting of these steps, SP 183 and SP 184, the distance ϵ_0 to ϵ_{m-1} between the pitch information and all the musical intervals on the axis of the absolute musical scale is found by calculation, and an affirmative result is found soon at the step SP 183. At this time, the CPU 1 detects the smallest of the distances regarding the individual musical intervals stored in the memory and decides this musical interval where the distance is in the minimum as the musical interval of the segment, and then sets the segment to be processed at the next segment, thereafter returning to the step SP 180 mentioned above (Steps SP 185 and SP 186).

By the repetition of the process in this manner, the musical intervals are identified for all the segments, and an affirmative result is obtained at the step SP 180, the CPU 1 thereupon bringing the particular processing program to a finish.

Therefore, the embodiment described above can identify the musical interval with a high degree of accuracy owing to its calculation of the distance between the pitch information on each segment and the axis of the absolute musical interval and its identification of the musical interval of the segment with such a musical interval on the axis of the absolute musical interval as results in the minimum distance.

Moreover, in the embodiment given above, the distance is calculated by the equation (2), but it is also acceptable to work out the distance by the following equation:

$$\epsilon_j = \sum_i |p_i - x_j| \quad \dots \quad (3)$$

Furthermore, the pitch information used in the process for identifying the musical interval may be expressed either in Hz, which is the unit of frequency, or in cent, which is a unit frequently used in the field of music.

Next, a detailed description is presented with reference to the flow chart in Fig. 23 about another process for the identification of musical intervals with the automatic music transcription system according to the present invention.

The CPU 1 first takes out the initial segment out of the segments obtained by the segmentation process and then finds by calculation the average value of all the pitch information present in that segment (Steps SP 190 and SP 191).

After that, the CPU 1 identifies the musical interval found on the axis of the absolute musical interval and closest to the calculated average value as the musical interval for the particular segment (Step SP 192). Moreover, the musical interval of each segment of the acoustic signal is identified with either one of the musical intervals different by a half step on the axis of the absolute musical interval. The CPU 1 distinguishes whether or not a given segment processed in this way, with its musical segment thereby identified, is the final segment (Step SP 193). If the CPU 1 finds as the result of this operation that the processing has been completed, it finishes the program for the particular program, but, if the process has not been completed yet, the CPU 1 takes up the next segment as the object of its processing and returns to the above-mentioned step SP 191 (Step SP 194).

With the repetition of this processing loop consisting of these steps, SP 191 to SP 194, the identification of musical intervals is executed with respect to all the segments on the basis of the pitch information in the segment.

In this regard, the system has been designed to utilize the average value for the musical interval identifying process on the ground that the acoustic signals will fluctuate in such a manner as to center around the musical interval intended by the singer or the like, even though those signals may have fluctuations, and that the average value corresponds to the intended musical interval.

Fig. 24 shows one example of the identification of a musical interval through such processing. The curve PIT in a dotted line represents the pitch information of the acoustic signal while the solid line VR in the vertical direction shows the division of each segment. The average value for each segment in this example is indicated by the solid line HR in the horizontal direction, and the identified musical interval is

represented by the dotted line HP in the horizontal direction. As it is evident from this Fig. 24, the average value has a very small deviation in relation to the musical interval on the axis of the absolute musical interval, and this makes it possible to perform the identification of the musical interval well.

Consequently, this embodiment finds the average value of the pitch information in respect of each segment and identifies the musical interval of the segment with such a musical interval on the axis of the absolute musical interval as is closest to the average value. Therefore, the system is capable of identifying the musical intervals with a high degree of accuracy. Moreover, as this system performs a tuning process on the acoustic signals prior to the identification of the musical interval, this method can find an average value assuming a value close to the musical interval on the axis of the absolute musical interval, providing considerable ease in the performance of the identification process.

In the example presented above, the musical interval of the segment is identified on the basis of the average value of the pitch, but the identification of segments is not limited to this. It can be based on the median value for the pitch. In other words, the process is performed as described below with respect to a flowchart shown in Fig. 25.

As shown in Fig. 25, the CPU 1 first takes out the initial segment out of the segments obtained by segmentation and then extracts the median value of all the pitch information present in the segment (Steps SP 190 and SP 195). Here, the median value is the value of the pitch information in the middle when the items of the pitch information for the particular segment are arranged in the order starting with the largest one, provided that the number of such items is an odd number, and the average value of the two items of such information positioned in the middle in case the number of such items is an even number.

The processes other than those at the steps SP 195, SP 196, and SP 196 are basically the same as those shown in Fig. 23.

By the repetition of the processing loop consisting of the steps, SP 195, SP 196, SP 193, and SP 194, the identification of the musical intervals on the basis of the pitch information in the particular segment is performed with respect to all the segments.

Here, the reason for which the system has been designed to utilize the median value for the process for identifying the musical intervals is that, even though acoustic signals have fluctuations, they are considered to fluctuate in a manner centering around the musical interval intended by the singer or the like, so that the median value corresponds to the intended musical interval.

Fig. 26 shows one example of the identification of musical intervals by this process, and the dotted-line curve PIT shows the pitch information of the acoustic signal while the solid line VR in the vertical direction indicates the division of the segment. The median value for each segment in this example is represented by the solid line HR in the horizontal direction, and the identified musical interval is shown by the dotted line HP in the horizontal direction. As it is evident from this Fig. 26, the median value has a very small deviation in relation to the musical interval on the axis of the absolute musical interval, making it possible for the system to perform the identifying process well. Also, it is possible to identify the musical interval without being affected by any unstable state of the pitch information immediately before or after the division of a segment (for example, the curve portions C1 and C2).

Thus, since the system in this embodiment extracts the median value of the pitch information on each segment and identifies the musical interval at such a musical interval on the axis of the absolute musical interval as is positioned closest to the median value, it is possible for the system to identify the musical interval with a high degree of accuracy. Moreover, prior to the identification of the musical interval, this system applies a tuning process to the acoustic signals. Therefore, by this method, the median value assumes a value close to the musical interval on the axis of the absolute musical interval, so that it has made it considerably easy to perform the identification.

Furthermore, the process for the identification of the musical interval may be executed on the basis of a peak point in the rise of power (Step SP 7 in Fig. 3). An explanation is provided on this feature with reference to Fig. 27 and Fig. 28. The processing procedure illustrated in Fig. 27 is basically the same as that given in Fig. 23, and only the steps, SP 197 and SP 198, are different.

The CPU 1 first takes out the initial segment out of those segments which have been obtained by segmentation and then takes out the sampling point which gives the initial maximum value (a peak in the rise) from the change in the power information on the segment (Steps SP 190 and SP 197).

After that, the CPU 1 identifies, as the musical interval for the particular segment, such a musical interval on the axis of the absolute musical interval as is closest to the pitch information on the sampling point giving rise to the peak in the rise of power (Step SP 198). In this regard, the musical intervals of the individual segments of the acoustic signals are identified with either one of the musical intervals different by a half step on the axis of the absolute musical interval.

Here, it has been designed to use the peak in the rise of the power information for the process for identifying the musical intervals because it is considered that, even though acoustic signals have fluctuations, the singer or the like will control the volume of voice in such a way as to attain the musical interval at a peak in volume, increasing the volume of voice at the time when the musical interval is shifted to a new sound. As a matter of fact, it has been conclusively verified that there is a very close correlation between a peak in the rise of the power information and the musical interval.

Fig. 28 illustrates one example of the identification of the musical interval by this process, and the first dotted-line curve PIT represents the pitch information of the acoustic signal, the second dotted-line curve POW represents the power information, and the solid line VR in the vertical direction indicates the division of segments. The pitch information at the peak in the rise in each segment in this example is shown by the solid line HR in the horizontal direction while the identified musical interval is shown by the dotted line HP in the horizontal direction. As it is evident from this Fig. 28, the pitch information in relation to the peak point in the rise of the power information has a very small deviation from the musical interval on the axis of the absolute musical interval, and it is observed that this feature makes it possible for the system to identify the musical interval well.

Therefore, according to the embodiment described above, the system extracts the pitch information on the peak point in the rise of the power information for each segment and identifies the musical interval of the segment with such a musical interval on the axis of the musical interval as is closest to this pitch information. Hence, the system is capable of identifying the musical interval with a high degree of accuracy. Moreover, prior to the identification of the musical interval, the system applies a tuning process to the acoustic signals, so that the pitch information in relation to the peak point in the rise of the power information assumes a value close to the musical interval on the axis of the absolute musical interval, and therefore it has become very easy for this system to perform the identification.

Moreover, since the system makes use of the peak point in the rise of the power information, it is possible for the system to identify the musical interval well even if the segment is so short that the number of sampling points is small in comparison with the case of the identification of a musical interval through the statistical processing of the pitch information in the segment, with the result that the identification of the musical interval by this system is little liable to be influenced by the segment length.

Furthermore, the embodiment described above shows a process for identifying the musical interval on the basis of the pitch information in relation to the peak point in the power information, however, it is also a workable process to perform the identification of the musical interval on the basis of the pitch information on the sampling point which gives the maximum value of the power information on this segment.

Next, a detailed description is given with reference to the flow chart in Fig. 29 concerning a still another arrangement of the musical interval identifying process and the reviewing process for the once identified musical intervals performed by this automatic music transcription system according to the present invention.

The CPU 1 first obtains an average value, for example, of the pitch information of the particular segment, with regard to the segment obtained through segmentation, and then identifies the musical interval of a given segment with such one of the musical intervals different from one another by a half step on the axis of the absolute musical interval as is closest to the average value (Step SP 200).

The musical interval thus identified is reviewed by this system in the following manner. Here, the review is made of those segments which are considered to have been identified with a musical interval independently of the segments respectively preceding and following the segments under review as the result of their division as separate segments in consequence of the instability of their musical interval at the time of their sound transition.

The CPU 1 first ascertains that the processing of the final segment has not been completed yet and judges whether or not the length of the segment to be taken as the object of the processing is shorter than the threshold value, and, in case the length exceeds the threshold value, the CPU 1 shifts the processing operation onto the next segment to take it up as the object of the processing, and then it returns to the step SP 200 (Steps SP 201 and SP 202).

The reason for this manner of processing is to be found in the fact that the length of a segment will be short in case it is identified as a separate segment despite its being a part of a single sound as at the beginning time or the ending time in the course of transition of the sound. When it is detected that the segment being processed is one with a short length, the CPU 1 determines the matching of the tendency of the change in the pitch information for the particular segment and the tendency of the change in the overshoot and also determines the matching of the tendency of the change in the pitch information for that segment and the tendency of the change in the undershoot, thereby judging whether or not the tendency of the change in the pitch information on that segment represents an overshoot or an undershoot (Steps SP

203 and SP 204).

Here, it is noted, at the time of a transition from one sound to another, that a gradual transition occurs in some cases from a somewhat higher musical interval level to the that of the sound in the proximity of the beginning of the next sound, that a gradual transition sometimes occurs from a somewhat lower musical interval level to that of the sound in the proximity of the beginning of the next sound, that a transition with a gradual decline in pitch sometimes occurs from the musical interval level of a sound to the next sound in the proximity of the ending of the sound, and that a transition with a gradual rise in pitch sometimes occurs from the musical interval level of a sound to the next sound in the proximity of the ending of the sound. Of the parts of segments where the musical interval changes with a tendency towards a gradual rise in pitch or a tendency towards a gradual fall in pitch by the effect of a sound transition although they are parts of single sounds, those parts which are higher in pitch than the proper musical interval are called "overshoots" and, of the parts of segments where the musical interval changes with a tendency towards a gradual rise in pitch or a tendency towards a gradual fall in pitch by the effect of a sound transition although they are parts of single sounds, those parts which are lower in pitch than the proper musical interval are called "undershoots".

Such overshoot parts and undershoot parts are sometimes distinguished as independent segments, and, in such a case, the CPU 1 judges whether or not the segment taken as the object of the process shows the possibility of its being a segment assuming any overshoot or any undershoot, the system determining the matching between the tendency of the change in the pitch information for the segment and the proper tendency towards a rise in pitch or the proper tendency towards a fall in pitch as just mentioned above.

When the CPU 1 obtains a negative result as the result of this judging process, it takes up the next segment as the object of the processing and returns to the above-mentioned step SP 201. On the other hand, if the CPU 1 judges that there is the possibility of the segment reflecting an overshoot or an undershoot, it finds the differences between the identified musical interval of the particular segment and the identified musical intervals of the immediately preceding segment and the immediately following segment in relation to the segment, placing a mark on the segment showing the smaller difference, and thereafter judges whether or not the difference in the musical interval of the segment so marked is smaller than the threshold value (Steps SP 205 and SP 206).

In case a sound has been divided into separate segments through the segmentation process even though they form a single sound, the musical interval of such a segment is not much different from the musical intervals of the preceding segments and the following segments, but, in case such a segment shows any considerable difference in musical interval from those of the segments preceding and following it, it is considered that the segment is not any segment reflecting any overshoot or any undershoot, in which case the CPU 1 takes up the next segment as the object of its processing and returns to the step SP 201 mentioned above.

On the other hand, in case the particular segment shows a small difference in musical interval from that of the marked segment, the CPU 1 judges whether or not there is any change in the power information in excess of the threshold value in the proximity of the boundary between the particular segment and the marked segment (Step SP 206). When a transition takes place from one sound to another, it often happens that also the power information changes, and, in case the change in the power information is large, it is considered that the particular segment is not any segment reflecting an overshoot or an undershoot. In this case, the CPU 1 takes up the next segment as the object of its processing and returns to the above-mentioned step, SP 201.

If an affirmative result is obtained by the judgment at this step, SP 207, it is considered that the particular segment is a segment reflecting an overshoot or an undershoot. Hence, the CPU 1 corrects the musical interval of the particular segment to that of the marked segment and taking up the next segment as the object of its processing, then returning to the step, SP 201, mentioned above (Step SP 208).

When the CPU 1 completes the review of the final segment by a process of a review of the musical intervals with respect to all the segments by the repetition of a process like this, it obtains an affirmative result at the step, SP 201, therewith completing the particular processing program.

Fig. 30 presents an example in which the identified musical interval is corrected by the process just described. Here, the curve expresses the pitch information PIT, and, in this example, the second segment S2 and the third segment S3 are intended to form the same musical interval. The second segment S2 was identified, prior to the correction, with the musical interval R2, which was at a level lower by a half step from the musical interval R3 with which the third segment S3 was identified, but the musical interval R3C of this segment S2 was later modified by this process to the musical interval R3 of the segment S3.

Therefore, this system can increase the accuracy of the musical score data owing to its improvement on the accuracy of the identified musical intervals and consequently to a higher degree of accuracy in the execution of the subsequent processes because the system has been designed thus to make a correction of the once identified musical interval through its detection of those segments erroneously identified with
 5 wrong musical intervals, using for the correction the segment length, the tendency of the change in the pitch information, the difference of the particular segment in musical interval from the preceding and following segments, and the difference of the particular segment in power information from the preceding and following segments.

Moreover, the above-mentioned embodiment has been designed to extract those segments identified
 10 with wrong musical intervals by taking account of the difference in power information between a particular segment and those sections preceding and following it, but it will be a workable method to extract such wrongly identified segments on the basis of at least the segment length, the tendency of the change in the pitch information, and the difference in musical interval between the particular segment and the preceding and following segments.

Moreover, it goes without saying that the method of detecting the presence of an overshoot or an undershoot on the basis of the change in the pitch information is not to be confined to the above-mentioned method of detecting them simply by a rising tendency or a falling tendency, but also another method, such as a comparison with a standard pattern, is applicable.

Also, as explained in the following part, the process for identifying musical intervals may be executed
 20 from a different viewpoint (Refer to the step SP 7 in Fig. 3). An explanation is given about this point with reference to Fig. 31 and Fig. 32.

The CPU 1 first takes out the first segment out of those obtained by segmentation, and then it prepares a histogram for all the pitch information in the particular segment (Steps SP 210 and SP 211).

Thereafter, the CPU 1 detects the value of the pitch information that occurs most frequently, i.e. the
 25 most frequent value, out of the histogram and identifies the musical interval of the particular segment with such a musical interval on the axis of the absolute musical interval as is closest to the detected most frequent value (Steps SP 212 and SP 213). Moreover, the musical interval of each segment of an acoustic signal is identified with either one of the musical intervals on the axis of the absolute musical interval with a difference by a half step between them. The CPU 1 then judges whether or not the segment identified with
 30 a musical interval by this process performed thereon is the final segment (Step SP 214). If it is found as the result that the process has been completed, the CPU 1 finishes the particular processing program and, if the process has not been completed yet, the CPU 1 takes up the next segment as the object of its processing and returns to the above-mentioned step, SP 211 (Step SP 215).

By repeating a processing loop consisting of these steps, SP 211 to SP 215, the identification of the
 35 musical interval is performed on the basis of the information on the most frequent value of the pitch information in each particular segment with respect to all the segments.

Here, the pitch information on the most frequent value is used in this system for its identification of the musical intervals in view of the fact that the pitch information showing the most frequent value can be considered to correspond to the intended musical interval because it is considered that the acoustic signals,
 40 which have fluctuations, fluctuate in a range centering around the musical interval intended by the singer or the like.

Moreover, in order to use the pitch information showing the most frequent value for the identification of the musical interval of sound segments, it is necessary to use a large number of sampling steps, and it is necessary to select a period for the acoustic signal for obtaining a piece of pitch information from the
 45 acoustic signal (the analytical cycle) to such an extent that the identification process will be performed well. Fig. 32 shows an example of the identification of musical intervals by a process like this, and the dotted-line curve PIT expresses the pitch information on the acoustic signal while the solid line VR in the vertical direction shows the division of the segment. The pitch information with the most frequent value for each segment in this example is represented by the solid line HP in the horizontal direction, and the identified
 50 musical interval is shown by the dotted line HP in the horizontal direction. As it is evident from Fig. 32, the pitch information with the most frequent value has a very minor deviation from the musical interval HR on the axis of the absolute musical interval and hence serves the purpose of performing the identifying process well. It is also understood clearly that this method is capable of identifying the musical intervals without being affected by the instability in the state of pitch information (for example, the curved sections C1 and
 55 C2) in the proximity of the segment division. Therefore, by the embodiment mentioned above, it is possible to determine the musical intervals with a high degree of accuracy because the most frequent value is extracted out of the pitch information on each segment and the musical interval of the segment is identified with such a musical interval on the axis of the absolute musical interval as is closest to the most frequent

value in the pitch information. Moreover, prior to the identification of the musical interval, a tuning process is applied to the acoustic signals, the pitch information with the most frequent value as processed by this method assumes the value closest to the musical interval on the axis of the absolute musical interval, making it very easy to perform the identifying process.

Also, it is possible to execute the process for the identification of the musical intervals by the processing procedure described below. Now, with regard to this process, an explanation is given with reference to Fig. 33 to Fig. 35.

The CPU 1 first takes the initial segment out of those segments obtained by the segmentation process (Step SP 6 in Fig. 3) and calculates the series length, $\text{run}(t)$, with respect to each analytical point in the segment (Steps SP 220 and SP 221).

Here, an explanation is given about the length of a series with reference to Fig. 34. The chronological change in the pitch information is presented in Fig. 34, in which the analytical points t are expressed along the horizontal axis while their pitch information is given on the vertical axis. As an example, the length of a series at the analytical point t_p is explained below.

The range of the analytical point which assumes the value between the pitch information h_0 and h_2 with a deviation by a very minor range Δh each upward or downward in relation to the pitch information on the particular analytical point t_p is the range from the analytical point t_0 to the analytical point t_s as shown in Fig. 34, and the period L from this analytical point t_0 to the analytical point t_s is to be referred to as the length of the series for the analytical point t_p .

When the length of the series, $\text{run}(t)$, is worked out by calculation in this manner with respect to all the analytical points in the segment, the CPU 1 extracts the analytical point where the length of the series, $\text{run}(t)$, is the longest (Step SP 222). Thereafter, the CPU 1 takes out the pitch information at the analytical point which gives the longest length of the series, $\text{run}(t)$, and identifies the musical interval of the particular segment with such a musical interval on the axis of the absolute musical interval as is the closest to this pitch information (Step SP 223). Moreover, the musical interval of each of the segments of acoustic signals is identified with either one of the musical intervals differing from one another by half a step on the axis of the absolute musical interval.

Next, the CPU 1 judges whether or not the segment identified with a musical interval as the result of this process performed on it is the final segment (Step SP 224). If the CPU 1 finds as the result of this operation that the process has been completed, it finishes the particular processing program and, if the process is not yet completed, it takes up the next segment as the object of its processing and returns to the above-mentioned step 221 (Step SP 225).

With the repetition of the processing loop consisting of the steps SP 221 to SP 225 in this manner, the CPU 1 executes the identification of the musical intervals on the basis of the pitch information on the analytical point which gives the length of the longest series in the segment with respect to all the segments.

In this regard, the system has been designed to utilize the length of the series, $\text{run}(t)$, for the process for identifying the musical intervals in view of the fact that, even though acoustic signals have fluctuations, they fluctuate within a narrow range in case the singer or the like intends to produce the same musical interval, and, as a matter of fact, it has been ascertained that there is a very high degree of correlation between the pitch information for the analytical point giving the length of the longest series and the intended musical scale.

In Fig. 35, an example is given for the identification of the musical intervals of the input acoustic signals by this process.

In Fig. 35, the distribution of the pitch information in respect of the analytical cycle is shown by a dotted-line curve PIT. The vertical lines VR1, VR2, VR3 and VR4 represent the divisions of segments as established by the segmentation process while the solid line HR in the horizontal direction expresses the pitch information on the analytical point which gives the length of the longest series in that segment. Moreover, the dotted line HP represents the musical interval identified by the pitch information. As it is evident from this Fig. 35, the pitch information which gives the length of the longest series has a very minor deviation in relation to the musical interval on the axis of the absolute musical interval, and it is thus understood that this method is capable of identifying the musical intervals well.

Accordingly, the embodiment described above can perform the identification of the musical intervals with less errors since it is designed to identify the musical interval of each segment on the basis of the section where the change in the pitch information in the segment is small and in continuum, i.e. the section where the change in the musical interval is small, by extracting the at the analytical point where the length of the series found with respect to the analytical point for each segment will be the largest.

Correction of Identified Musical Interval

Next, a detailed description is presented, with reference to the flow chart in Fig. 36, about the process (the step, SP 10, in Fig. 3) for correcting the musical intervals identified by the musical interval identifying process at the above- mentioned step, SP 7.

Before executing such a process for correcting the musical intervals, the CPU 1 first obtains, for example, the average value of the pitch information in the particular segment, with respect to the segments obtained by segmentation, and identifies the musical interval of the segment with such one of the musical intervals with a difference by a half step on the axis of the absolute musical interval as is closest to the average value obtained of the pitch information in the segment (Step SP 230), and thereafter prepares a histogram with regard to the twelve-step musical scale for all the pitch information, finding the weighing coefficient determined for each step in the musical scale by the key and its product sum with the frequency of occurrence of each musical scale, and determines the key which gives the maximum product sum as the key for the particular acoustic signal (Step SP 231).

In the correcting process, the CPU 1 first ascertains that the processing of the final segment has not been completed yet, and then, judging whether or not the musical interval identified for the segment taken as the object of the processing is any of those musical intervals (for example, mi, fa, si, do, if on the C-major key) which are different by a half step from the musical intervals mutually adjacent on the musical interval on the determined key, and, in case it is different, the CPU 1 takes up the next segment as the object of its processing, without making any correction of the musical interval, and returns to the step, SP 232 (Steps SP 232 to SP 234).

On the other hand, if the identified musical interval in the segment being processed is any of those musical intervals, the CPU 1 works out the classified totals of the items of the pitch information existing between the identified musical interval of the segment and the musical interval different therefrom by a half step on the musical scale for the key so determined (Step SP 235). For example, in case the musical interval for the segment being processed is "mi" on the C-major key, the CPU 1 finds the distribution of the pitch information present between the sets of information respectively corresponding to "mi" and "fa" in the particular segment being processed. It follows from this that the pitch information not present between these half steps will not be calculated for determining the classified total, even if it is part of the pitch information within this segment. Then, the CPU 1 finds whether there are more items of pitch information larger than the pitch information on this half-step intermediate section or there are more items of pitch information smaller than the pitch information on this half- step intermediate section and identifies the musical interval which is closer to the pitch information present in a greater number of items on the axis of the absolute musical interval as the musical interval for the segment (Step SP 236).

Upon completion of the review and correction of the results of the identification process, the CPU takes up the next segment as the object of its processing and returns to the above-mentioned step, SP 232.

It is in view of the greater possibility of mistakes in identification due to the difference by a half step from the adjacent musical intervals that the system has been thus designed to review the musical intervals in case the identified musical intervals are those with a half-step difference from the adjacent musical intervals on the key determined for them.

With the repetition of the above-mentioned process, thereby executing the review of the musical intervals with respect to all the segments until the review of the final segment is completed, the CPU 1 obtains an affirmative result at the step SP 232 and finishes the particular processing program.

Fig. 37 shows one example of the correction of a once identified musical interval, in which the determined key is the C-major key and the musical interval identified on the basis of the average value of the pitch information is "mi". This segment is put to the correcting process as its identified musical interval is "mi" and the pitch information present between "mi" and "fa" - consequently, only the pitch information in the period T1 - is calculated to determine the classified totals and the pitch information upward and downward of the pitch information value PC for the section intermediate between "mi" and "fa" is calculated to work out the classified total, and, since the pitch information greater than the pitch information value PC is predominant in this period T1, the musical interval of this segment is re-identified with the musical interval for "fa".

Therefore, the embodiment given above is capable of accurately identifying the musical interval of each segment because it is designed to perform a more detailed review of the musical interval of the segment in the case of any musical interval in which the difference between the adjacent musical intervals is a half step on the key determined for the identified musical interval. Moreover, the embodiment given above shows a system which identifies a segment with the musical interval to which the average value of the pitch information is found to be closest, but it is also possible to apply a similar manner of review to those

musical intervals identified by another method of identifying musical intervals.

Also, the above-mentioned embodiment has been designed to re-identify the musical intervals, depending on the relative volume of the larger pitch information and the smaller pitch information than the pitch information in the section intermediate between the two segments taken as the objects of the review, but another method may be employed to conduct such a review. For example, the review may be done on the basis of the average value or the most frequent value of the pitch information present in the section between the two musical intervals taken as the objects of such a review out of the pitch information on the particular segment being processed.

10 Process for Determining A Key

Next, a detailed description is provided, with reference to the flow chart in Fig. 38, about the process for determining the key inherent in the acoustic signals (Step SP 9 in Fig. 3) by the automatic music transcription system like this.

15 The CPU 1 develops histograms on the musical scale from all the pitch information as tuned by the above-mentioned tuning process (Step SP 240). At this juncture, the musical scale histogram means the histograms relating to the twelve musical scales on the axis of the absolute musical interval, i.e. those in "C (do)," "C sharp: D flat (do#: reb)," "D (re)," ..., "A (la)," "A sharp: B flat (la#: sib)," "B (si)," and, in case the pitch information is not present on the axis of the absolute musical interval, the histograms will represent the classified totals of the values as allocated to those musical scales on the two musical intervals on the axis of the absolute musical interval to which the pitch information is closest in proportion to the distance to those intervals. For this reason, the musical interval which is different by one octave is to be treated as the same musical interval.

Next, the CPU 1 obtains product sum of the weighing coefficients as illustrated in Fig. 39 and as determined by the respective keys and the above-mentioned musical scale histograms with respect to all of the 24 keys in total, which are the twelve major keys, "C major," "D flat major," "D major," ..., "B flat major," "B major," and the twelve minor keys, "A minor," "B flat minor," "B minor," ..., "G minor," "A flat minor" (Step SP 241).

Moreover, Fig. 39 indicates the weighing coefficient for "C major" in the first column, COL 1, that for "A minor" in the second column, COL 2, that for "D flat major" in the third column, COL 3, and that for "B flat minor" in the fourth column, COL 4. For the other keys, the system applies the same process, using the weighing coefficient, "202021020201," as from the keynote (do) for the major keys and using the weighing coefficient, "202201022010," as from the keynote (la) for the minor keys.

Here, the weighing coefficients are determined in such a way that a weight other than "0" is given to those musical intervals which can be expressed without the temporary signatures (#, b) for the particular key and also that "2" is used for the matching of the pentatonic and septitonic musical scales in the major keys and the minor keys, i.e. for the musical scales in which there will be an agreement in the musical interval difference from the keynote when the keynotes are brought into agreement between a major key and a minor key, and that "1" is used for the musical scales with no agreement of the difference in musical interval. Furthermore, these weighing coefficients are in correspondence to the degrees of importance of the individual musical intervals in the particular key.

When the CPU 1 has obtained the product sums for all the 24 keys in this manner, it determines the key in which the product sum is the largest as the key for the particular acoustic signals, and it finishes the particular process for determining the key (Step SP 242).

45 Therefore, the embodiment mentioned above prepares histograms for musical scales, captures the frequency of occurrence in respect of the musical scales for the individual musical intervals, finds the product sum with the weighing coefficient as the parameter of importance for the musical interval to be determined in accordance with the frequency of occurrence and the key, and determines the key in which the product sum is the largest as the key for the acoustic signals, and consequently the system is capable of accurately determining the key for such signals and reviewing the musical intervals identified on the basis of such a key, thereby making a further improvement on the accuracy of the musical score data.

Moreover, the weighing coefficients are not confined to those cited in the embodiment mentioned above, and it is feasible, for example, to give a heavier weight to the keynote.

Moreover, the means of determining the key are not limited to those mentioned above, and the determination of the key may be executed by the processing procedure shown in Fig. 40. It is omitted to explain this procedure since it is the same as the procedure shown in Fig. 38 up to the step, SP 241.

When the CPU 1 obtains the product sums for the 24 keys at the step, SP 241, it extracts the key with the largest product sum for the major key and the key with the largest product sum for the minor key,

respectively (Step SP 243). Thereafter, the CPU 1 extracts the key in which the dominant key (the key higher by five degrees from the keynote) in the candidate key is the keynote for the extracted major key and the key in which the subdominant key (i.e. the key lower by five degrees from the keynote) in the candidate key is the keynote for the extracted major key and also extracts the key in which the dominant key (i.e. the key higher by five degrees from the keynote) in the candidate key is the keynote for the extracted minor key and the key in which the subdominant key (i.e. the key lower by five degrees from the keynote) in the candidate key is the keynote for the extracted minor key (Step SP 244).

The CPU 1 finally determines the proper key by selecting one key out of a total of the six candidate keys extracted in this way on the basis of the relationship between the initial note (i.e. the musical interval of the initial segment) and the final note (i.e. the musical interval of the final segment) (Step SP 245).

The system has been thus designed not to determine the key having the largest product sum at once as the key which the acoustic signal has in view of the fact that the keynote, the dominant note, and the subdominant note frequently occur in the melody of a piece of music and that it may be quite frequent in some cases for the dominant note and the subdominant note to be generated from the keynote, and that the determination of the key merely by the largest value for the product sum could result in the determination not of the real key but of the key in which the dominant note or the subdominant note in the real key serves as the keynote. Therefore, now that it is found from an empirical rule that the initial sound and final sound in a piece of music have a unique relationship in respect of the key, as mentioned above, it has been designed to make the final determination of the key on the basis of this relationship. In the case of the C major key, for example, it is observed that music frequently starts with either one of the notes, "do," "mi," and "so" and ends with "do," and, also in the other keys, music often ends with the keynote. Therefore, the system according to the embodiment given above is capable of accurately determining the key, reviewing the musical interval identified on the basis of such a key, and further improving the accuracy of the musical score data because there has been designed to prepare musical scale histograms, thereby capturing the frequency of occurrence of each musical scale, to find the product sum with the weighing coefficient as the parameter for the degree of importance of the musical scales as determined in accordance with the frequency and the key, to extract six keys as the candidate keys on the basis of the product sum, and finally to determine the key with reference to the initial note and final note in the piece of music.

Furthermore, the embodiment mentioned above has been so designed as to obtain a total of six candidate keys through its extraction of the key with the maximum product sum for the major key and the minor key, respectively, and yet it is a feasible method finally to determine the key out of a total of three candidate keys to be extracted out of those keys with the maximum product sum to be extracted without any regard to the distinction between the major key and the minor key.

Tuning Process

Next, a detailed description is presented with reference to the detailed flow chart in Fig. 41 about the tuning process (Step SP 3 in Fig. 3) in an automatic music transcription system which performs the transcription of musical scores by its execution of this process.

The CPU 1 first converts the input pitch information expressed in Hz, which is a unit for frequency, into pitch data expressed in cent (in a value derived by multiplying with 1,200 the ratio of the frequency of a given musical interval to the standard musical interval as expressed in terms of a logarithm with 2 forming its base), which is a unit for the musical scale (Step SP 250). In this regard, a difference by 100 cents corresponds to the half-step difference in the musical interval. After that, the CPU 1 prepares a histogram like the one shown in Fig. 42 calculating the classified totals of the individual sets of pitch data with identical numerical values forming the lowest two digits of the cent values (Step SP 251). In specific terms, the CPU 1 performs arithmetic operations to work out the classified totals, treating the data with the cent values of 0, 100, 200, ... as identical data, treating the data with the cent values of 101, 201, ... as identical data, and treating the data with the cent values of 2, 102, 202, ... as identical data, until it completes the calculation to find the classified totals of the group of data with the cent values of 99, 199, 299, Thus, the system develops a histogram for the pitch information with a full-width of 100 cents varying by one cent as illustrated in Fig. 42.

At this juncture, the pitch information different by every 100 cents but calculated as identical for the calculation of the classified totals contains differences by the integral times of the half step, and the acoustic signals take the half step and the full step as the standards for a difference in the musical interval. Hence, the histograms developed by this system do not assume any uniform distribution, but indicate the peak of frequency in the proximity of the cent value which corresponds to the axis of musical interval held

by the singer who has uttered the acoustic signals or by the particular musical instrument which has generated such signals.

Next, the CPU 1 clears the parameters i and j to zero and sets the parameter MIN at A, which is a sufficiently large value (Step SP 252). Then, the CPU 1 performs arithmetic operations for determining a statistical dispersion, VAR, centering around the cent value i , using the histogram information obtained (Step SP 253). After that, the CPU 1 judges whether or not the dispersion value VAR obtained by the calculation is larger than the parameter MIN, and it renews the dispersion value VAR at the value of the parameter MIN in case the VAR value is smaller than the parameter and also modifies the parameter j to assume the value of the parameter i , thereafter proceeding to the step, SP 256. In case the VAR value is larger than the parameter MIN, the CPU 1 proceeds immediately to the step, SP 256, without performing the renewal operation (Steps SP 254 to SP 255). After that, the CPU 1 judges whether or not the parameter i has the value 99, and, in case it is different in value, it increments the parameter i , thereafter returning to the above-mentioned step, SP 253 (Step SP 257).

In this manner, the CPU 1 obtains the cent information (j) with the minimum dispersion from the classified total information obtained on the pitch information. Here, since the dispersion around the cent information is the smallest, it can be judged to be a cent group (j , $100 + j$, $200 + j$, ...) by every half step forming the center of the acoustic signal. In other words, it can be interpreted that the cent group expresses the axis of the musical interval for the singer or the musical instrument.

Therefore, the CPU 1 slides the axis of the musical interval by the value of this cent information, thereby fitting this axis into that of the absolute musical interval. First, the CPU 1 judges whether or not the parameter j is smaller than 50 cents, i.e. to which of the axes of the absolute musical interval, that of the higher tones or that of the lower tones, the parameter j is closer, and, in case the parameter is closer to the higher-tone axis, the CPU 1 modifies all the pitch information by sliding it towards the higher-tone axis by the obtained value of the cent j , but, in case the parameter is closer to the lower-tone axis, the CPU 1 modifies all the pitch information by sliding it towards the lower-tone axis by the value obtained of the cent j (Step SP 258 to SP 260).

In this manner, the axis of the acoustic signals is fitted almost exactly into the axis of the absolute musical interval, and the pitch information developed in this way is used for the subsequent processes.

Therefore, the embodiment mentioned above is capable of attaining higher accuracy in the musical score data to be obtained, whatever the source of the acoustic signal may be, because the system does not apply the obtained information as it is to the segmentation process or to such processes as that for identifying the musical intervals, but finds the classified totals by every half step on the same axis, detecting the amount of the deviation from the axis of the absolute musical interval out of the information on the classified totals by applying the dispersion as the parameter, and modifying the axis of the musical interval for the acoustic signal by the amount of the deviation, so that the modified pitch information may be used for the subsequent processes.

Moreover, the embodiment mentioned above presents a system which performs a tuning process on the pitch information obtained through autocorrelation analysis, but the method of extracting the pitch information is, of course, not to be confined to this.

In the above-mentioned embodiment, moreover, the system obtains the axis of the musical interval for the acoustic signal by the application of dispersion, and yet another statistical technique may be applied to the detecting process for the axis.

Furthermore, the embodiment given above uses cents as the unit for the pitch information subjected to the statistical processing in the tuning process, but it goes without saying that the applicable units are not limited to this.

Extraction of Pitch Information

Next, a further description is given with regard to the extraction of pitch information (Refer to the step, SP 1, in Fig. 3) in an automatic music transcription system which performs musical score transcription by performing this process.

A detailed flow chart for such a process of extracting the pitch information is presented in Fig. 43. First, from the N -pieces of acoustic signal $y(t)$ ($t=0, \dots, N-1$; where t expresses the sampling number with the sampling point s being set at 0) which is located inside the analytical windows at the noted sampling point s and the subsequent sampling points, the CPU 1 finds the autocorrelation function $\phi(\tau)$ ($\tau=0, \dots, N-1$; $\mu=0, \dots, N-1-\tau$) as expressed in the following equation (Step SP 270):

$$\phi(\tau) = \sum_{u} y(u) y(u+\tau) \quad \dots \quad (4)$$

5 which expresses the above-mentioned acoustic signal, $y(t)$, and the acoustic signal obtained by sliding the acoustic signal by the amount of τ pieces in relation to the noted sampling point s . Moreover, the autocorrelation function curve obtained in this manner is presented in Fig. 44.

Next, the CPU 1 detects the amount of deviation, z , which gives the maximum of the local maximum for the autocorrelation functions $\phi(\tau)$ by an amount of deviation other than 0, i.e. the pitch cycle for the acoustic
 10 signal as expressed in terms of the scale for the sampling number, from the value of the autocorrelation functions $\phi(\tau)$ for the N -pieces, and the CPU 1 takes out the autocorrelation functions, $\phi(z-1)$, $\phi(z)$, $\phi(z+1)$ regarding the three preceding and following amounts of deviation, $z-1$, z , $z+1$, in total, including this amount of deviation z (Step SP 271). Upon completion of this extraction, the CPU 1 performs a interpolation process for normalizing these autocorrelation functions, $\phi(z-1)$, $\phi(z)$, $\phi(z+1)$ in the manner expressed in the
 15 following equations (Step SP 272):

$$p_1 = \phi(z-1) / (N - z + 1) \quad (5)$$

$$p_2 = \phi(z) / (N - z) \quad (6)$$

$$20 \quad p_3 = \phi(z+1) / (N - z - 1) \quad (7)$$

The reason why this system employs this procedure is that, because of the analytical windows provided here, the number of pieces to be added, $(N - \tau)$ pieces, in the calculation of the sum of products would
 25 decrease, according as the amount of deviation τ becomes larger, if the arithmetic operations to find the autocorrelation functions according to the equation (4) were performed and that each of the maximums for the autocorrelation functions, which should become equal when the amount of deviation τ is enlarged, would decline gradually along with the passage of time as shown in Fig. 44 under the influence of such a decrease in the number of pieces for addition. Therefore, the interpolation process for normalization is
 30 performed in order to eliminate such influence.

Then, the CPU 1 obtains the pitch cycle τ_p expressed for the acoustic signal on the scale of the sampling number as smoothed through arithmetic operations performed with the following equation (Step SP 273):

$$35 \quad \tau_p = z - (p_3 - p_1) / [2 \{ (p_1 - p_2) (p_2 - p_3) \}] \quad (8)$$

Here, the equation (8) is to be used for calculating the amount of deviation, τ_p as expressed on the scale of the sampling number giving the maximum value on a parabola CUR conceived as a parabola passing through the autocorrelation values for the amount of deviation z , which is considered to represent
 40 the pitch cycle for the acoustic signal expressed on the scale of the sampling number once obtained, and for the amounts of deviation, $z-1$, and $z+1$, respectively preceding and following the amount of deviation z (Refer to Fig. 44). In other words, the system extracts the amount of deviation which gives the maximum value out of the information contained in the parabola by drawing the parabola in approximation of the curve in the proximity of the first maximum value for the autocorrelation function $\phi(\tau)$.

45 This feature has been adopted in order to avoid the inadequacy that it has hitherto been impossible to extract the pitch information accurately because the pitch cycle (z) where the maximum value will become the largest, if found, clarifies only its position in a sampling point and because the conventional approach could not detect the local maximum even when it exists between the sampling points, so that the resulting information would contain errors to that extent, because the autocorrelation function $\phi(\tau)$ is obtained at
 50 each sampling point.

Furthermore, since the autocorrelation function $\phi(\tau)$ can be expressed by a cosine function, which, with Maclaurin's expansion applied thereto, can be expressed in an even function, it is possible to express the same in a parabolic function if the terms upward of the fourth-degree can be ignored and the amount of deviation which gives the local maximum can be found with little difference from the actual amount of
 55 deviation even if the amount of deviation is calculated by approximation in a parabola.

Next, the CPU 1 calculates the pitch frequency f_p from the pitch cycle τ_p of the acoustic signal expressed with reference to the scale for the sampling number in accordance with the equation given in the following:

$$f_p = f_s / \tau_p \quad (9)$$

and then the CPU 1 moves on to the next process (Step SP 274). Moreover, f_s represents the sampling frequency. Accordingly, the embodiment mentioned above can find the local maximum of the autocorrelation function even if the maximum is positioned between the sampling points and can therefore extract the pitch frequency more accurately in comparison with the conventional method without raising the sampling frequency, so that the system can more accurately execute such subsequent processes as the segmentation, the musical interval identification, and the key determination.

In the embodiment given above, the interpolation process for normalization for eliminating the influence of the analytical windows is performed prior to the interpolation of the pitch cycle, and yet it is acceptable to make the interpolation of the pitch cycle while omitting such a normalizing process.

Moreover, another embodiment described above shows a system which perform the correction of the pitch cycle by applying a parabola. Such a correction may be made with another function. For example, such a correction may be made with an even function of the fourth degree by applying the autocorrelation functions for the five preceding and following points of the amount of deviation corresponding to the once obtained pitch frequency.

Moreover, the process for extracting the pitch information (Step SP 1 in Fig. 3) may be performed also by the procedure shown in the flow chart in Fig. 45. First, from the N-pieces of acoustic signal $y(t)$ ($t=0, \dots, N-1$; where t expresses the sampling number with the sampling point s being set at 0) which is located inside the analytical windows at the noted sampling point s and the subsequent sampling points, the CPU 1 finds the autocorrelation function, the CPU 1, operating by this procedure, first finds by arithmetic operation the autocorrelation function $\phi(\tau)$ ($\tau=0, \dots, N-1$; $u=0, \dots, N-1-\tau$) expressed in the equation (4) (step SP 280).

The equation (4) expresses the above-mentioned acoustic signal, $y(t)$, and the acoustic signal obtained by sliding the acoustic signal by the amount of r pieces in relation to the noted sampling point s . Moreover, the autocorrelation function curve obtained in this manner is presented in Figs. 46A and 46B, respectively.

Next, the CPU 1 detects the amount of deviation, z , which gives the maximum value for the autocorrelation functions $\phi(\tau)$ by an amount of deviation other than 0, i.e. the pitch cycle for the acoustic signal as expressed in terms of the scale for the sampling number, from the values of the N-pieces of the autocorrelation functions $\phi(\tau)$ (Step SP 281).

Thereafter, the CPU 1 takes out the autocorrelation functions, $\phi(z-1)$, $\phi(z)$, $\phi(z+1)$ for the three preceding and following amounts of deviation, $z-1$, z , $z+1$, including this amount of deviation z and calculates the parameter A expressed in the following equation (Steps SP 282 and SP 283). Moreover, the parameter A is the weighing average for the autocorrelation functions, $\phi(z-1)$, $\phi(z)$, and $\phi(z+1)$.

$$A = \{\phi(z-1) + 2\phi(z) + \phi(z+1)\}/4 \quad (10)$$

After the completion of this process, the CPU 1 takes out the autocorrelation functions, $\phi(y)$ and $\phi(y+1)$, for the amounts of deviation y and $y+1$, which are closest to the one half amount of deviation, $z/2$, for the amount of deviation, z , and works out the parameter B expressed in the following equation:

$$B = \{\phi(y) + \phi(y+1)\}/2 \quad (11)$$

(Steps SP 284 and SP 285). Moreover, the parameter B represents the average of the autocorrelation functions, $\phi(y)$ and $\phi(y+1)$. After that, the CPU 1 compares both the parameters A and B to determine which of these has the larger value, and, in case the parameter A is larger than the parameter B , the CPU 1 selects the amount of deviation z as the amount of deviation τ_p (Steps SP 286 and SP 287). On the other hand, in case the parameter B is larger than the parameter A , the CPU 1 selects the amount of deviation, $z/2$, as the amount of deviation τ_p corresponding to the pitch (Step SP 288).

In this way, the system has been designed not to use the amount of deviation which gives the maximum value for the autocorrelation function directly as the pitch cycle in view of the observation that the autocorrelation function in the proximity of the second local maximum point is detected as the function which gives the maximum value, provided that the amount of deviation two times as large as the amount of deviation which gives the real maximum value coincides almost exactly with the sampling point and that the amount of deviation which gives the real maximum value, so that it may be judged on the basis of the relative size of the parameters A and B may be used for finding whether or not the information being processed is such a case as mentioned above and that one half of the amount of deviation is to be taken as that corresponding to the pitch cycle in case the value does not corresponds to the amount of deviation

which gives the real maximum value. Moreover, Fig. 46 (B) shows a case in which the value in the proximity of the first local maximum is detected as the maximum value, and, in this case, the parameter A will always be larger than the parameter B as shown in Fig. 46 (B), and the obtained amount of deviation z is used as it is for the pitch cycle to be used in the subsequent process.

5 The CPU 1 finds the pitch frequency f_p by arithmetic operation, in accordance with the equation (9), from the pitch frequency τ_p expressed in terms of the scale for the sampling number obtained in this manner. Then, the CPU moves on to the next process (Step 289).

Consequently, in the embodiment mentioned above, the system has been designed, for the sampling frequency, to detect the occurrence of the maximum value even when the autocorrelation function in the proximity of the second local maximum point attains the maximum value and to apply interpolation to the pitch cycle, so that the system is capable of extracting the pitch information with a higher level of accuracy in comparison with the state in the past, without raising the sampling frequency, and the system can therefore execute the subsequent processes, such as the segmentation, the musical interval identifying process, and the key determining process.

15 Furthermore, the embodiment described above features a system for which the parameters A and B used for judging whether or not the amount of deviation which gives the maximum value is what corresponds to any point in the proximity of the real peak are weighted average values, but another parameter may be used for such a judgment.

Furthermore, in the embodiment given above shows the present invention applied to an automatic music transcription system, but the present invention may be applied also to those various kinds of apparatus which require the process of extracting pitch information from acoustic signals.

20 In the above-mentioned embodiment, moreover, the CPU 1 executes all the processes shown in Fig. 3 according to the programs stored in the main storage device 3, but the system may be so designed as to make the CPU 1 execute all the processes with a hardware construction. For example, as shown in Fig. 47, where those parts in correspondence to their counterparts in Fig. 2 are represented with the same reference codes, the system may be so constructed that the acoustic signal transmitted from the acoustic signal input device 8 is amplified through the amplifying circuit 10 and thereafter converted into a digital signal by feeding it into the digital/ analog converter 12 via a pre-filter circuit 11, the acoustic signal thus converted into a digital signal being processed for autocorrelation analysis by the signal processor 13 for extracting the pitch information and being also processed for finding the sum of the square value thereby to extract the power information to be given to the processing system working with software. For a signal processor 13 to be used for a hardware construction (10 to 13) like this, it is possible to use a processor (for example, μ PD 7720 made by Nippon Electric Corporation) which is capable of performing its real-time processing of signals in the vocal sound zone and also has interfacing signals provided for the CPU 1 in the host computer. A system according to the present invention is capable of performing highly accurate segmentation without being influenced by noises or fluctuations in the power information, even if they are present, determining the key well and identifying the musical interval of each segment accurately, and generating the final musical score data with accuracy.

Moreover, a system according to the present invention is capable of providing a pitch extracting method and pitch extracting apparatus which are capable of extracting pitch information with a higher degree of accuracy, in comparison with the state in the past, without raising the sampling frequency through the utilization of autocorrelation functions.

Still further, a system according to the present invention is capable of further improving the accuracy of the post-treatment such as the process for identifying the musical intervals and thereby improving the accuracy of the finally generated musical score data.

Claims

1. A method for transcribing music, comprising the steps of:

50 inputting acoustic signal;

extracting a pitch information and an acoustic power information from said input acoustic signal;

correcting said pitch information in proportion to the amount of deviation of the pitch information for said acoustic signal from an absolute musical interval axis;

55 first dividing said acoustic signal into single sound segments on the basis of said corrected pitch information while second dividing said acoustic signal into single sound segments on the basis of changes in said power information;

third dividing said acoustic signal on the basis of both of said segment information obtained at said first and second dividing steps;

identifying musical intervals of said acoustic signals in each of said segments along the absolute musical interval axis with reference to said pitch information;

fourth dividing said acoustic signal again into single-sound segments on the basis of the point whether or not said identified musical intervals of said segments in continuum are identical;

determining a key of said acoustic signal on the basis of said extracted pitch information;

determining a time and tempo of said acoustic signal on the basis of said segment information; and

compiling musical score data from said information of said determined musical interval, sound length, key, time and tempo.

2. The method for transcribing music of Claim 1, further comprising a step of eliminating noises of and interpolation said extracted pitch and power information after said extraction of said pitch and power information.

3. The method for transcribing music of Claim 1 or 2 wherein said second dividing step comprising steps of:

comparing said power information to a predetermined value and dividing said acoustic signal into a first section larger than said predetermined value while recognizing said first section as an effective section and into a second section smaller than said predetermined value while recognizing said second section as an invalid section;

extracting a point of change in rising of said power information with respect to said effective section;

dividing said effective segment into smaller parts at said point of change in rising;

measuring length of said segments of both of said effective and invalid sections; and

connecting any segment with a length shorter than a predetermined length to the preceding segment to form one segment.

4. The method for transcribing music of Claim 2 wherein said second dividing step comprising steps of:

extracting a point of change in rising of said power information with respect to said effective section; and

dividing said acoustic signal on the basis of said extracted point of change in rising.

5. The method for transcribing music of Claim 1 or 2 wherein said second dividing step comprises the steps of:

dividing said acoustic signal into a first section larger than a predetermined acoustic power value while recognizing said first section as an effective section and into a second section smaller than said

predetermined acoustic power value while recognizing said section as an invalid section;

measuring the length of both said first and second sections; and

connecting any segment with a length shorter than a predetermined length to the preceding segment.

6. The method for transcribing music of Claim 1 or 2 wherein said second dividing step comprising steps of:

extracting a point of change in rising of said power information; and

dividing said acoustic signal with respect to said point of change in rising.

7. The method for transcribing music of Claim 1 or 2 wherein said second dividing step comprising steps of:

extracting a point of change in rising of said power information;

dividing said acoustic signal with respect to said point of change in rising; and

connecting any segment with a length shorter than a predetermined length to preceding segment.

8. The method for transcribing music of one of the preceding claims wherein said first dividing step comprises the steps of:

calculating a length of each of a series of sampling points on the basis of said extracted pitch information;

detecting a section in which said calculated length of said series exceeding a predetermined value continues;

extracting a sampling point within the series of points having the maximum length in respect of each of said detected sections and recognizing said sampling point as a typical point;

detecting the amount of the variation in said pitch information between said typical points with respect

to the individual sampling points between them when the difference in said pitch information at two adjacent typical points exceeds a predetermined value; and dividing said acoustic signals at said sampling point where the amount of the pitch variation is in the maximum.

5 9. The method for transcribing music of one of the preceding claims wherein said third dividing step comprising steps of:
determining a standard length corresponding to a predetermined duration of time of a note on the basis of each of the length of said segment divided at said first dividing step; and
10 dividing said first divided segment on the basis of said standard length and dividing again in detail said divided segment having a length longer than said predetermined duration of time of said note.

10. The method for transcribing music of one of the preceding claims wherein said musical intervals identifying step comprises the steps of:
15 calculating the distance along the absolute musical interval axis between each of said segment of said pitch information and said absolute musical interval;
detecting the smallest distance; and
recognizing said musical interval of the smallest distance as an actual musical interval of said segment.

20 11. The method for transcribing music of one of claims 1 to 9, wherein said musical intervals identifying step comprises the steps of:
calculating an average value of all said pitch information of said segment; and
identifying said musical interval of said segment found on the absolute musical interval axis and closest to said calculated average value as an actual musical interval for the particular segment.

25 12. The method for transcribing music of one of claims 1 to 9, wherein said musical intervals identifying step comprises the steps of:
extracting an intermediate value of said pitch information of each segments; and
identifying the musical interval having an intermediate value closest to said absolute musical interval as
30 an actual musical interval.

13. The method for transcribing music of one of claims 1 to 9, wherein said musical intervals identifying step comprising steps of:
extracting the most frequent value of said pitch information; and
35 identifying the musical interval the most frequent value of its pitch information is the closest to that of the absolute musical interval as an actual musical interval.

14. The method for transcribing music of one of claims 1 to 9, wherein said musical intervals identifying step comprises the steps of:
40 extracting a pitch information on the peak point in the rise of said power information for each segment; and
identifying the musical interval having a peak point closest to said pitch information as an actual musical interval.

45 15. The method for transcribing music of one of claims 1 to 9, wherein said musical intervals identifying step comprises the steps of:
calculating the length of the series found with respect to an analytical point for each segment;
extracting a segment having the maximum length of the series; and
identifying the extracted musical interval to the absolute musical interval based on said pitch information for the analytical point having said maximum length of the series.
50

16. The method for transcribing music of one of claims 10 to 15, wherein said musical intervals identifying step comprises the steps of:
extracting segments a length of which is lower than a predetermined value;
55 extracting segments in which a pitch changes at a constant rate;
detecting difference in identified musical interval between said extracted segment and adjacent segments; and
identifying the musical interval, one of the difference of which is smaller than a predetermined value, as

an actual musical interval.

- 17.** The method for transcribing music of claim 16, wherein said musical interval identifying step comprising steps of:

5 extracting segments of said musical interval different from adjacent musical interval by a half step on the musical scale for the key;
 classifying totals of the items of said pitch information existing between said identified musical interval of said segment and said musical interval different therefrom by the half step on the musical scale for the key; and
 10 identifying an actual musical interval of said segment in accordance with said classified totals of the items of said pitch information.

- 18.** The method for transcribing music of one of the preceding claims, wherein said key determining step comprises the steps of:

15 classifying totals of the items of said pitch information with respect to the absolute musical interval axis;
 extracting frequency of occurrence of the musical scale of said musical interval in said acoustic signal;
 calculating a product sum with a predetermined weighing coefficient and said extracted frequency of occurrence of the musical scale of said musical interval with respect to all possible keys; and
 identifying said key having the maximum product sum as an actual key of said acoustic signal.

- 19.** The method for transcribing music of one of the preceding claims, wherein said pitch information extracting step comprising steps of:

converting an analogue signal of said inputted acoustic signal into digital form;
 calculating an autocorrelation function of said acoustic signal in the digital form;
 25 detecting an amount of deviation giving the maximum of the local maximum for said calculated autocorrelation functions by an amount of deviation other than 0;
 detecting an approximate curve through which said autocorrelation functions of a plurality of sampling points including that giving said amount of deviation pass;
 determining an amount of deviation giving the local maximum of said autocorrelation on said calculated
 30 approximate curve; and
 detecting a pitch frequency in accordance with said determined amount of deviation.

- 20.** The method for transcribing music of one of claims 1 to 19, wherein said pitch information extracting step comprising steps of:

35 converting an analogue signal of said inputted acoustic signal into digital form;
 calculating an autocorrelation function of said acoustic signal in the digital form;
 detecting a pitch information in accordance with the maximum information of said calculated autocorrelation function;
 judging whether the local maximum point of said autocorrelation function exists approximate to two-
 40 times of a frequency component of said detected pitch information; and
 outputting an actual pitch information corresponding to said local maximum if the result of said judge is positive.

- 21.** The method for transcribing music of one of claims 1 to 20, wherein said pitch information correcting step comprises the steps of:

45 classifying totals of said pitch information;
 detecting an amount of the deviation from the absolute musical interval axis of said pitch information on said classified totals; and
 modifying the musical interval for said acoustic signal by the amount of said deviation.

- 22.** An apparatus for transcribing music, comprising:

means (8) for inputting an analogue acoustic signal;
 means (10) for amplifying said inputted acoustic signal;
 means (12) for converting the analogue signal into digital form;
 55 means for processing (13) said digital acoustic signal for extracting a pitch information and an acoustic power information, said processing means including,
 means for correcting said pitch information in proportion to the amount of deviation of the pitch information for said acoustic signal from an absolute musical interval axis;

first means for dividing said acoustic signal into single sound segments on the basis of said corrected pitch information;

second means for dividing said acoustic signal into single sound segments on the basis of changes in said power information;

5 third means for dividing said acoustic signal on the basis of both of said segment information obtained by said first and second dividing means;

means for identifying musical intervals of said acoustic signals in each of said segments along the absolute musical interval axis with reference to said pitch information;

10 fourth means for dividing said acoustic signal again into single-sound segments on the basis of the point whether or not said identified musical intervals of said segments in continuum are identical;

means for determining a key of said acoustic signal on the basis of said extracted pitch information;

means for determining a time and tempo of said acoustic signal on the basis of said segment information; and

15 means for compiling musical score data from said information of said determined musical interval, sound length, key, time and tempo;

means (3) for storing the processing program;

means (1) for controlling said signal processing program; and

means (5) for displaying the transcribed music.

20 Patentansprüche

1. Verfahren zur Musiktranskribierung, das die folgenden Schritte umfaßt:

Eingabe eines akustischen Signals;

25 Ermittlung einer Tonhöheninformation und einer Schallstärkeinformation aus dem eingegebenen akustischen Signal;

Korrektur der Tonhöheninformation im Verhältnis zu dem Ausmaß der Abweichung der Tonhöheninformation für das akustische Signal von einer absoluten Tonintervallachse;

30 erste Unterteilung des akustischen Signals in einzelne Tonsegmente auf der Basis der korrigierten Tonhöheninformation, während die zweite Unterteilung des akustischen Signals in einzelne Tonsegmente auf der Basis der Veränderungen in der Schallstärkeninformation erfolgt;

dritte Unterteilung des akustischen Signals auf der Basis beider Segmentinformationen, die in dem ersten und zweiten Unterteilungsschritt erhalten wurden;

Identifizierung der Tonintervalle der akustischen Signale in jedem der Segmente entlang der absoluten Tonintervallachse mit Bezugnahme auf die Tonhöheninformation;

35 vierte Unterteilung des akustischen Signals wieder in Einzeltonsegmente aufgrund der Tatsache, ob die identifizierten Tonintervalle der fortlaufenden Segmente identisch sind oder nicht;

Bestimmung einer Tonart des akustischen Signals auf der Basis der ermittelten Tonhöheninformation;

Bestimmung eines Takts und Tempos des akustischen Signals auf der Basis der Segmentinformation; und

40 Zusammenstellung von Partiturdaten aus der erhaltenen Information über Tonintervall, Tonlänge, Tonart, Takt und Tempo.

2. Verfahren zur Musiktranskribierung nach Anspruch 1, welches ferner einen Schritt zur Beseitigung von Störgeräuschen aus der ermittelten Tonhöheninformation und Tonstärkeninformation und Interpolation derselben nach der Ermittlung der Tonhöhen- und Tonstärkeninformation umfaßt.

45

3. Verfahren zur Musiktranskribierung nach Anspruch 1 oder 2, wobei der zweite Unterteilungsschritt folgende Schritte umfaßt:

Vergleich der Tonstärkeninformation mit einem vorbestimmten Wert und Unterteilung des akustischen Signals in einen ersten Abschnitt, der größer als der vorbestimmte Wert ist, wobei der erste Abschnitt als effektiver Abschnitt erkannt wird, und in einen zweiten Abschnitt, der kleiner als der vorbestimmte Wert ist, wobei der zweite Abschnitt als ungültiger Abschnitt erkannt wird;

50 Ermittlung eines Veränderungspunktes im Anstieg der Tonstärkeninformation in Bezug auf den effektiven Abschnitt; Unterteilung des effektiven Segments in kleinere Teile an diesem Veränderungspunkt im Anstieg;

55 Messung der Segmentlänge sowohl der effektiven als auch der ungültigen Abschnitte; und Verbinden jedes Segments mit einer Länge, die geringer als eine vorbestimmte Länge ist, mit dem vorangehenden Segment zur Bildung eines Segments.

4. Verfahren zur Musiktranskribierung nach Anspruch 2, wobei der zweite Unterteilungsschritt folgende Schritte umfaßt:
Ermittlung eines Veränderungspunktes im Anstieg der Tonstärkeninformation in Bezug auf den effektiven Abschnitt; und
5 Unterteilung des akustischen Signals auf der Basis des ermittelten Veränderungspunktes im Anstieg.
5. Verfahren zur Musiktranskribierung nach Anspruch 1 oder 2, wobei der zweite Unterteilungsschritt folgende Schritte umfaßt:
Unterteilung des akustischen Signals in einen ersten Abschnitt, der größer als ein vorbestimmter
10 Tonstärkenwert ist, wobei der erste Abschnitt als effektiver Abschnitt erkannt wird, und in einen zweiten Abschnitt, der kleiner als der vorbestimmte Tonstärkenwert ist, wobei der zweite Abschnitt als ungültiger Abschnitt erkannt wird; Messung der Länge sowohl des ersten als auch des zweiten Abschnitts; und
Verbinden jedes Segments mit einer Länge, die geringer als eine vorbestimmte Länge ist, mit dem
15 vorangehenden Segment.
6. Verfahren zur Musiktranskribierung nach Anspruch 1 oder 2, wobei der zweite Unterteilungsschritt folgende Schritte umfaßt:
Ermittlung eines Veränderungspunktes im Anstieg der Tonstärkeninformation; und
20 Unterteilung des akustischen Signals in Bezug auf den Veränderungspunkt im Anstieg.
7. Verfahren zur Musiktranskribierung nach Anspruch 1 oder 2, wobei der zweite Unterteilungsschritt folgende Schritte umfaßt:
Ermittlung eines Veränderungspunktes im Anstieg der Tonstärkeninformation;
25 Unterteilung des akustischen Signals in Bezug auf den Veränderungspunkt im Anstieg; und
Verbinden jedes Segments mit einer Länge, die geringer als eine vorbestimmte Länge ist, mit dem vorangehenden Segment.
8. Verfahren zur Musiktranskribierung nach einem der vorangehenden Ansprüche, wobei der erste
30 Unterteilungsschritt folgende Schritte umfaßt:
Berechnen einer Länge einer jeden Serie von Abtastpunkten auf der Basis der ermittelten Tonhöheninformation;
Feststellen eines Abschnitts, in dem die berechnete Länge der Serien, die einen vorbestimmten Wert übersteigen, weitergeht;
35 Ermittlung eines Abtastpunktes in der Serie von Punkten mit maximaler Länge in Bezug auf jeden der nachgewiesenen Abschnitte und Erkennen des Abtastpunktes als charakteristischen Punkt;
Feststellen des Ausmaßes der Schwankung in der Tonhöheninformation zwischen den charakteristischen Punkten in Bezug auf die einzelnen dazwischenliegenden Abtastpunkte, wenn der Unterschied in der Tonhöheninformation an zwei angrenzenden charakteristischen Punkten einen vorbestimmten Wert
40 übersteigt; und
Unterteilung der akustischen Signale an dem Abtastpunkt, an dem das Ausmaß der Tonhöhenschwankung maximal ist.
9. Verfahren zur Musiktranskribierung nach einem der vorangehenden Ansprüche, wobei der dritte
45 Unterteilungsschritt folgende Schritte umfaßt:
Bestimmung einer Standardlänge entsprechend einer vorbestimmten Zeitdauer einer Note auf der Basis jeder der Längen des im ersten Unterteilungsschritt unterteilten Segments; und
Unterteilung des ersten unterteilten Segments auf der Basis der Standardlänge und neuerliche genaue
Unterteilung des unterteilten Segments, das eine Länge aufweist, die größer ist als die vorbestimmte
50 Zeitdauer der Note.
10. Verfahren zur Musiktranskribierung nach einem der vorangehenden Ansprüche, wobei der Tonintervall-Identifizierungsschritt folgende Schritte umfaßt:
Berechnen des Abstandes entlang der absoluten Tonintervallachse zwischen jedem Segment der
55 Tonhöheninformation und dem absoluten Tonintervall;
Feststellen des geringsten Abstandes; und
Erkennen des Tonintervalls mit dem geringsten Abstand als ein tatsächliches Tonintervall des Segments.

11. Verfahren zur Musiktranskribierung nach einem der Ansprüche 1 bis 9, wobei der Tonintervall-Identifizierungsschritt folgende Schritte umfaßt:
Berechnen eines Durchschnittswertes aller Tonhöheninformationen des Segments; und
Identifizierung des Tonintervalls des Segments, das auf der absoluten Tonintervallachse gefunden wurde und dem berechneten Durchschnittswert am nächsten ist, als ein tatsächliches Tonintervall für das bestimmte Segment.
12. Verfahren zur Musiktranskribierung nach einem der Ansprüche 1 bis 9, wobei der Tonintervall-Identifizierungsschritt folgende Schritte umfaßt:
Ermittlung eines Zwischenwertes der Tonhöheninformation jedes Segments; und
Identifizierung des Tonintervalls, das einen dem absoluten Tonintervall am nächsten liegenden Zwischenwert aufweist, als tatsächliches Tonintervall.
13. Verfahren zur Musiktranskribierung nach einem der Ansprüche 1 bis 9, wobei der Tonintervall-Identifizierungsschritt folgende Schritte umfaßt:
Ermittlung des häufigsten Wertes der Tonhöheninformation; und Identifizierung des Tonintervalls, dessen häufigster Wert der Tonhöheninformation jenem des absoluten Tonintervalls am nächsten ist, als tatsächliches Tonintervall.
14. Verfahren zur Musiktranskribierung nach einem der Ansprüche 1 bis 9, wobei der Tonintervall-Identifizierungsschritt folgende Schritte umfaßt:
Ermittlung einer Tonhöheninformation über den Spitzenpunkt im Anstieg der Tonstärkeninformation für jedes Segment; und
Identifizierung des Tonintervalls mit einem Spitzenpunkt, der der Tonhöheninformation am nächsten ist, als ein tatsächliches Tonintervall.
15. Verfahren zur Musiktranskribierung nach einem der Ansprüche 1 bis 9, wobei der Tonintervall-Identifizierungsschritt folgende Schritte umfaßt:
Berechnen der Länge der Serie, die in Bezug auf einen analytischen Punkt für jedes Segment ermittelt wird;
Ermittlung eines Segments mit der maximalen Serienlänge; und Identifizierung des ermittelten Tonintervalls mit dem absoluten Tonintervall auf der Basis der Tonhöheninformation für den analytischen Punkt mit der maximalen Serienlänge.
16. Verfahren zur Musiktranskribierung nach einem der Ansprüche 10 bis 15, wobei der Tonintervall-Identifizierungsschritt folgende Schritte umfaßt:
Ermittlung von Segmenten, deren Länge geringer als ein vorbestimmter Wert ist;
Ermittlung von Segmenten, bei welchen sich eine Tonhöhe in konstantem Maße verändert;
Feststellen eines Unterschieds in dem identifizierten Tonintervall zwischen dem ermittelten Segment und den angrenzenden Segmenten; und
Identifizierung des Tonintervalls, dessen Unterschied kleiner als ein vorbestimmter Wert ist, als tatsächliches Tonintervall.
17. Verfahren zur Musiktranskribierung nach Anspruch 16, wobei der Tonintervall-Identifizierungsschritt folgende Schritte umfaßt:
Ermittlung von Segmenten des Tonintervalls, die sich von dem angrenzenden Tonintervall um einen Halbtonschritt auf der Tonleiter für die Tonart unterscheiden;
Klassifizierung der Summen der Tonhöheninformationen, die zwischen dem identifizierten Tonintervall des Segments und dem Tonintervall, das sich davon um einen Halbtonschritt auf der Tonleiter für die Tonart unterscheidet, bestehen; und Identifizierung eines tatsächlichen Tonintervalls des Segments in Übereinstimmung mit den klassifizierten Summen der Tonhöheninformationen.
18. Verfahren zur Musiktranskribierung nach einem der vorangehenden Ansprüche, wobei der Tonart-Bestimmungsschritt folgende Schritte umfaßt:
Klassifizierung der Summen der Tonhöheninformationen in Bezug auf die absolute Tonintervallachse;
Ermittlung der Häufigkeit des Auftretens der Tonleiter des Tonintervalls in dem akustischen Signal;
Berechnung einer Produktsumme mit einem vorbestimmten Wichtungsfaktor und der ermittelten Häufigkeit des Auftretens der Tonleiter des Tonintervalls in Bezug auf alle möglichen Tonarten; und

Identifizierung der Tonart mit der maximalen Produktsumme als tatsächliche Tonart des akustischen Signals.

- 5 19. Verfahren zur Musiktranskribierung nach einem der vorangehenden Ansprüche, wobei der Tönhöheninformations-Ermittlungsschritt folgende Schritte umfaßt:
Umwandlung eines analogen Signals des eingegebenen akustischen Signals in eine digitale Form;
Berechnen einer Autokorrelationsfunktion des akustischen Signals in der digitalen Form;
Feststellen eines Abweichungsausmaßes, das das Maximum des lokalen Maximums für die berechneten Autokorrelationsfunktionen bei einem Abweichungsausmaß ungleich 0 ergibt;
10 Feststellen einer Näherungskurve, durch welche die Autokorrelationsfunktionen einer Mehrzahl von Abtastpunkten einschließlich jenes, der dieses Abweichungsausmaß angibt, gehen;
Bestimmung eines Abweichungsausmaßes, das das lokale Maximum der Autokorrelation auf der berechneten Näherungskurve angibt; und
Feststellen einer Tönhöhenfrequenz in Übereinstimmung mit dem bestimmten Abweichungsausmaß.
15
20. Verfahren zur Musiktranskribierung nach einem der Ansprüche 1 bis 19, wobei der Tönhöheninformations-Ermittlungsschritt folgende Schritte umfaßt:
Umwandlung eines analogen Signals des eingegebenen akustischen Signals in eine digitale Form;
Berechnen einer Autokorrelationsfunktion des akustischen Signals in der digitalen Form;
20 Feststellen einer Tönhöheninformation in Übereinstimmung mit der maximalen Information der berechneten Autokorrelationsfunktion;
Beurteilung, ob der lokale Maximalpunkt der Autokorrelationsfunktion etwa das Zweifache einer Frequenzkomponente der erfaßten Tönhöheninformation ausmacht; und
Ausgabe einer tatsächlichen Tönhöheninformation entsprechend dem lokalen Maximum, wenn das
25 Ergebnis dieser Beurteilung positiv ist.
21. Verfahren zur Musiktranskribierung nach einem der Ansprüche 1 bis 20, wobei der Tönhöheninformations-Korrekturschritt folgende Schritte umfaßt:
Klassifizierung der Summen der Tönhöheninformation; Feststellen eines Abweichungsausmaßes von
30 der absoluten Tonintervallachse der Tönhöheninformation bei den klassifizierten Summen; und
Modifizierung des Tonintervalls für das akustische Signal um das Abweichungsausmaß.
22. Gerät zur Musiktranskribierung, bestehend aus einem:
Mittel (8) zur Eingabe eines analogen akustischen Signals; Mittel (10) zur Verstärkung des eingegebenen akustischen Signals;
35 Mittel (12) zur Umwandlung des analogen Signals in die digitale Form;
Mittel zur Verarbeitung (13) des digitalen akustischen Signals zur Ermittlung einer Tönhöheninformation und einer Tonstärkeninformation, wobei dieses Verarbeitungsmittel folgendes enthält:
Mittel zur Korrektur der Tönhöheninformation im Verhältnis zu dem Abweichungsausmaß der Tönhöheninformation des akustischen Signals von einer absoluten Tonintervallachse;
40 ein erstes Mittel zur Unterteilung des akustischen Signals in einzelne Tonsegmente auf der Basis der korrigierten Tönhöheninformation;
ein zweites Mittel zur Unterteilung des akustischen Signals in einzelne Tonsegmente auf der Basis der Veränderungen in der Tonstärkeninformation;
45 ein drittes Mittel zur Unterteilung des akustischen Signals auf der Basis beider Segmentinformationen, die durch das erste und zweite Unterteilungsmittel erhalten wurden;
Mittel zur Identifizierung der Tonintervalle der akustischen Signale in jedem der Segmente entlang der absoluten Tonintervallachse in Bezug auf die Tönhöheninformation;
ein viertes Mittel zur neuerlichen Unterteilung des akustischen Signals in Einzeltonsegmente aufgrund
50 der Tatsache, ob die identifizierten Tonintervalle der fortlaufenden Segmente identisch sind;
Mittel zur Bestimmung einer Tonart des akustischen Signals auf der Basis der ermittelten Tönhöheninformation;
Mittel zur Bestimmung eines Takts und Tempos des akustischen Signals auf der Basis der Segmentinformation; und
55 Mittel zur Zusammenstellung von Partiturdaten aus der erhaltenen Information über Tonintervall, Tonlänge, Tonart, Takt und Tempo;
Mittel (3) zur Speicherung des Verarbeitungsprogramms;
Mittel (1) zur Steuerung des Signalverarbeitungsprogramms; und

Mittel (5) zur Anzeige der transkribierten Musik.

Revendications

- 5 1. Procédé pour transcrire de la musique comprenant les étapes qui consistent à:
 - entrer un signal acoustique;
 - extraire une information de ton et une information de puissance acoustique dudit signal acoustique d'entrée;
 - corriger ladite information de ton proportionnellement à la valeur de déviation de l'information de
 - 10 ton pour ledit signal acoustique par rapport à un axe des intervalles musicaux absolus;
 - diviser en premier ledit signal acoustique en segments sonores uniques selon ladite information de ton corrigée tout en divisant en deuxième ledit signal acoustique en segments sonores uniques selon les variations de ladite information de puissance;
 - diviser en troisième ledit signal acoustique selon à la fois lesdites informations de segment
 - 15 obtenues dans lesdites première et deuxième étapes de division;
 - identifier des intervalles musicaux desdits signaux acoustiques dans chacun desdits segments le long de l'axe des intervalles musicaux absolus en référence à ladite information de ton;
 - diviser en quatrième ledit signal acoustique à nouveau en segments sonores uniques selon le point que lesdits intervalles musicaux identifiés desdits segments in continuum sont ou non identiques;
 - 20 déterminer une clé dudit signal acoustique selon ladite information de ton extraite;
 - déterminer une mesure et un tempo dudit signal acoustique selon lesdites informations de segment; et
 - compiler les données de partition musicale à partir desdites informations dudit intervalle musical, de ladite longueur de son, de ladite clé, de ladite mesure et dudit tempo déterminés.
 - 25
2. Procédé pour transcrire de la musique selon la revendication 1 comprenant en outre une étape qui consiste à éliminer les bruits desdites et à interpoler lesdites informations de ton et de puissance extraites après ladite extraction desdites informations de ton et de puissance.
- 30 3. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 et 2, dans lequel ladite deuxième étape de division comprend les étapes qui consistent à:
 - comparer ladite information de puissance à une valeur prédéterminée et diviser ledit signal acoustique en une première partie plus grande que ladite valeur prédéterminée tout en identifiant ladite première partie comme une partie effective et en une seconde partie plus petite que ladite valeur
 - 35 prédéterminée tout en identifiant ladite seconde partie comme une partie incorrecte;
 - extraire un point de changement dans la croissance de ladite information de puissance par rapport à ladite partie effective;
 - diviser ladite partie effective en plus petites portions audit point de changement dans la croissance;
 - mesurer la longueur desdits segments à la fois desdites parties effectives et incorrectes; et à
 - 40 lier tout segment d'une longueur plus petite qu'une longueur prédéterminée au segment précédent pour former un segment.
4. Procédé pour transcrire de la musique selon la revendication 2, dans lequel ladite deuxième étape de division comprend les étapes qui consistent à:
 - 45 extraire un point de changement dans la croissance de ladite information de puissance par rapport à ladite partie effective; et à
 - diviser ledit signal acoustique selon ledit point de changement extrait dans la croissance.
5. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 et 2, dans lequel ladite deuxième étape de division comprend les étapes qui consistent à:
 - 50 diviser ledit signal acoustique en une première partie plus grande qu'une valeur de puissance acoustique prédéterminée tout en identifiant ladite première partie comme une partie effective et en une seconde partie plus petite que ladite valeur de puissance acoustique prédéterminée tout en identifiant ladite partie comme une partie incorrecte;
 - 55 mesurer la longueur à la fois desdites première et seconde parties; et à
 - lier tout segment de longueur inférieure à une longueur prédéterminée au segment précédent.

6. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 et 2, dans lequel ladite deuxième étape de division comprend les étapes qui consistent à:
extraire un point de changement dans la croissance de ladite information de puissance; et à
diviser ledit signal acoustique par rapport audit point de changement dans la croissance.
- 5 7. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 et 2, dans lequel ladite deuxième étape de division comprend les étapes qui consistent à:
extraire un point de changement dans la croissance de ladite information de puissance;
diviser ledit signal acoustique par rapport audit point de changement dans la croissance; et à
10 lier tout segment de longueur inférieure à une longueur prédéterminée au segment précédent.
8. Procédé pour transcrire de la musique selon l'une des revendications 1 à 7, dans lequel ladite première étape de division comprend les étapes qui consistent à :
calculer la longueur de chacun d'une série de points d'échantillonnage selon ladite information de
15 ton extraite;
détecter une partie dans laquelle ladite longueur de ladite série calculée dépassant une valeur prédéterminée continue;
extraire un point d'échantillonnage dans la série de points ayant la longueur maximale à l'égard de
chacune desdites parties détectées et identifier ledit point d'échantillonnage comme un point typique;
20 détecter la valeur de la variation de ladite information de ton entre lesdits points typiques par rapport aux points d'échantillonnage individuels entre eux quand la différence desdites informations de ton en deux points typiques voisins dépasse une valeur prédéterminée; et à
diviser lesdits signaux acoustiques audit point d'échantillonnage où la valeur de la variation de ton est au maximum.
- 25 9. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 8, dans lequel ladite troisième étape de division comprend les étapes qui consistent à:
déterminer une longueur de référence correspondant à une durée prédéterminée d'une note selon
chacune des longueurs dudit segment divisé dans ladite première étape de division; et à
30 diviser ledit premier segment divisé selon ladite longueur de référence et diviser à nouveau en détail ledit segment divisé ayant une longueur plus grande que ladite durée prédéterminée de ladite note.
- 35 10. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 9, dans lequel ladite étape d'identification d'intervalles musicaux comprend les étapes qui consistent à :
calculer la distance le long de l'axe des intervalles musicaux absolus entre chacun dudit segment de ladite information de ton et dudit intervalle musical absolu;
détecter la plus petite distance; et à
40 identifier ledit intervalle musical de la plus petite distance comme intervalle musical présent dudit segment.
11. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 9, dans lequel ladite étape d'identification d'intervalles musicaux comprend les étapes qui consistent à :
calculer une valeur moyenne de toutes les dites informations de ton dudit segment; et à
45 identifier ledit intervalle musical dudit segment trouvé sur l'axe des intervalles musicaux absolus et le plus proche de ladite valeur moyenne calculée comme intervalle musical présent pour le segment particulier.
- 50 12. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 9, dans lequel ladite étape d'identification d'intervalles musicaux comprend les étapes qui consistent à :
extraire une valeur intermédiaire de ladite information de ton de chacun des segments; et à
identifier l'intervalle musical ayant une valeur intermédiaire la plus proche dudit intervalle musical absolu comme intervalle musical présent.
- 55 13. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 9, dans lequel ladite étape d'identification d'intervalles musicaux comprend les étapes qui consistent à :
extraire la valeur la plus fréquente de ladite information de ton; et à
identifier l'intervalle musical dont la valeur la plus fréquente de son information de ton est la plus

proche de celle de l'intervalle musical absolu comme intervalle musical présent.

14. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 9, dans lequel ladite étape d'identification d'intervalles musicaux comprend les étapes qui consistent à:

extraire une information de ton au point de crête dans la croissance de ladite information de puissance pour chaque segment; et à identifier l'intervalle musical ayant un point de crête le plus proche de ladite information de ton comme intervalle musical présent.

15. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 9, dans lequel ladite étape d'identification d'intervalles musicaux comprend les étapes qui consistent à:

calculer la longueur de la série trouvée par rapport à un point analytique pour chaque segment; extraire un segment ayant la longueur maximale de la série; et à identifier l'intervalle musical extrait par rapport à l'intervalle musical absolu selon ladite information de ton pour le point analytique ayant ladite longueur maximale de la série.

16. Procédé pour transcrire de la musique selon l'une quelconque des revendications 10 à 15, dans lequel ladite étape d'identification d'intervalles musicaux comprend les étapes qui consistent à:

extraire les segments dont la longueur est inférieure à une valeur prédéterminée; extraire les segments dans lesquels le ton varie à une fréquence constante; détecter la différence d'intervalle musical identifié entre ledit segment extrait et les segments voisins; et à identifier l'intervalle musical dont la différence est inférieure à une valeur prédéterminée comme intervalle musical présent.

17. Procédé pour transcrire de la musique selon la revendication 16, dans lequel ladite étape d'identification d'intervalles musicaux comprend les étapes qui consistent à:

extraire les segments dudit intervalle musical différent d'un intervalle musical voisin d'un demi-ton dans la gamme musicale pour la clé; classer les totaux des éléments de ladite information de ton existant entre ledit intervalle musical identifié dudit segment et ledit intervalle musical différent de celui-ci du demi-ton dans la gamme musicale pour la clé; et à identifier un intervalle musical présent dudit segment conformément auxdits totaux classés des éléments de ladite information de ton.

18. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 17, dans lequel l'étape de détermination de la clé comprend les étapes qui consistent à:

classer les totaux des éléments de ladite information de ton par rapport à l'axe des intervalles musicaux absolus; extraire la fréquence d'apparition de la gamme musicale dudit intervalle musical dans ledit signal acoustique; calculer une somme de produits avec un coefficient de pondération prédéterminé et ladite fréquence d'apparition extraite de la gamme musicale dudit intervalle musical par rapport à toutes les clés possibles; et à identifier ladite clé ayant la somme de produits maximale comme clé présente dudit signal acoustique.

19. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 18, dans lequel ladite étape d'extraction d'information de ton comprend les étapes qui consistent à:

convertir un signal analogique dudit signal acoustique d'entrée sous forme numérique; calculer une fonction d'autocorrélation dudit signal acoustique sous la forme numérique; détecter la valeur de déviation donnant le maximum du maximum local pour lesdites fonctions d'autocorrélation calculées par une valeur de déviation différente de 0; détecter une courbe approximative suivie par lesdites fonctions d'autocorrélation d'une pluralité de points d'échantillonnage y compris celle donnant ladite valeur de déviation; déterminer la valeur de déviation donnant le maximum local de ladite autocorrélation sur ladite courbe approximative calculée; et à détecter une fréquence de ton selon ladite valeur de déviation déterminée.

20. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 19, dans lequel ladite étape d'extraction d'information de ton comprend les étapes qui consistent à :
- convertir un signal analogique dudit signal acoustique d'entrée sous forme numérique;
 - calculer une fonction d'autocorrélation dudit signal acoustique sous la forme numérique;
 - 5 détecter une information de ton conformément à l'information de maximum de ladite fonction d'autocorrélation calculée;
 - juger si le point de maximum local de ladite fonction d'autocorrélation est présent à proximité de deux fois la composante de fréquence de ladite information de ton détectée; et à
 - 10 sortir une information de ton présent correspondant audit maximum local si le résultat du jugement est positif.
21. Procédé pour transcrire de la musique selon l'une quelconque des revendications 1 à 20, dans lequel ladite étape de correction d'information de ton comprend les étapes qui consistent à :
- classer les totaux desdites informations de ton;
 - 15 détecter la valeur de déviation par rapport à l'axe des intervalles musicaux absolus de ladite information de ton sur lesdits totaux classés; et à
 - modifier l'intervalle musical pour ledit signal acoustique de la valeur de déviation.
22. Dispositif pour transcrire de la musique, comprenant:
- 20 un moyen (8) pour entrer un signal acoustique analogique;
 - un moyen (10) pour amplifier ledit signal acoustique entré;
 - un moyen (12) pour convertir le signal analogique sous forme numérique;
 - un moyen (13) pour traiter ledit signal acoustique numérique pour l'extraction d'une information de ton et d'une information de puissance acoustique, ledit moyen de traitement incluant:
 - 25 un moyen pour corriger ladite information de ton proportionnellement à la valeur de déviation de l'information de ton pour ledit signal acoustique par rapport à un axe des intervalles musicaux absolus;
 - un premier moyen pour diviser ledit signal acoustique en segments sonores uniques selon ladite information de ton corrigée;
 - un deuxième moyen pour diviser ledit signal acoustique en segments sonores uniques selon les
 - 30 variations de ladite information de puissance;
 - un troisième moyen pour diviser ledit signal acoustique selon à la fois lesdites informations de segment obtenues par lesdits premier et deuxième moyens;
 - un moyen pour identifier des intervalles musicaux dudit signal acoustique dans chacun desdits segments le long de l'axe des intervalles musicaux absolus en référence à ladite information de ton;
 - 35 un quatrième moyen pour diviser ledit signal acoustique à nouveau en segments sonores uniques selon le point que lesdits intervalles musicaux identifiés desdits segments in continuum sont ou non identiques;
 - un moyen pour déterminer une clé dudit signal acoustique selon ladite information de ton extraite;
 - un moyen pour déterminer une mesure et un tempo dudit signal acoustique selon lesdites
 - 40 informations de segment; et
 - un moyen pour compiler les données de partition musicale à partir desdites informations dudit intervalle musical, de ladite longueur de son, de ladite clé, de ladite mesure et dudit tempo déterminés;
 - un moyen (3) pour mémoriser le programme de traitement;
 - un moyen (1) pour contrôler ledit programme de traitement de signaux; et
 - 45 un moyen (5) pour visualiser la musique transcrite.

50

55

FIG. 1

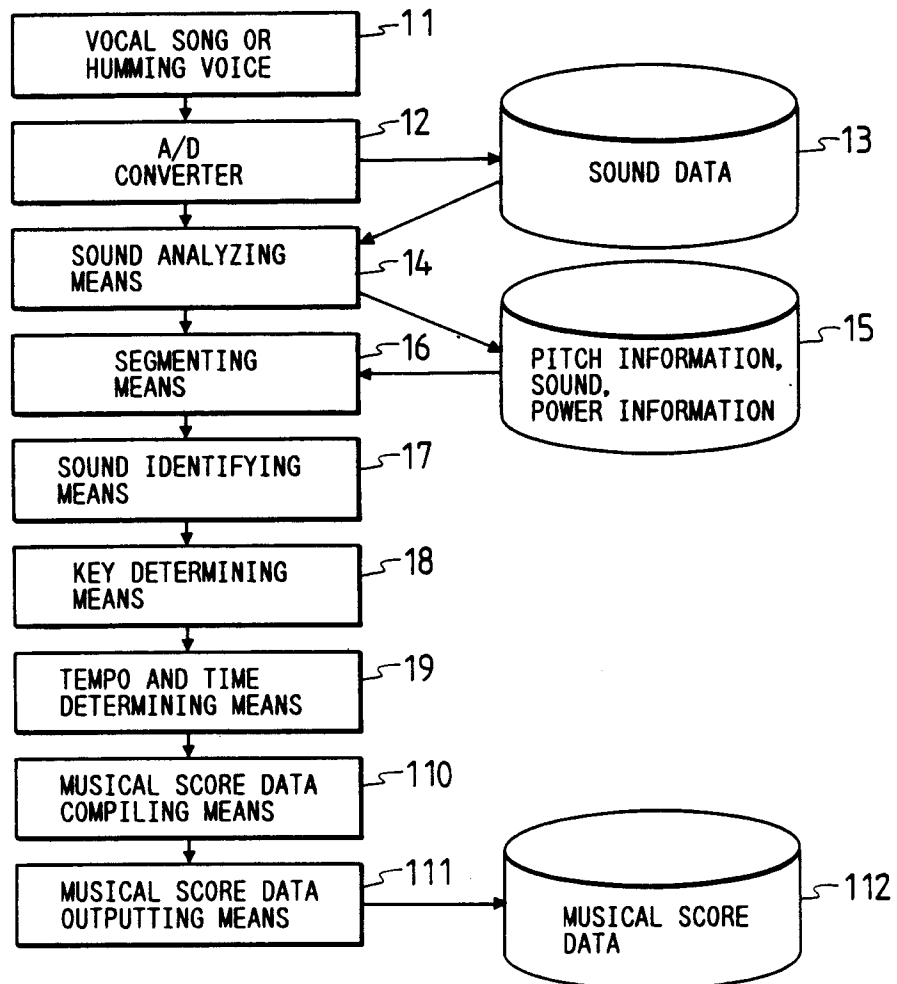


FIG. 2

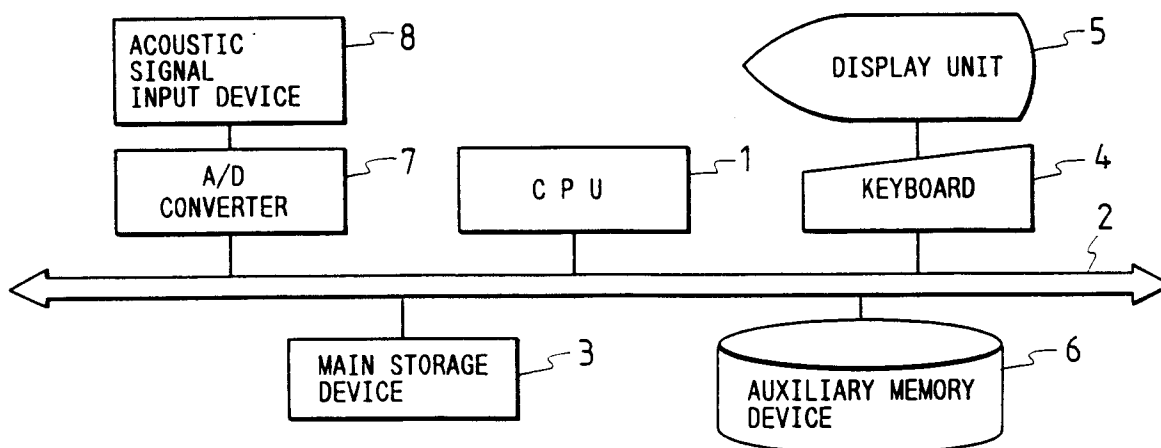


FIG. 3

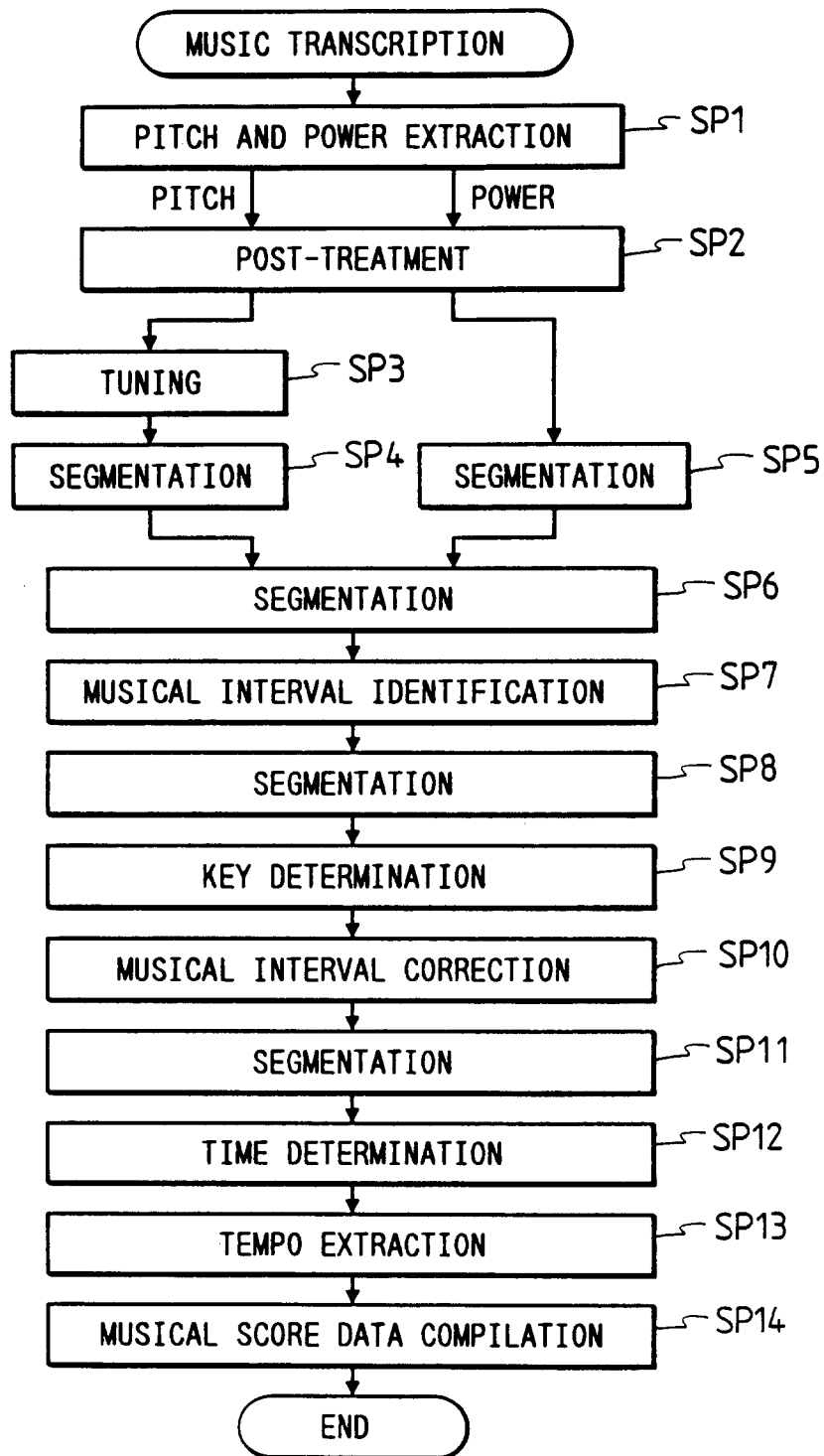


FIG. 4

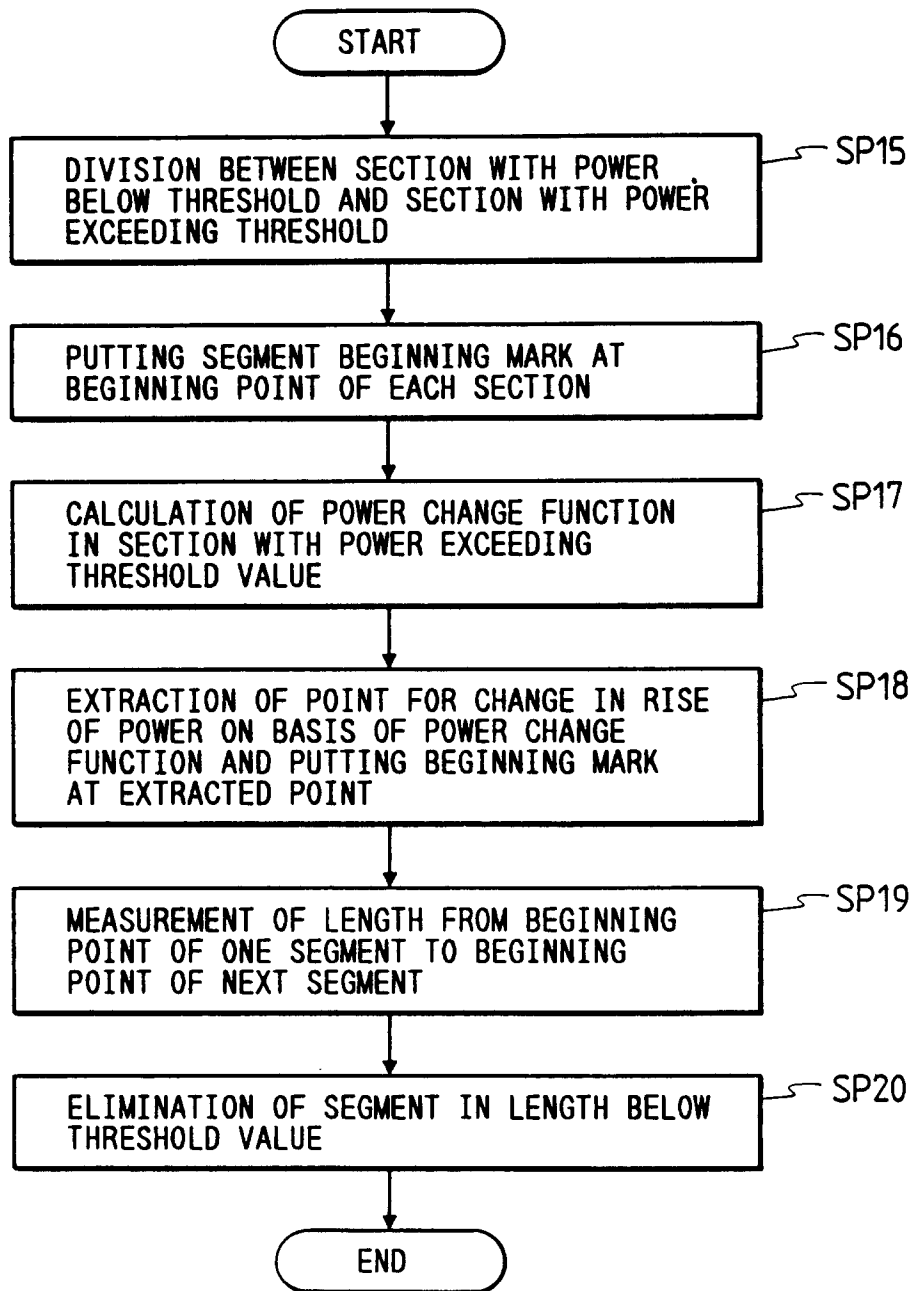


FIG. 5

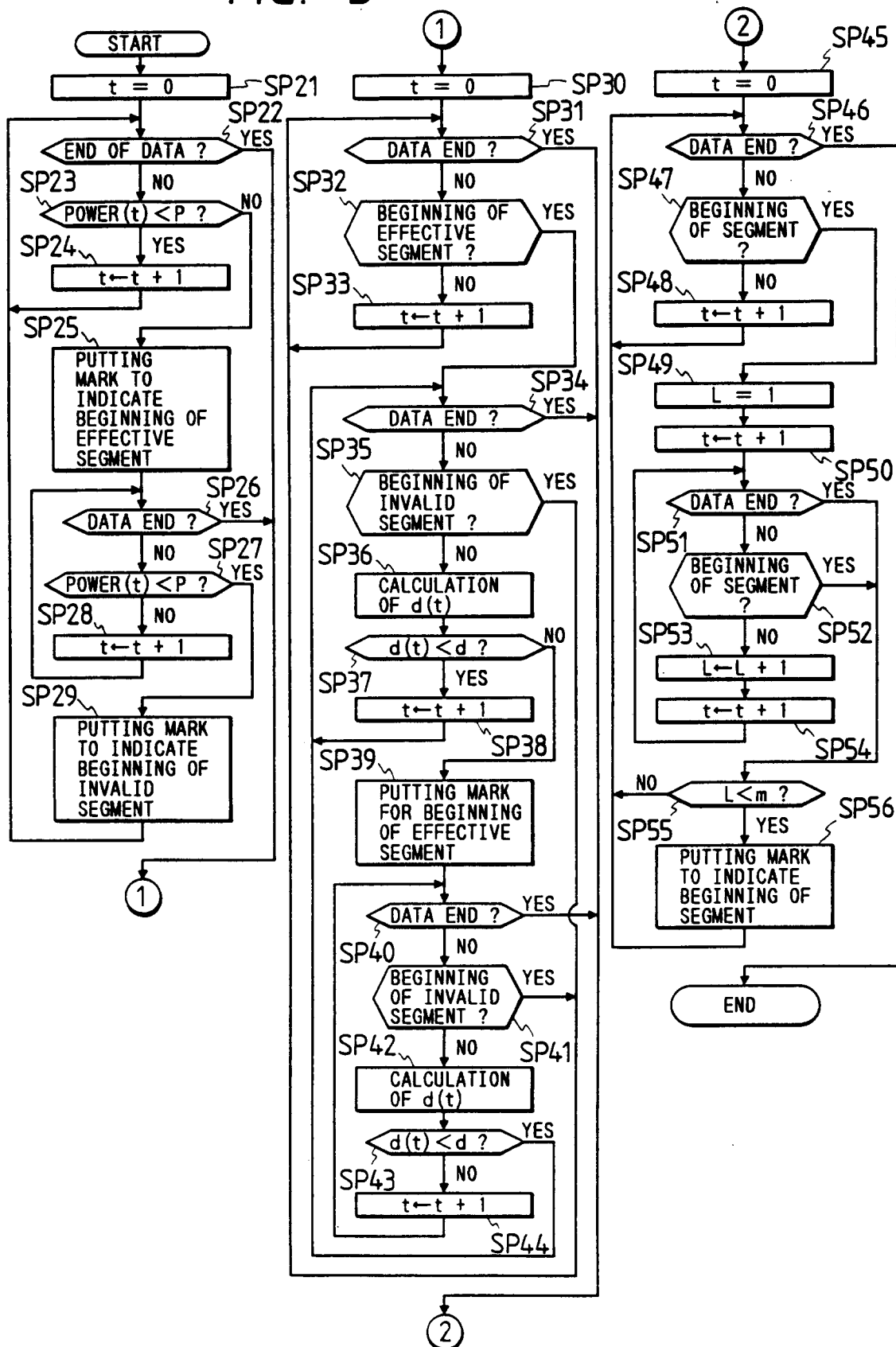


FIG. 6

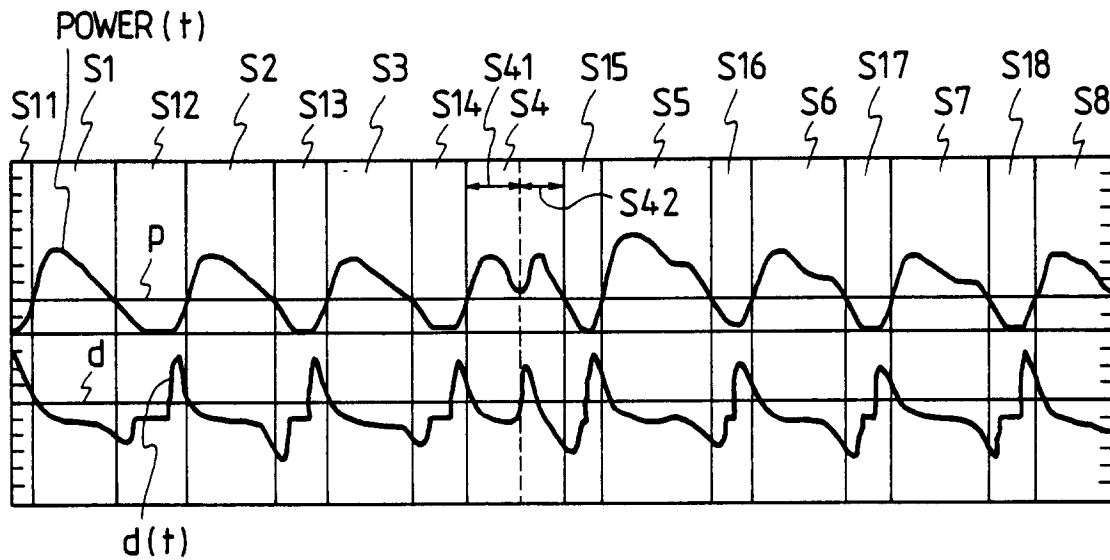


FIG. 7

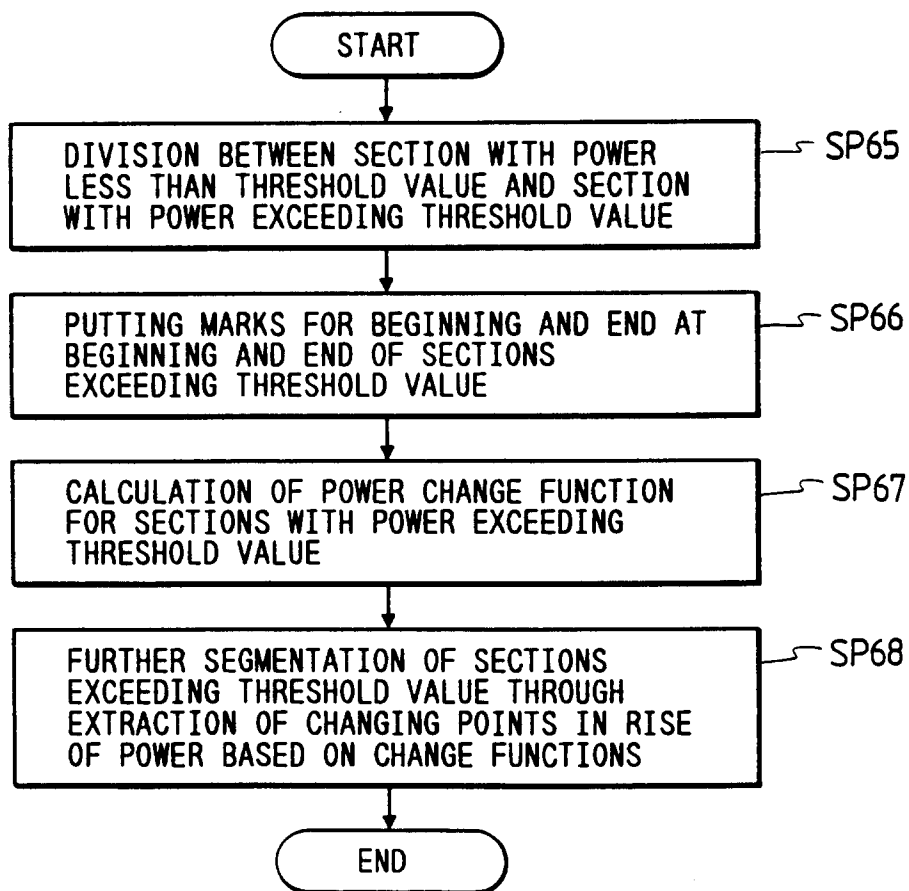


FIG. 8

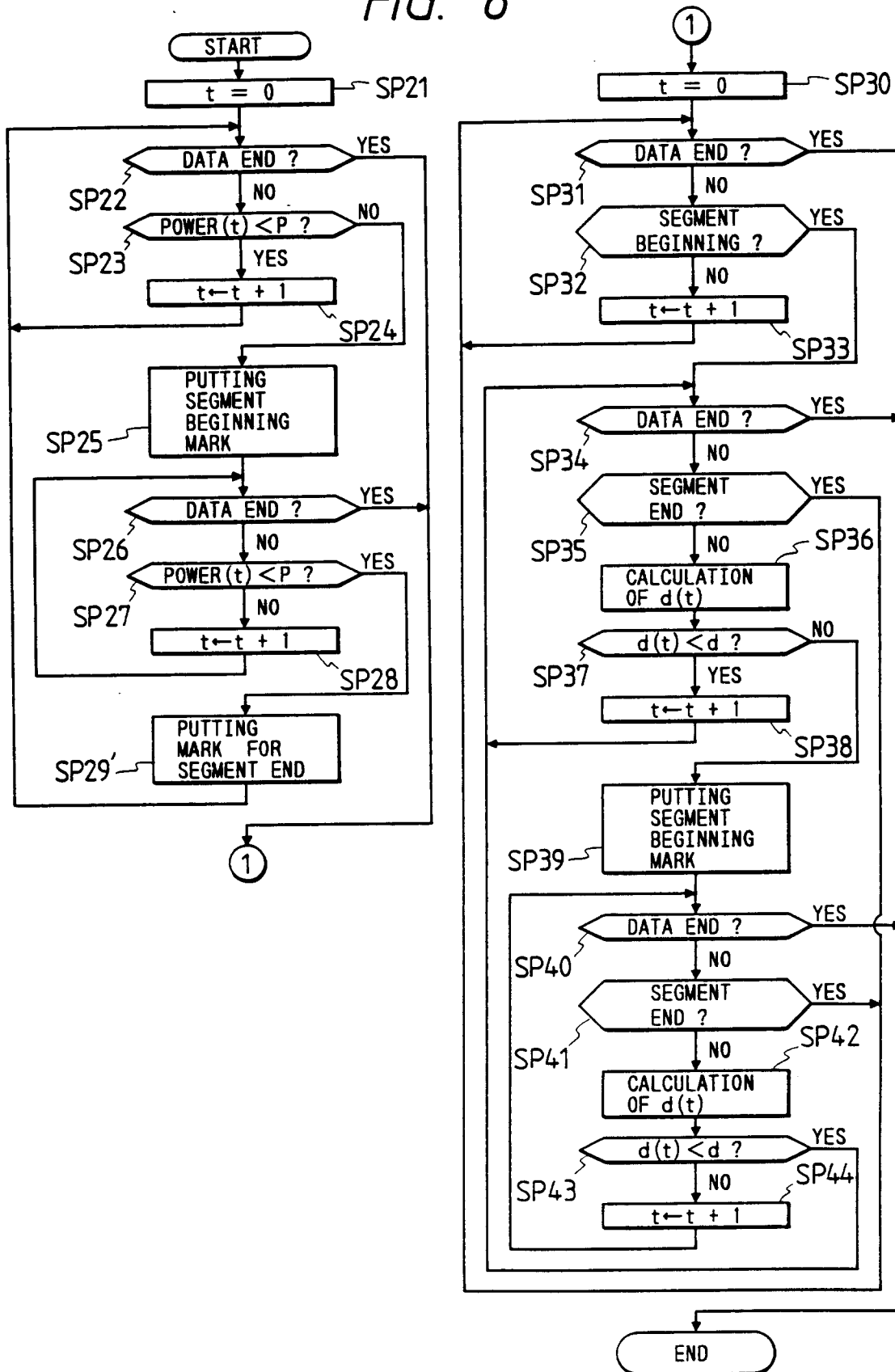


FIG. 9

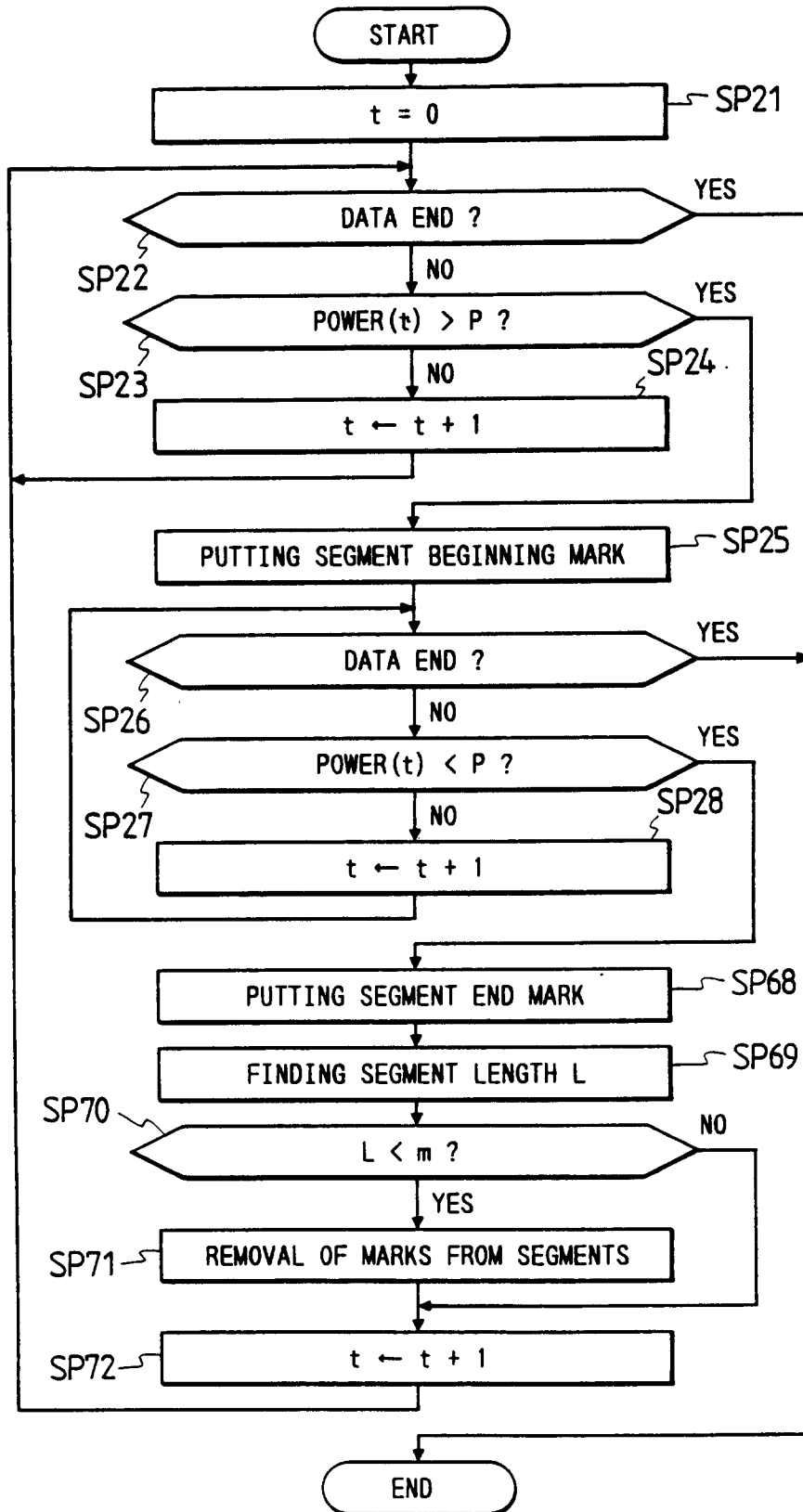


FIG. 10

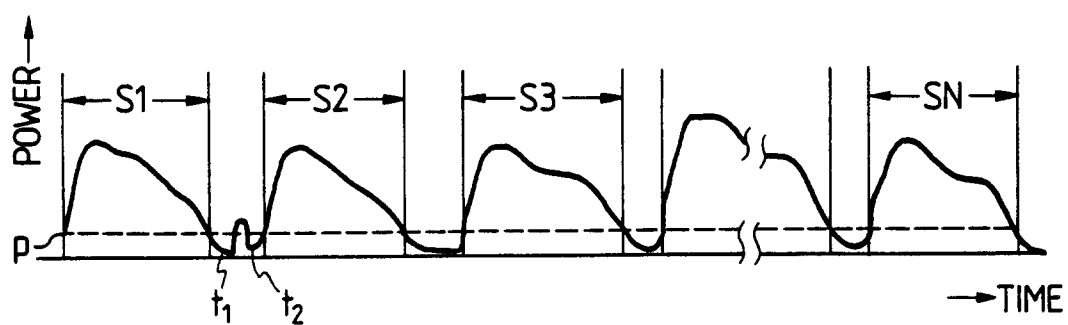


FIG. 12

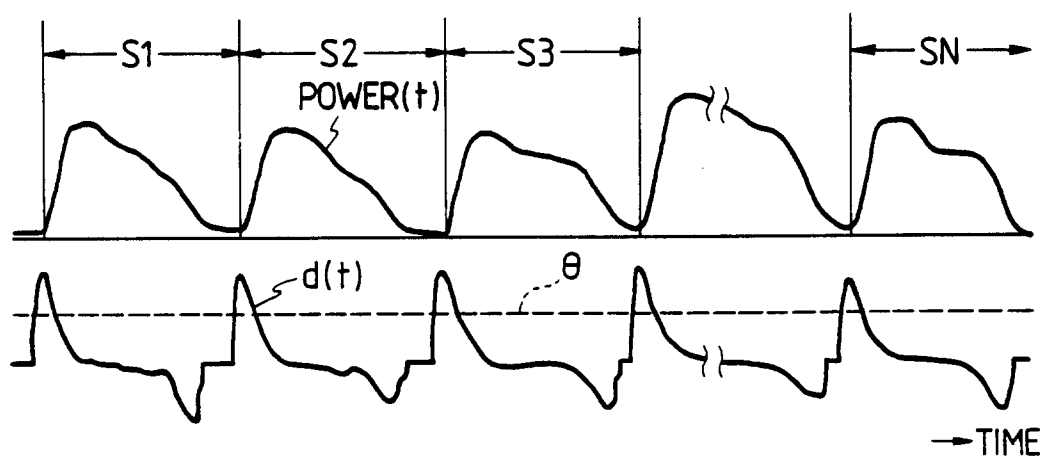


FIG. 11

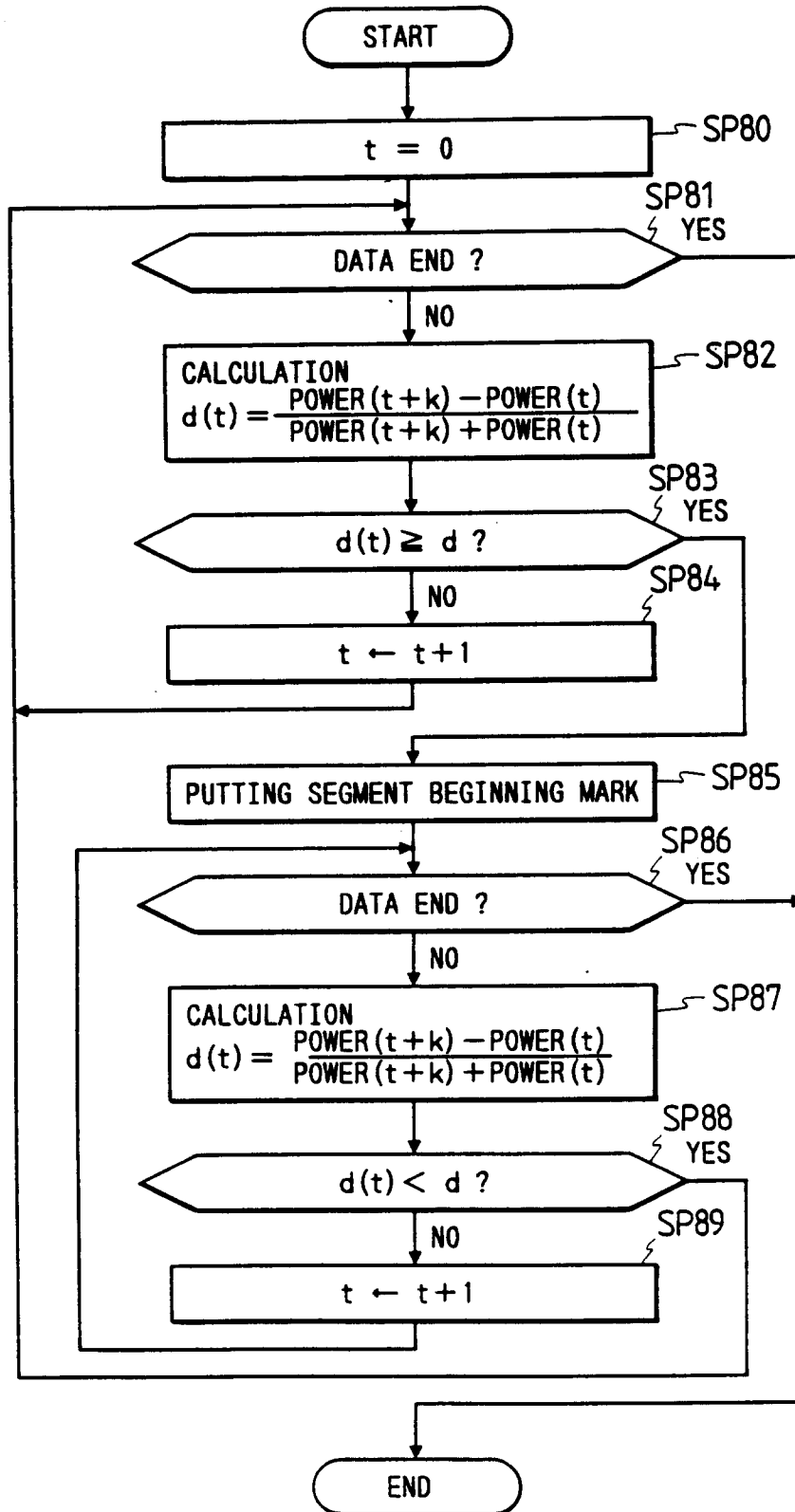


FIG. 13

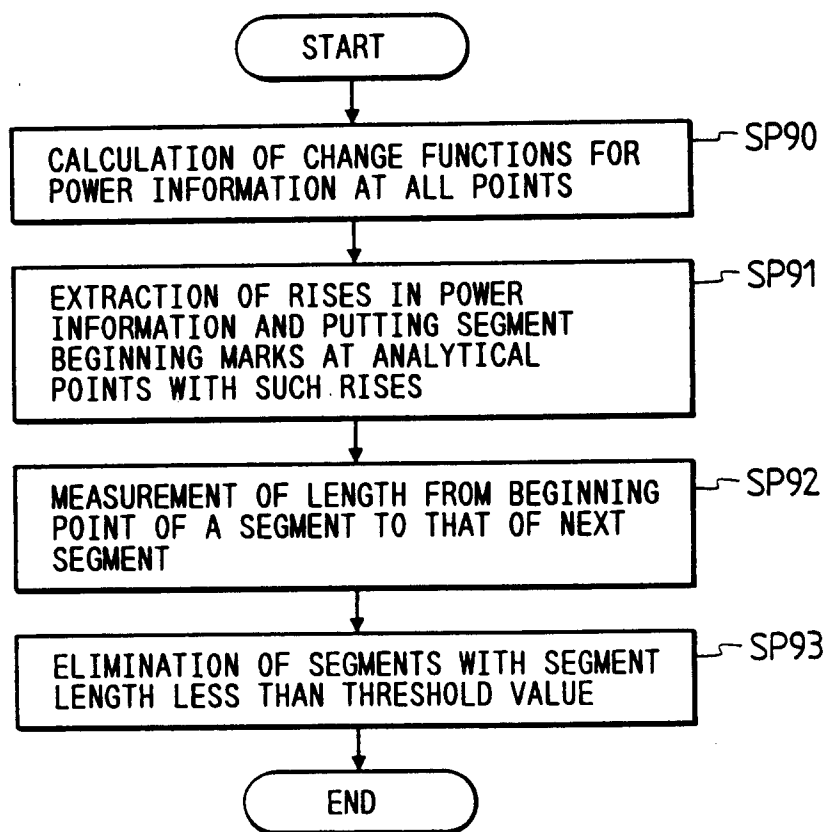


FIG. 14

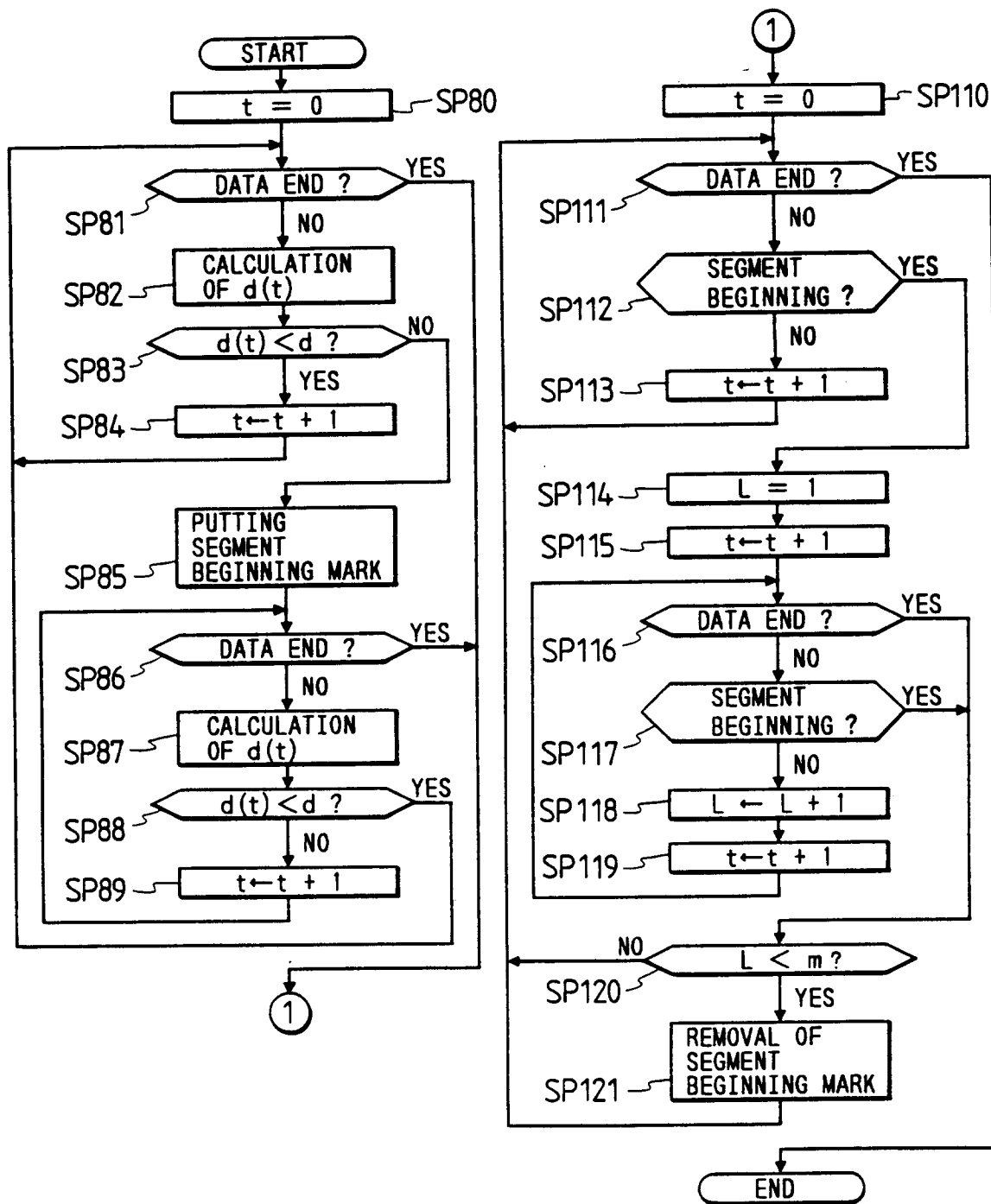


FIG. 15

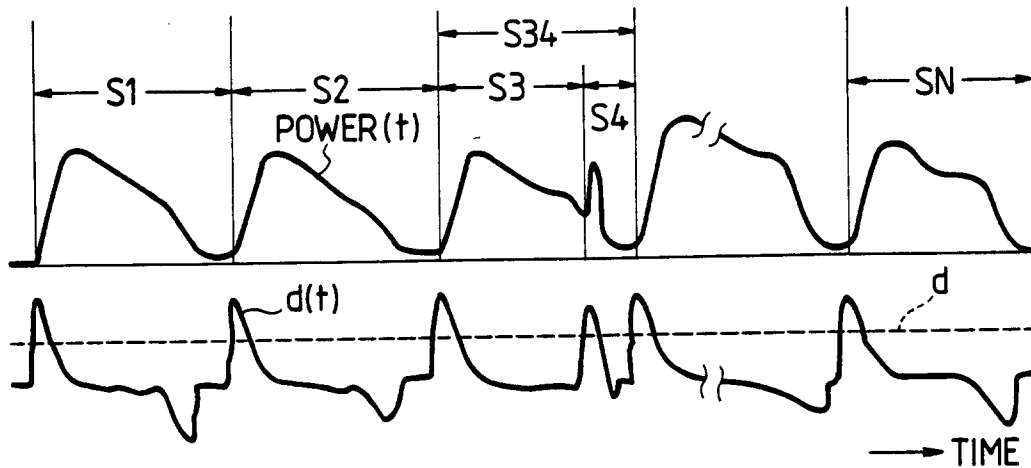


FIG. 16

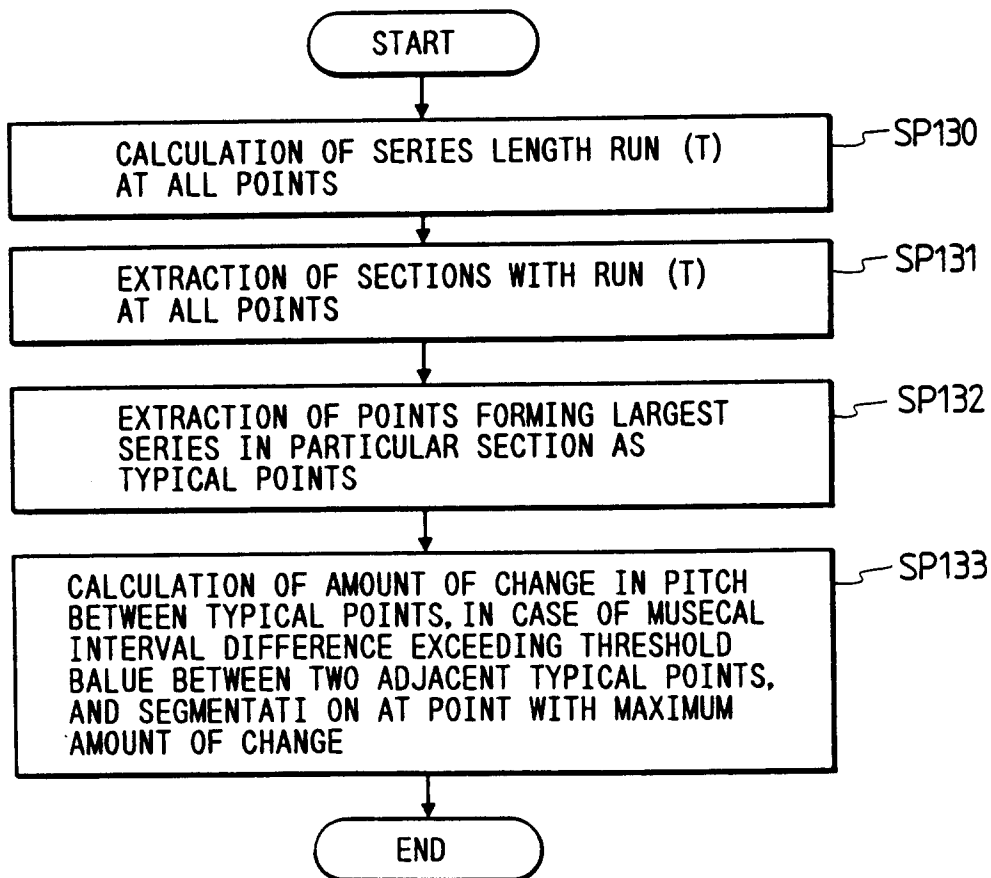


FIG. 17

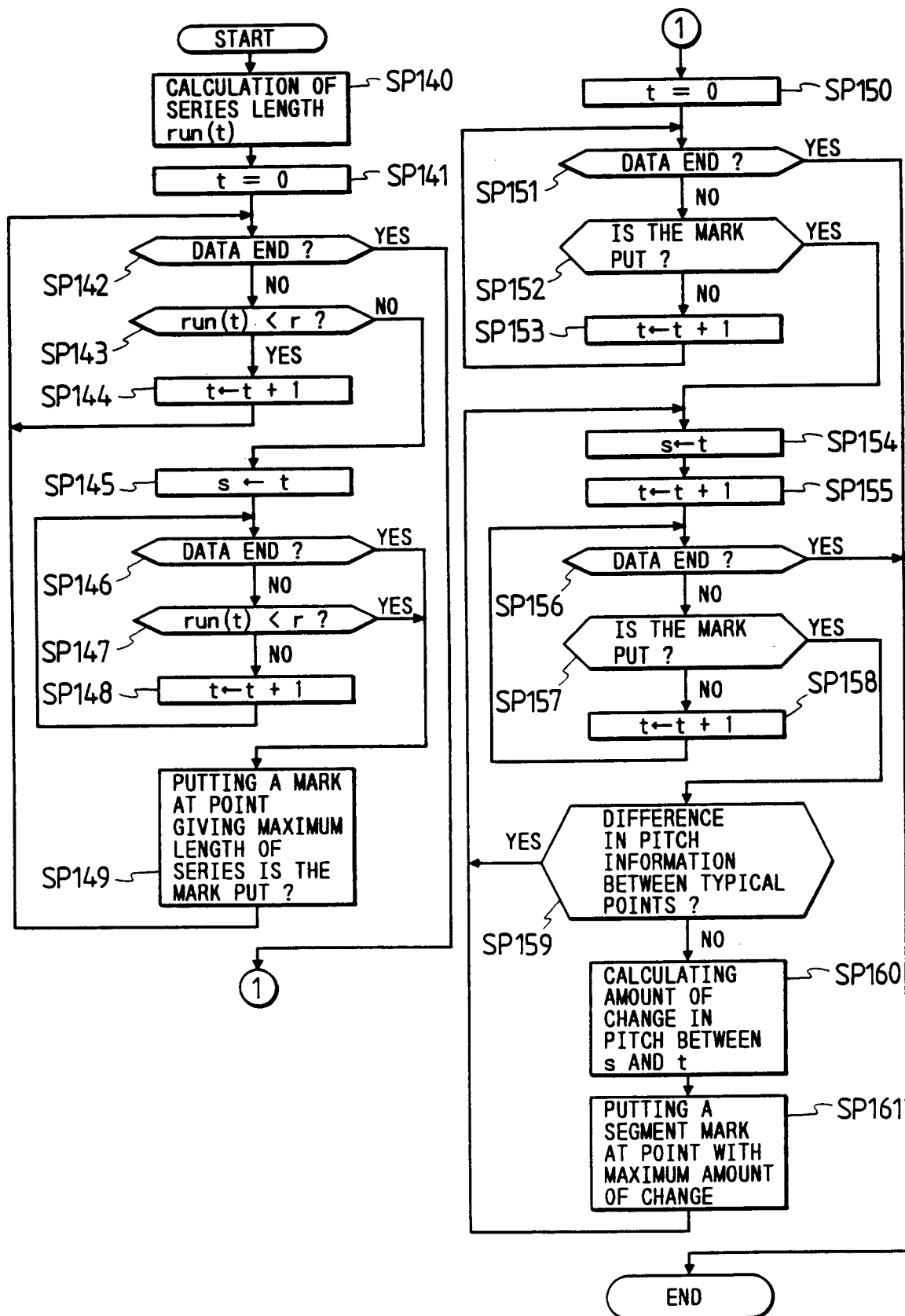


FIG. 19

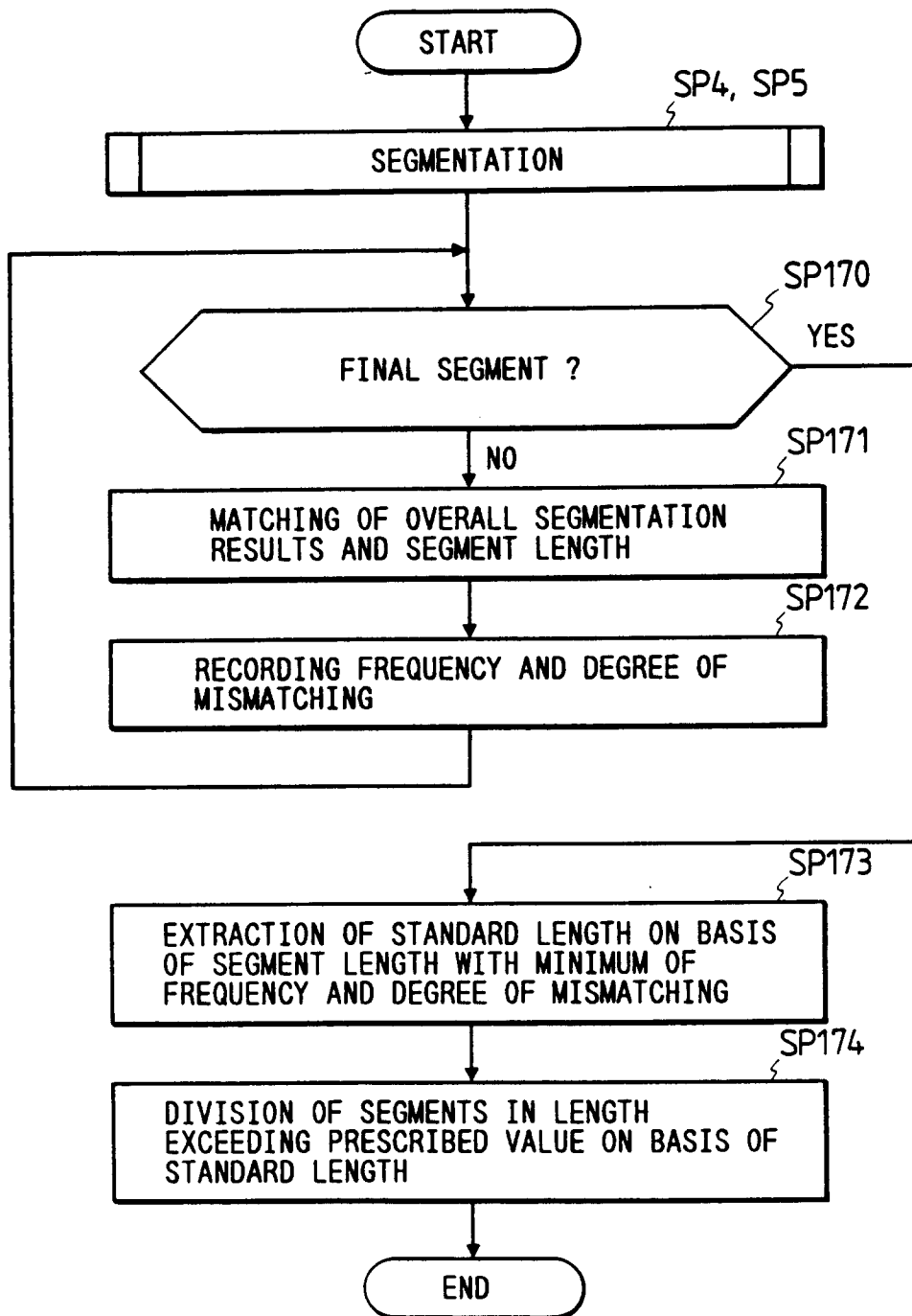


FIG. 18

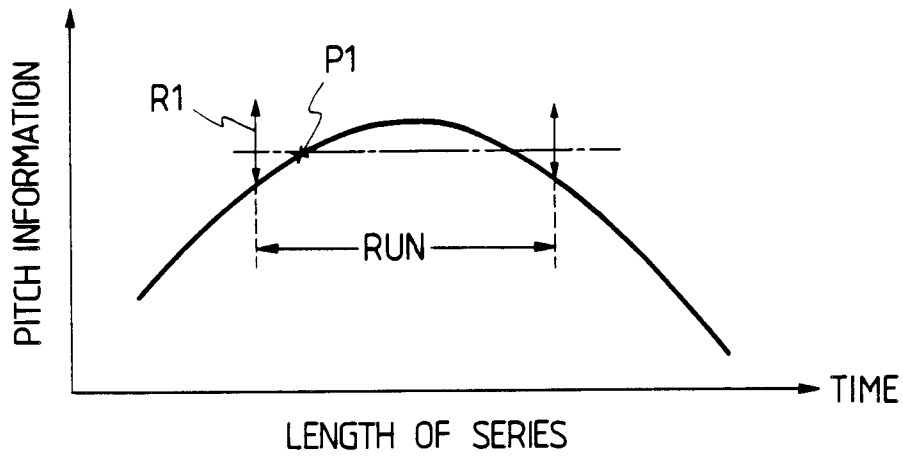


FIG. 20

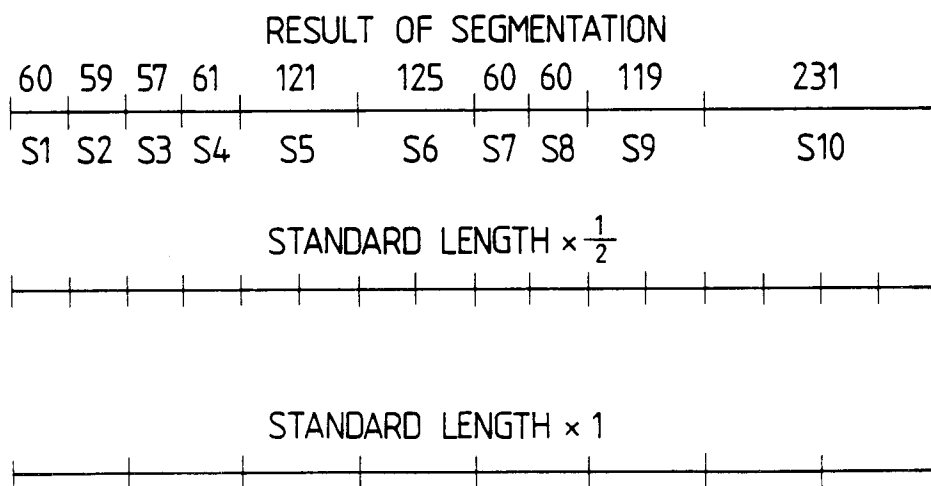


FIG. 21

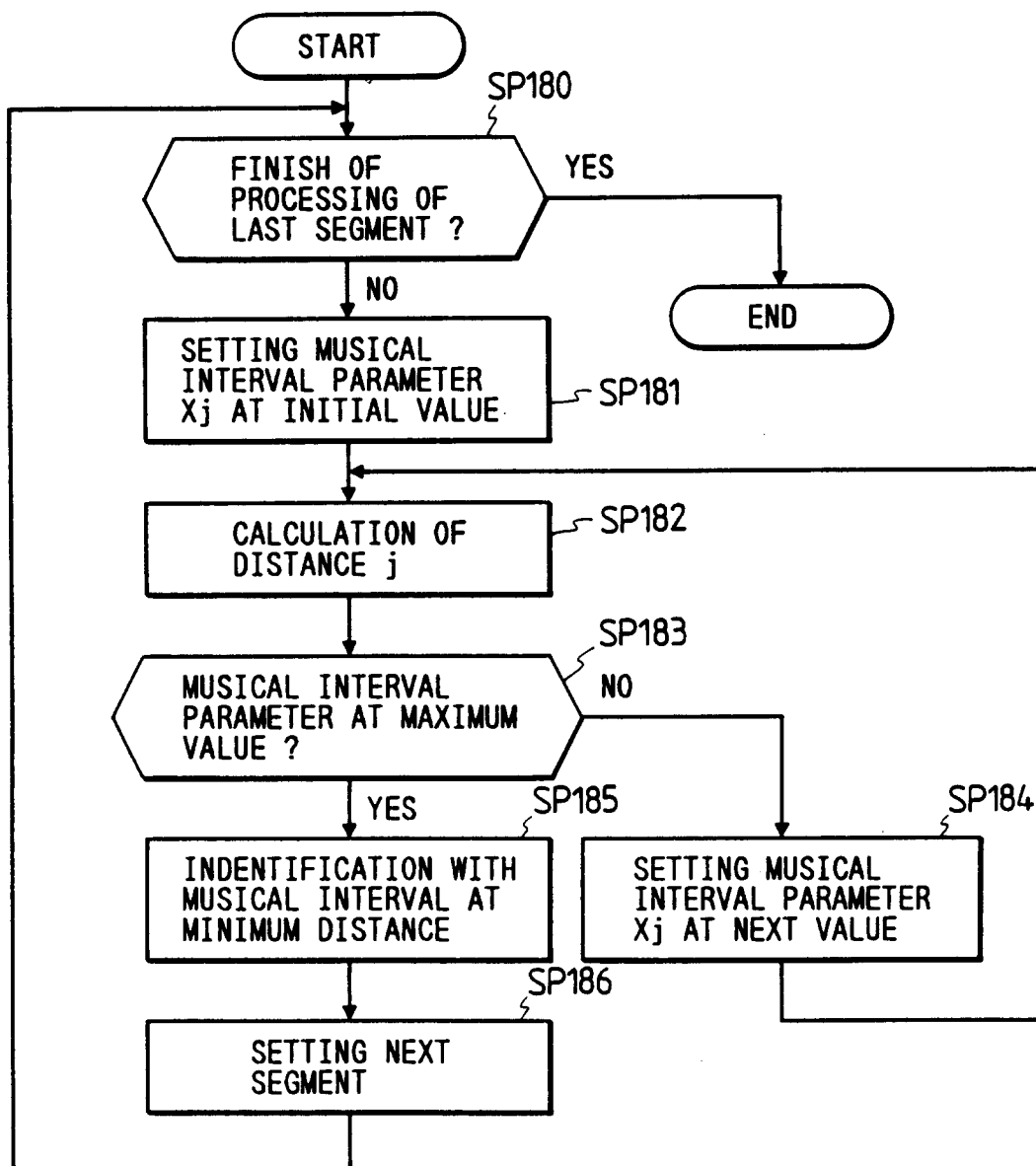


FIG. 22

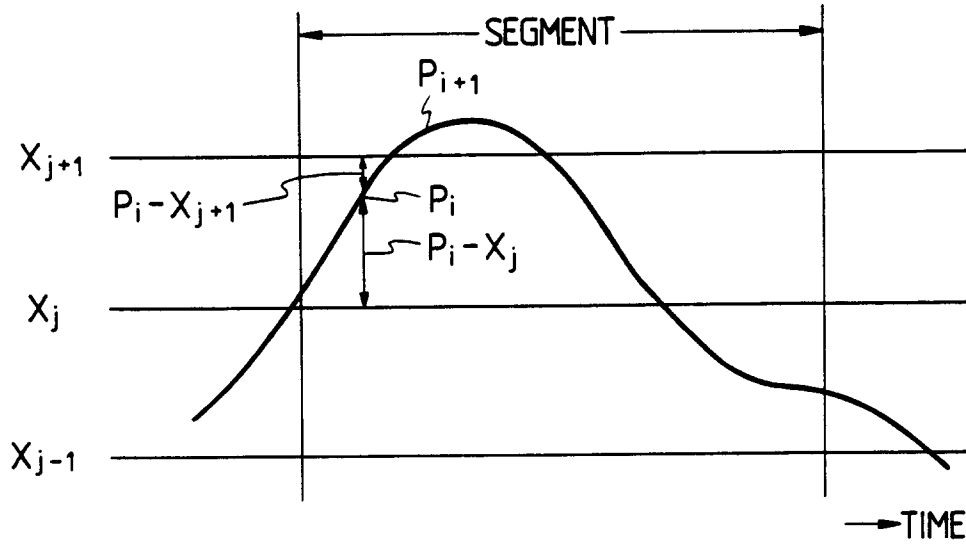


FIG. 24

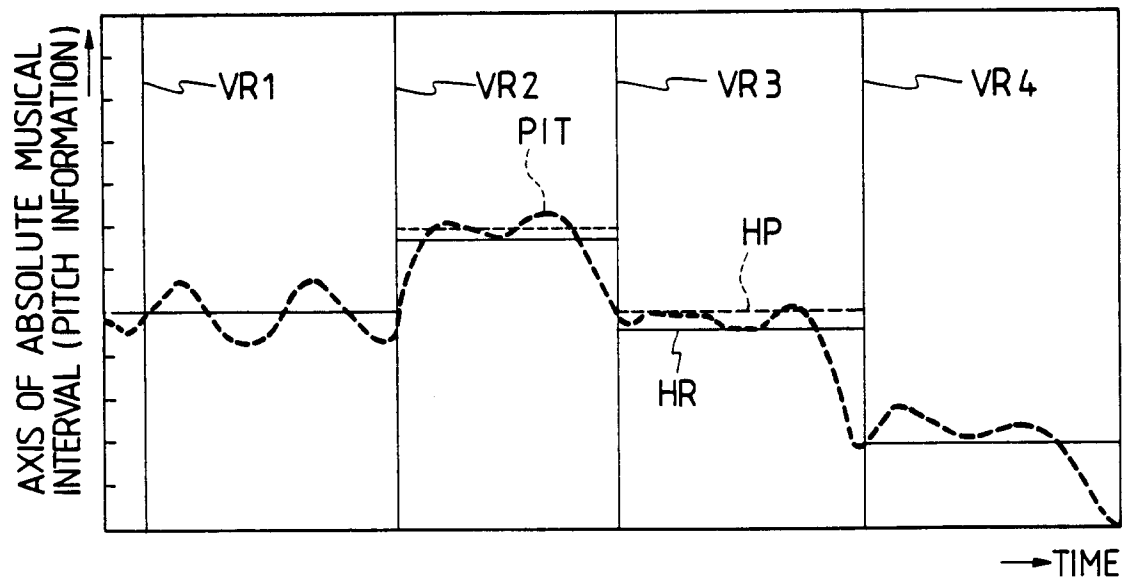


FIG. 23

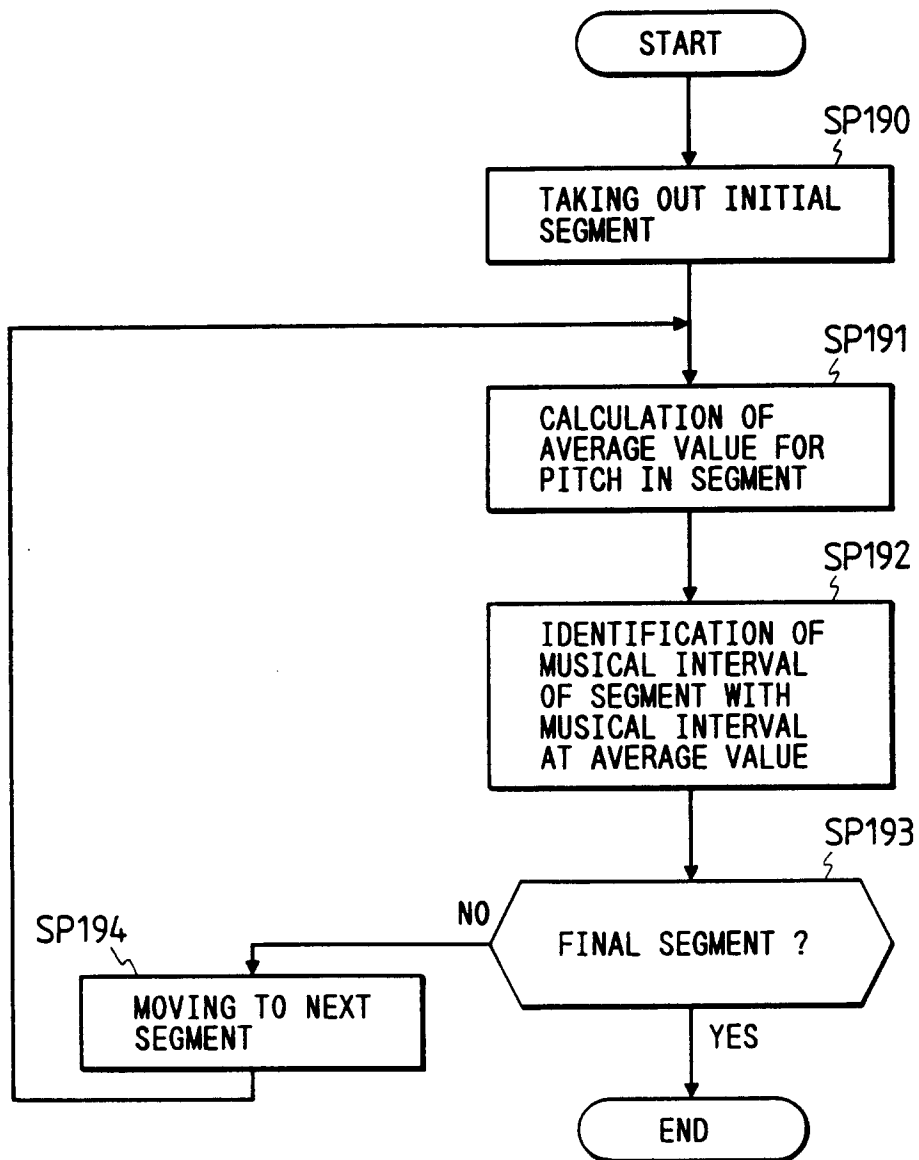


FIG. 25

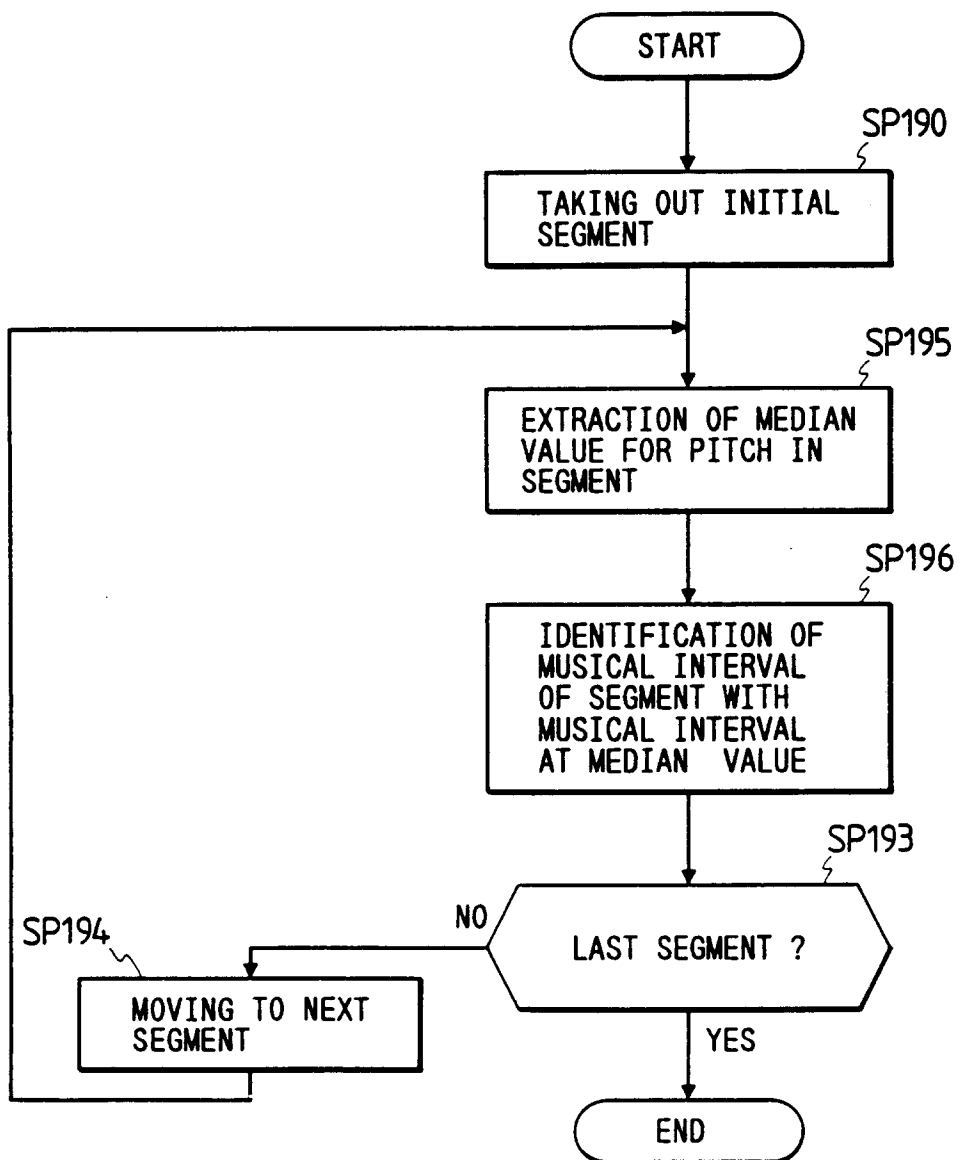


FIG. 26

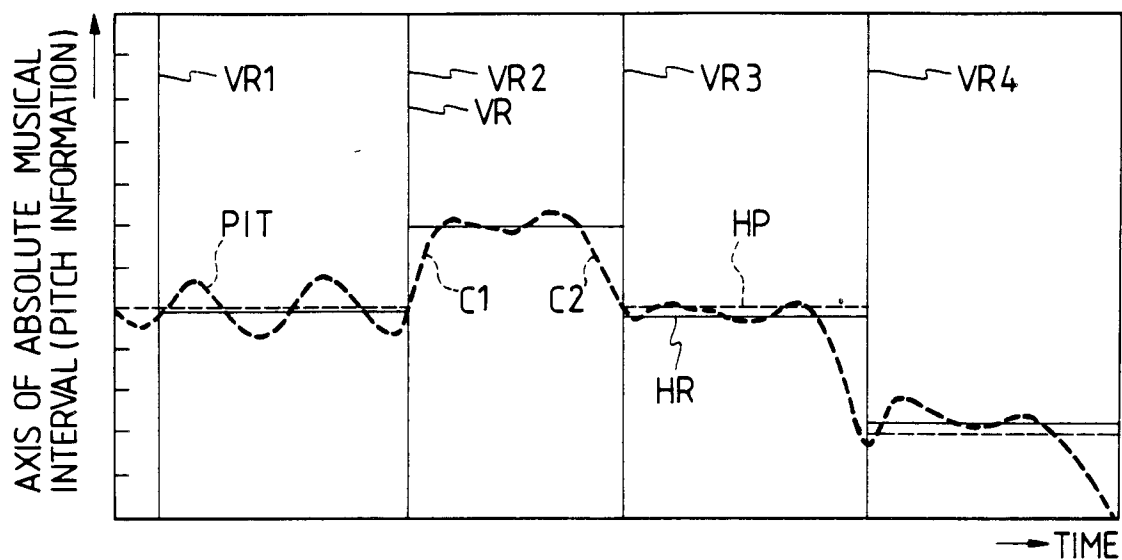


FIG. 28

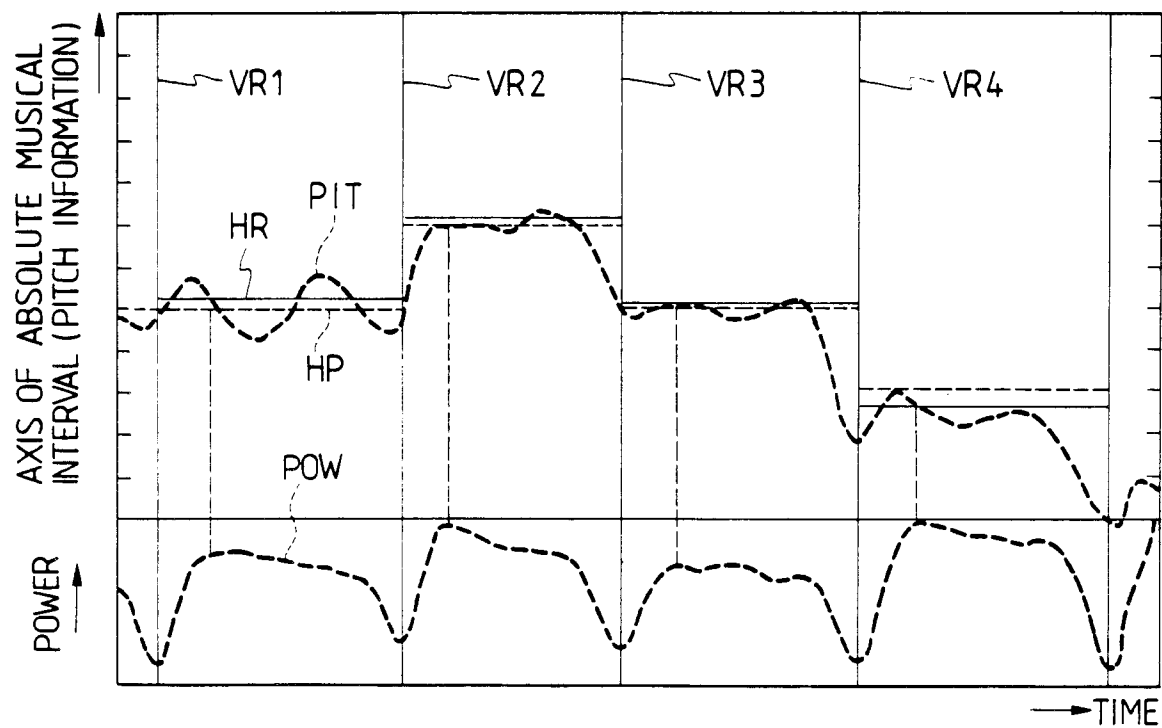


FIG. 27

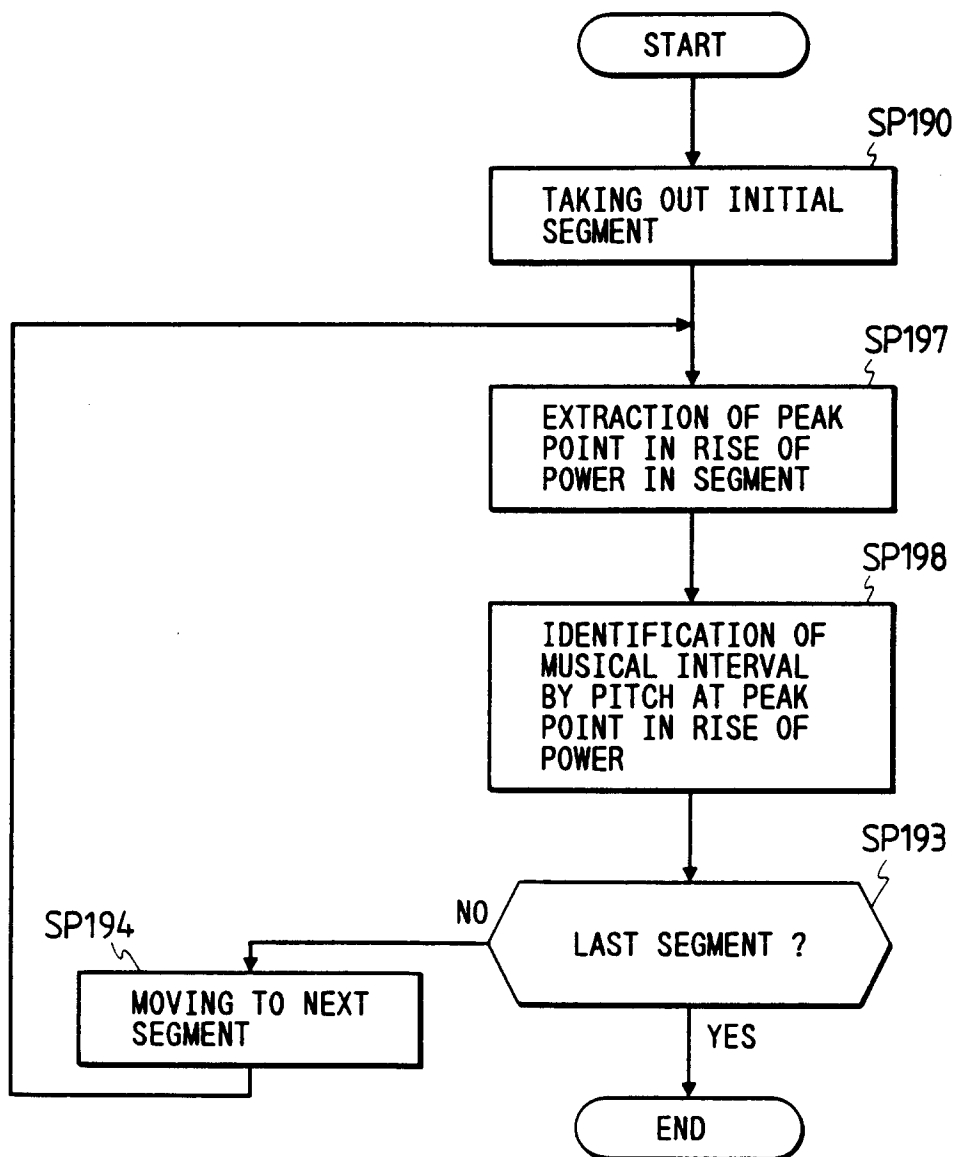


FIG. 29

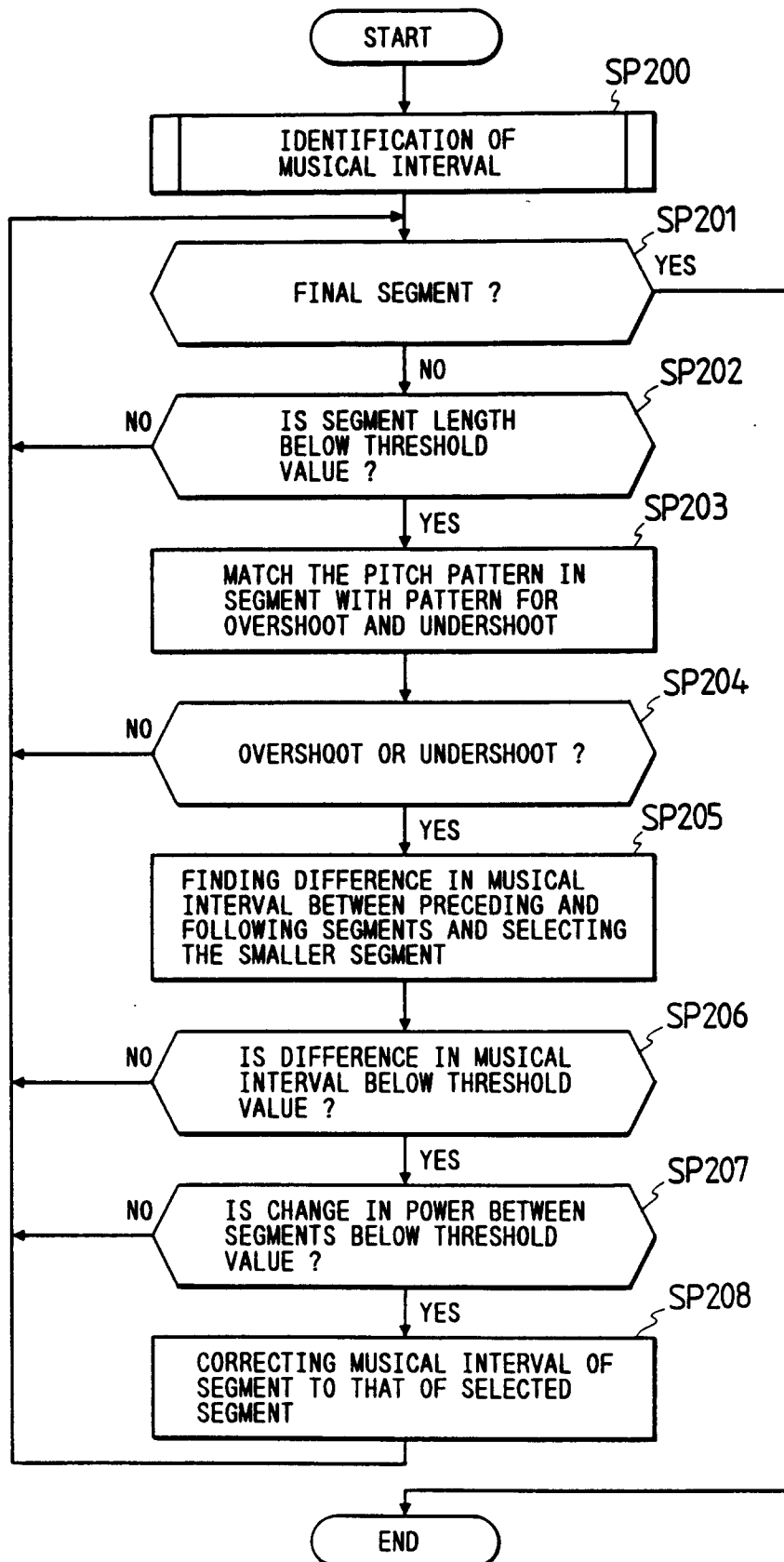


FIG. 31

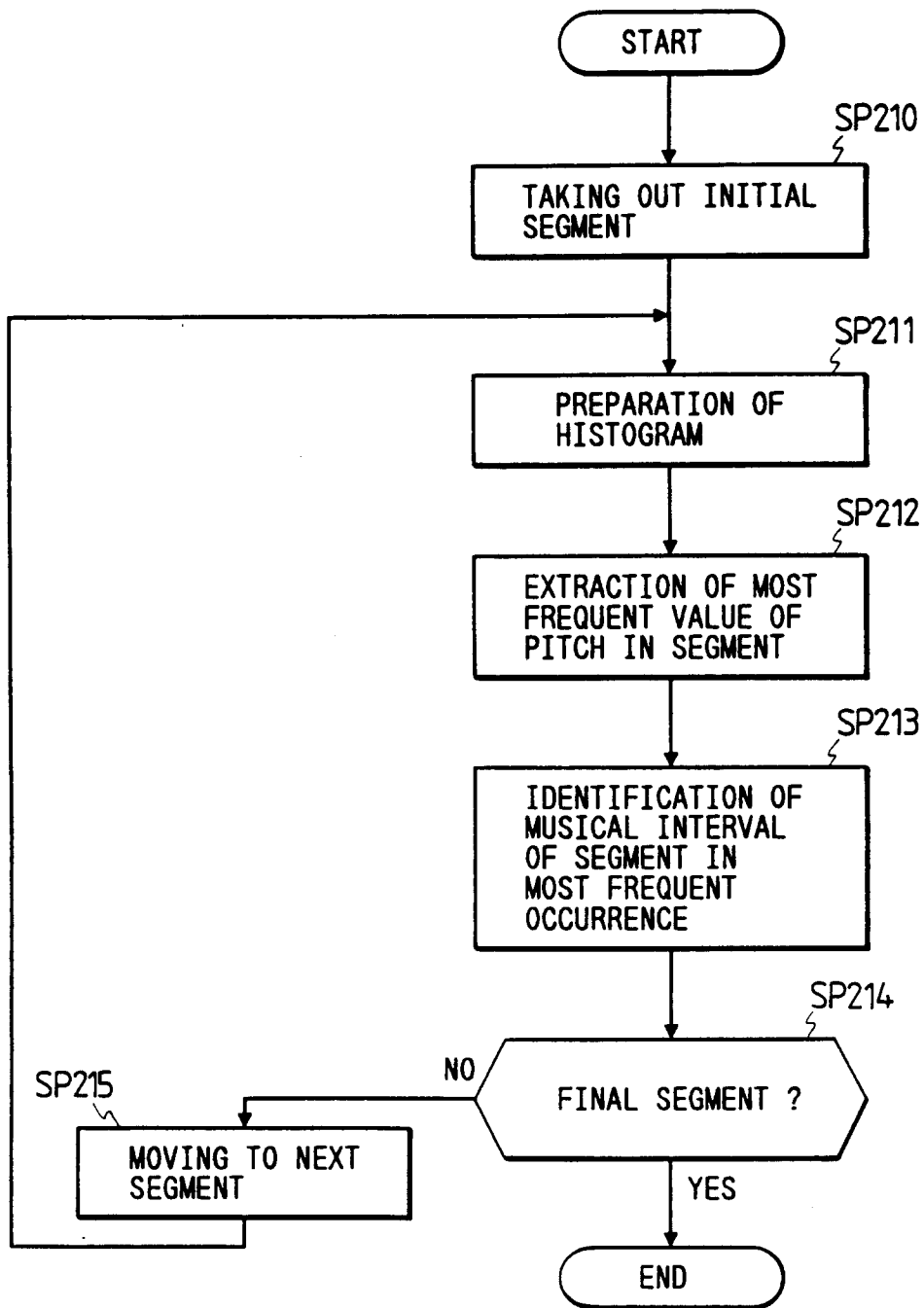


FIG. 30

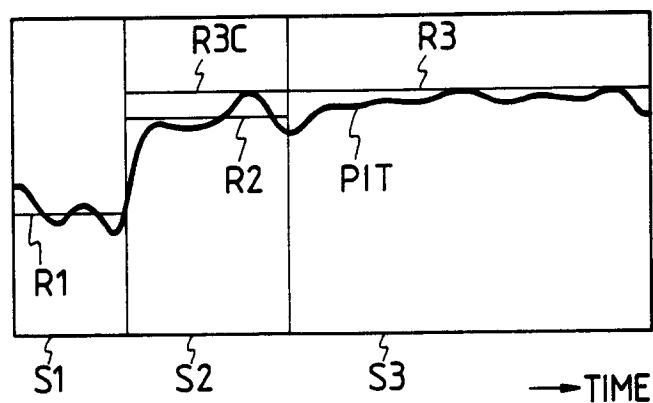


FIG. 32

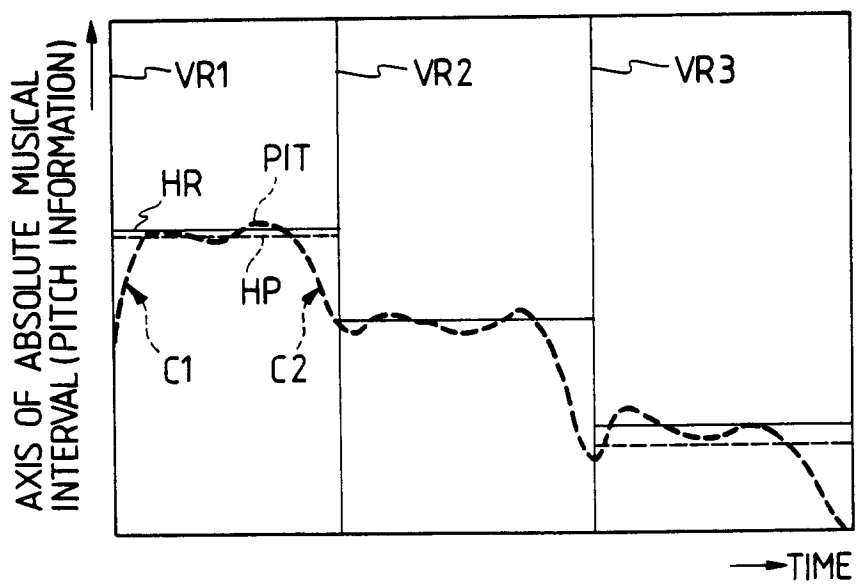


FIG. 33

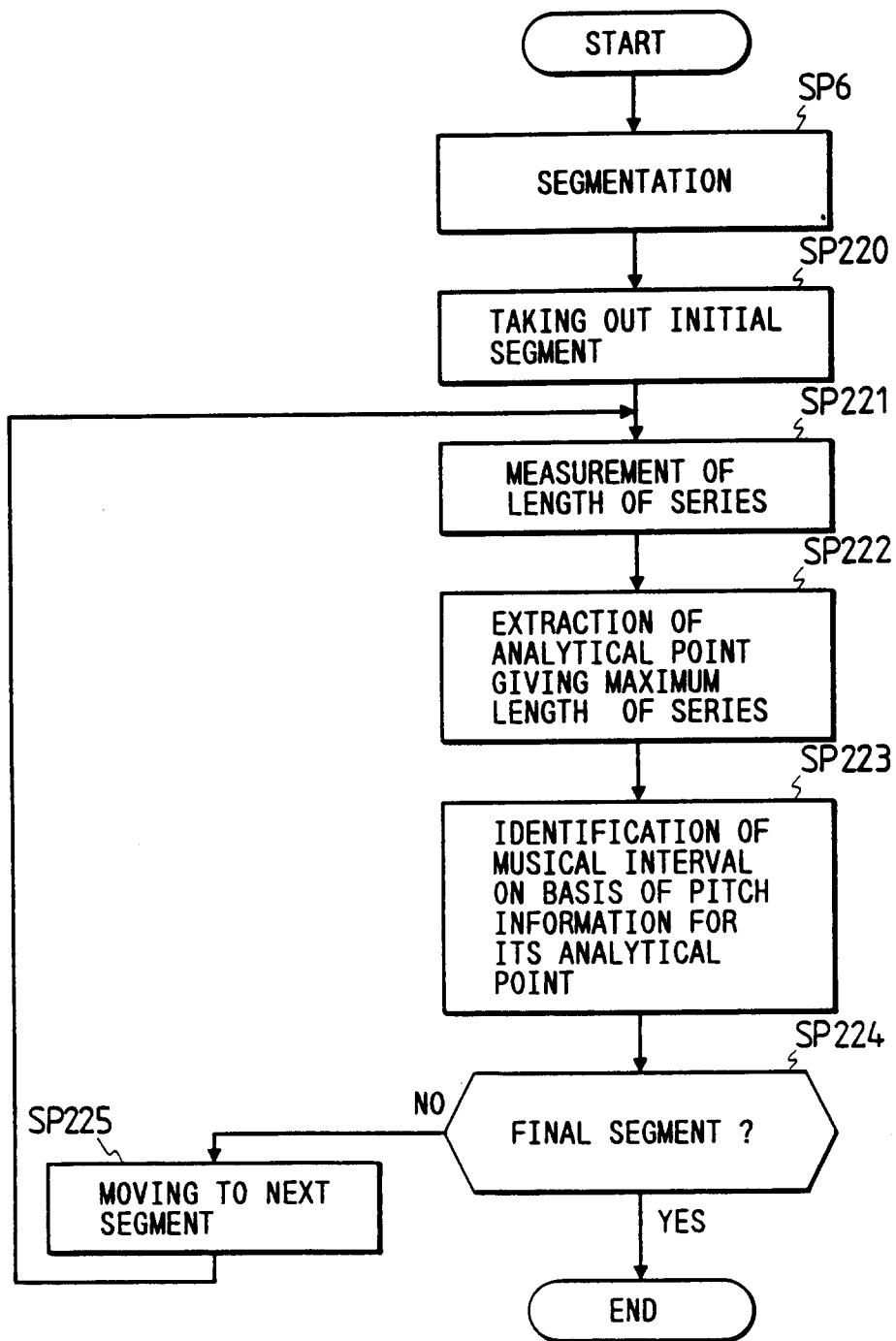


FIG. 34

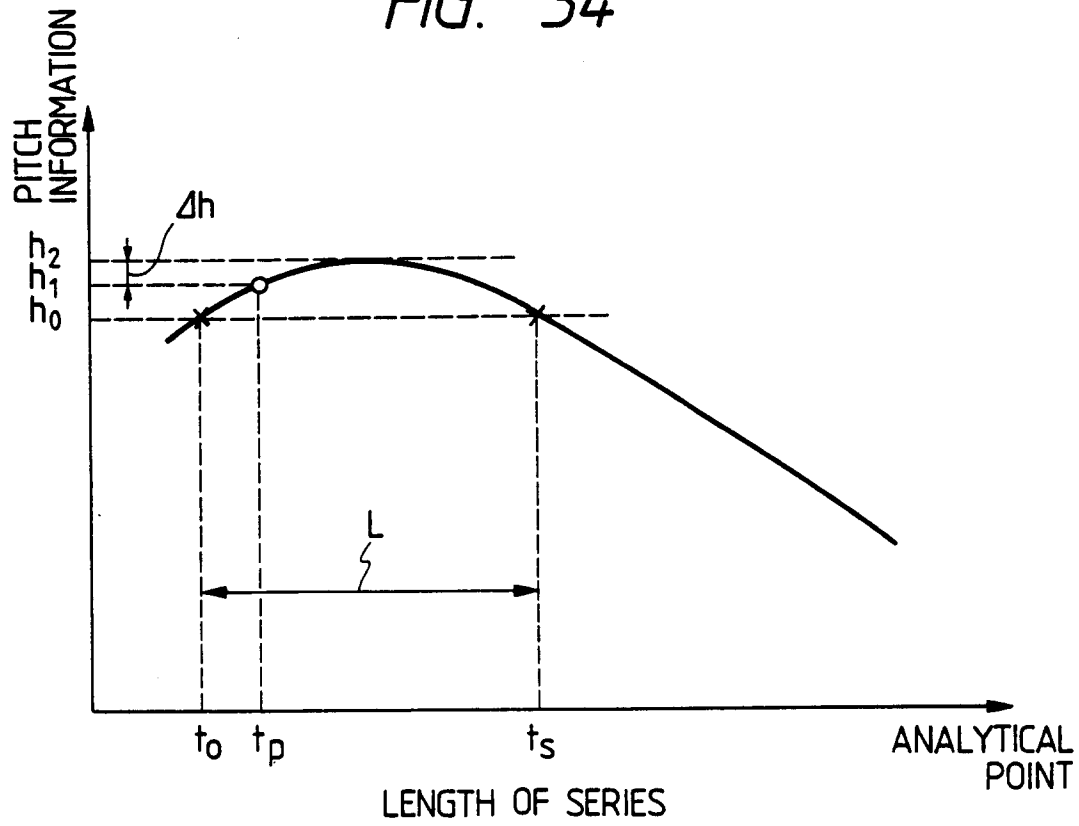


FIG. 35

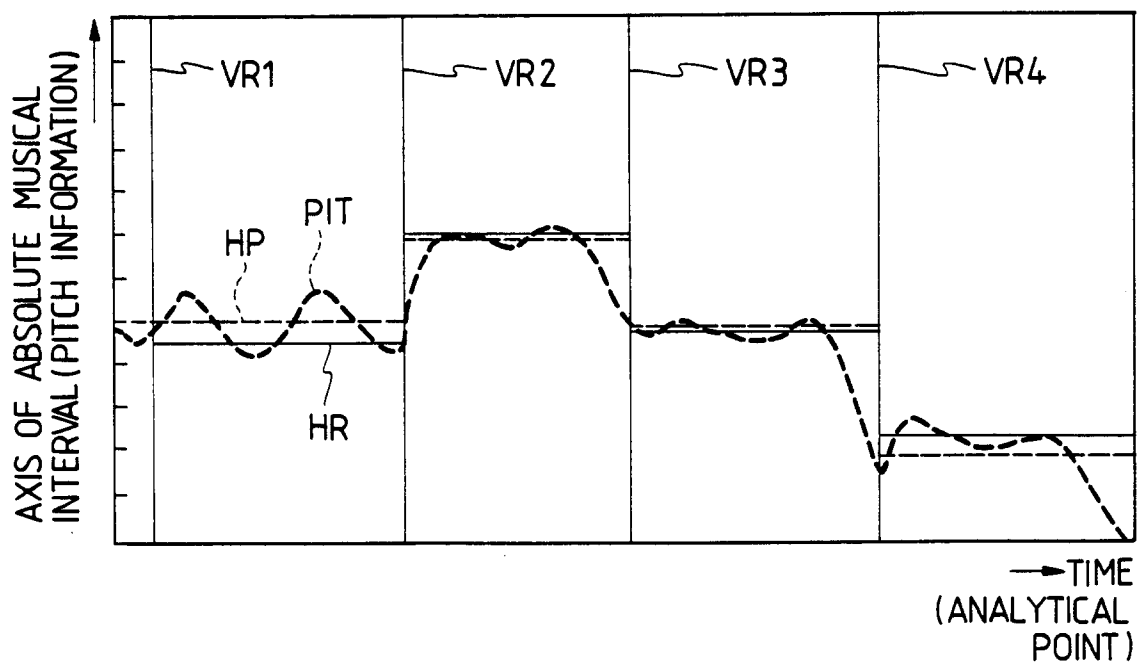


FIG. 36

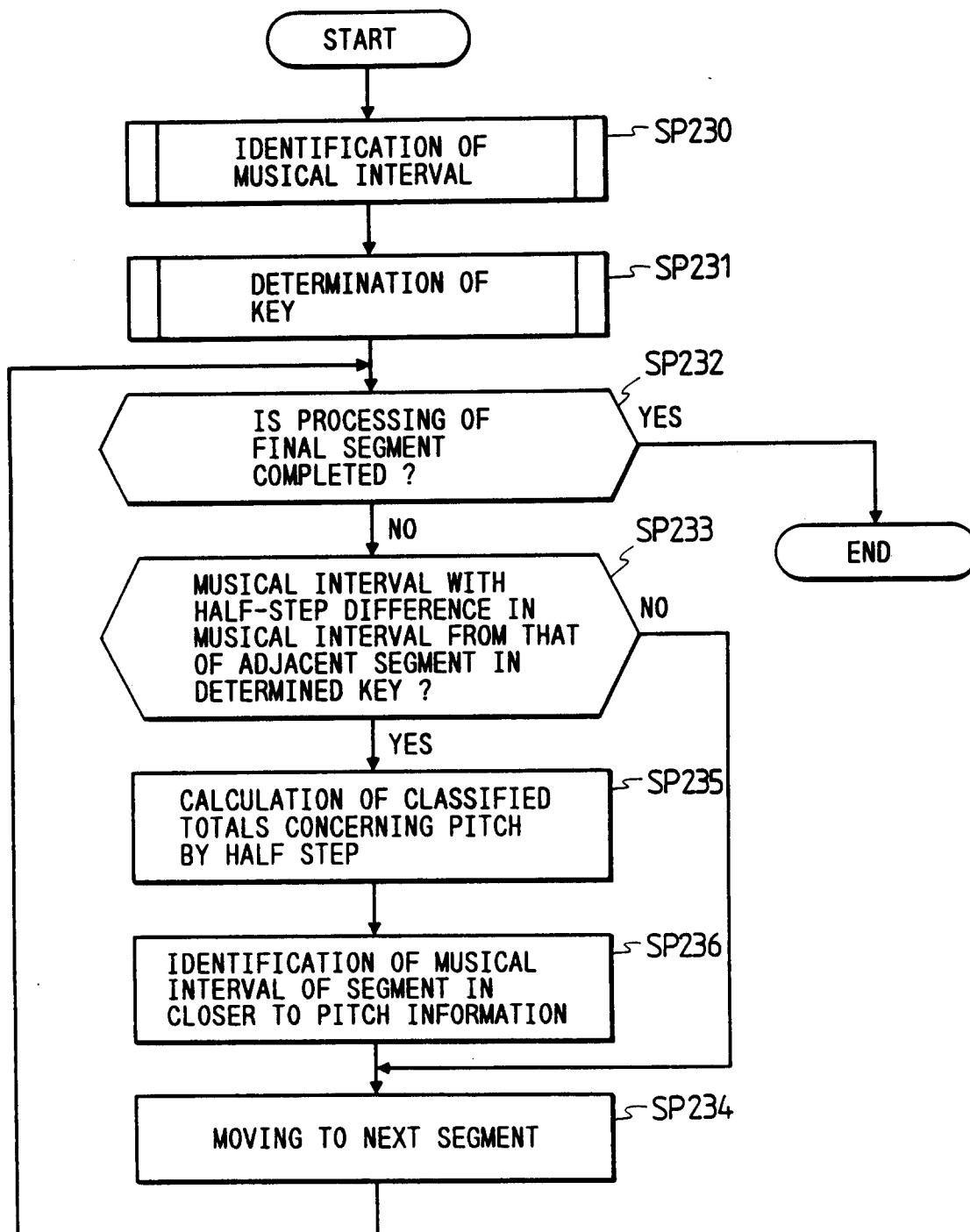


FIG. 37

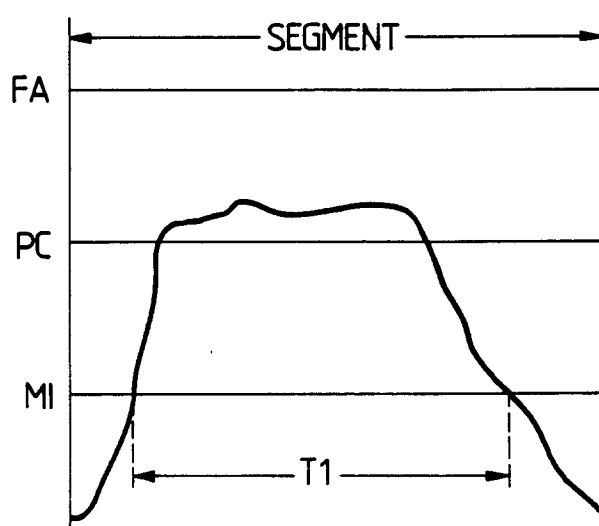


FIG. 38

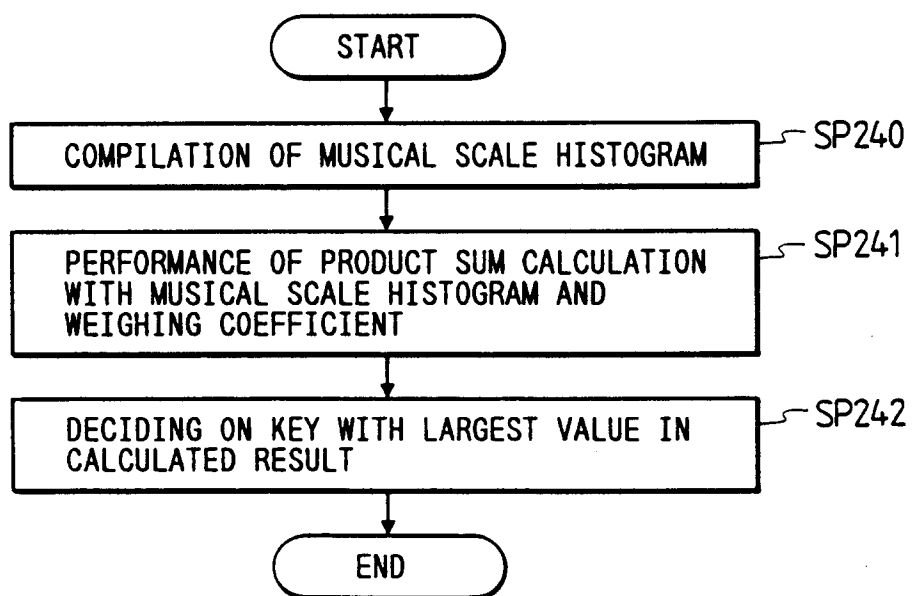


FIG. 39

	C	C SHARP D FLAT	D	D SHARP E FLAT	E	F	F SHARP G FLAT	G	G SHARP A FLAT	A	A SHARP B FLAT	B
COL 1 C MAJOR	2	0	2	0	2	1	0	2	0	2	0	1
COL 2 A MINOR	2	0	1	0	2	2	0	1	0	2	0	2
COL 3 D FLAT MAJOR	1	2	0	2	0	2	1	0	2	0	2	0
COL 4 B FLAT MINOR	2	2	0	1	0	2	2	0	1	0	2	0

FIG. 40

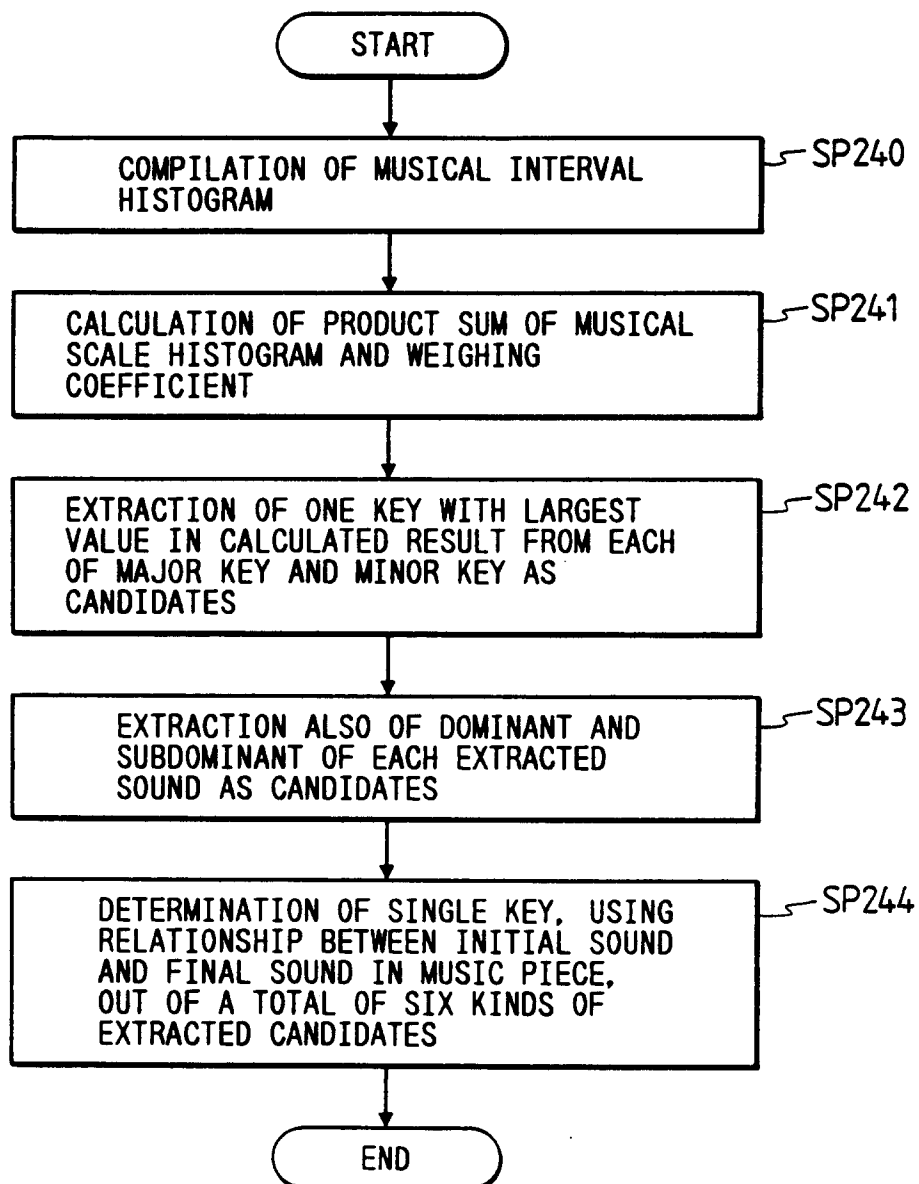


FIG. 41

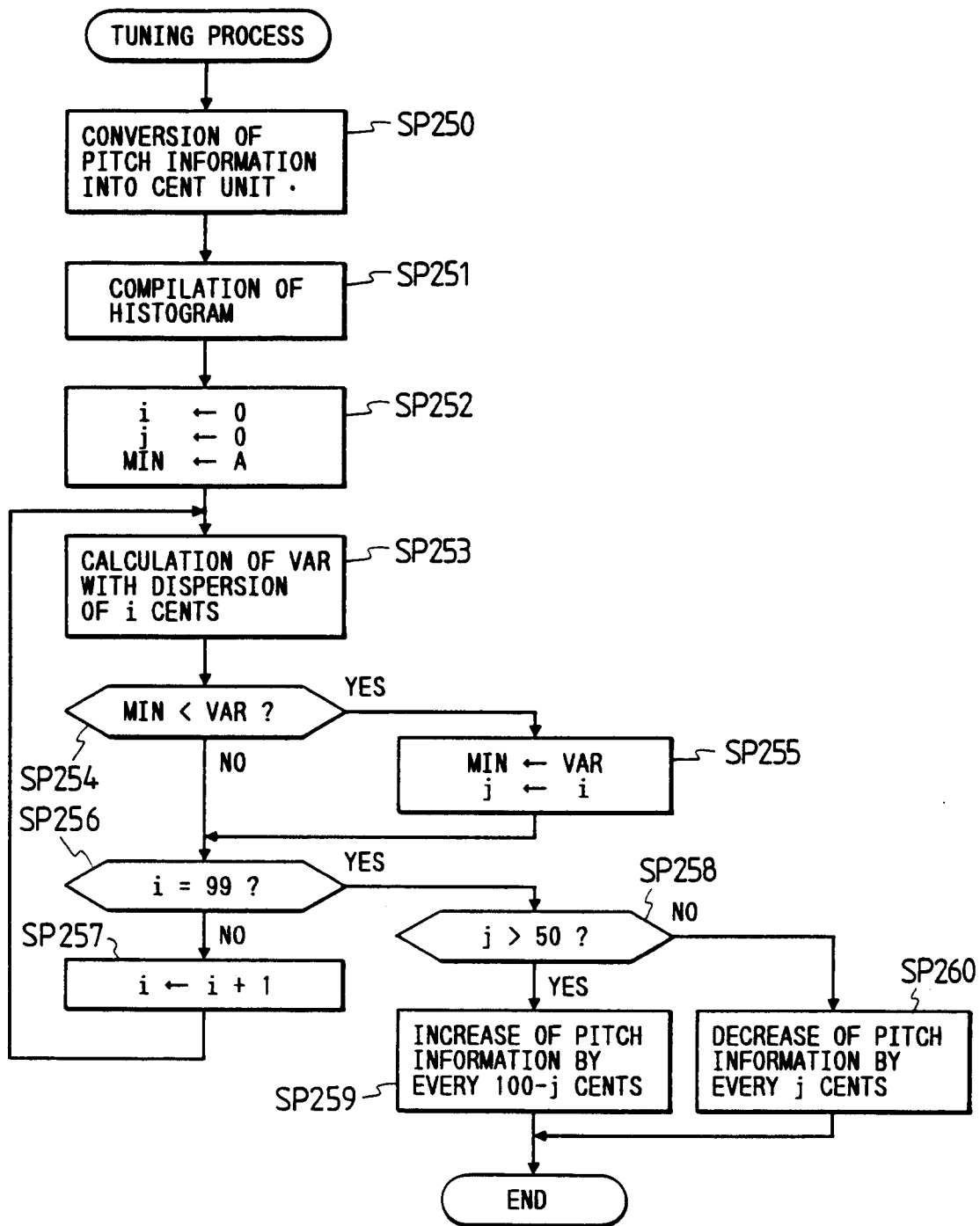


FIG. 42

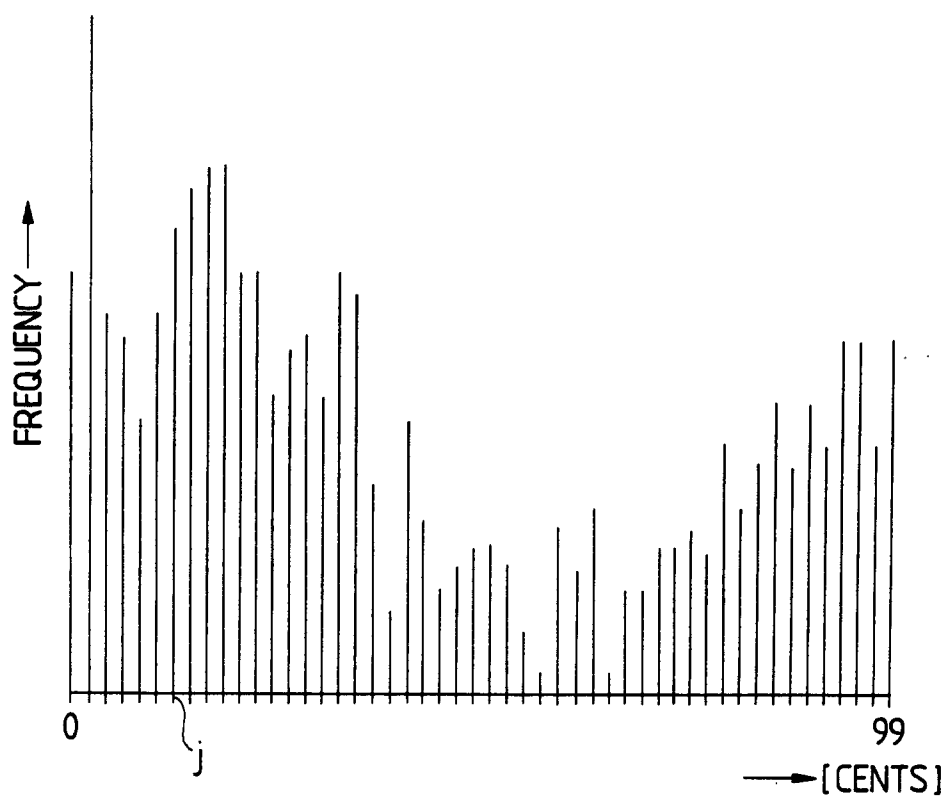


FIG. 44

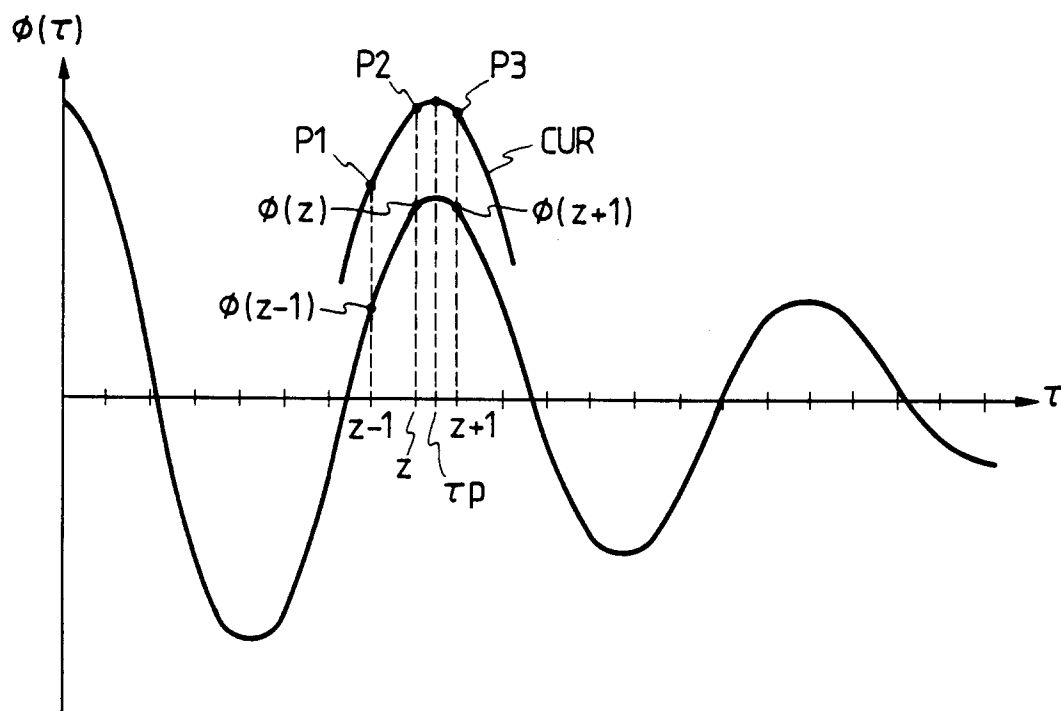


FIG. 43

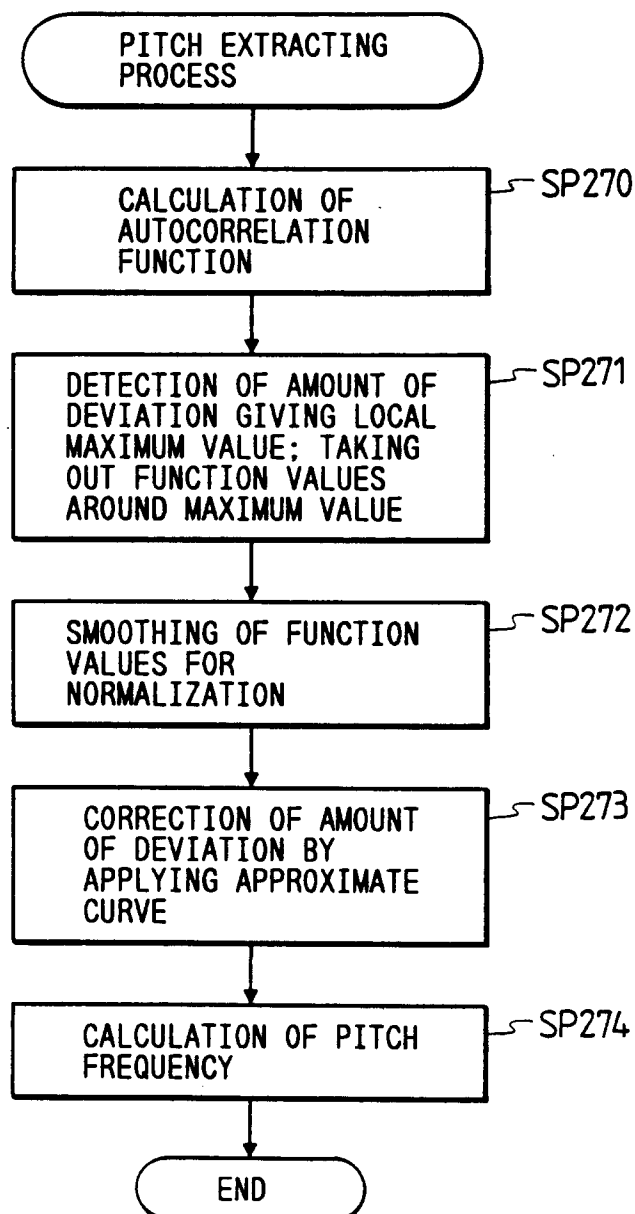


FIG. 45

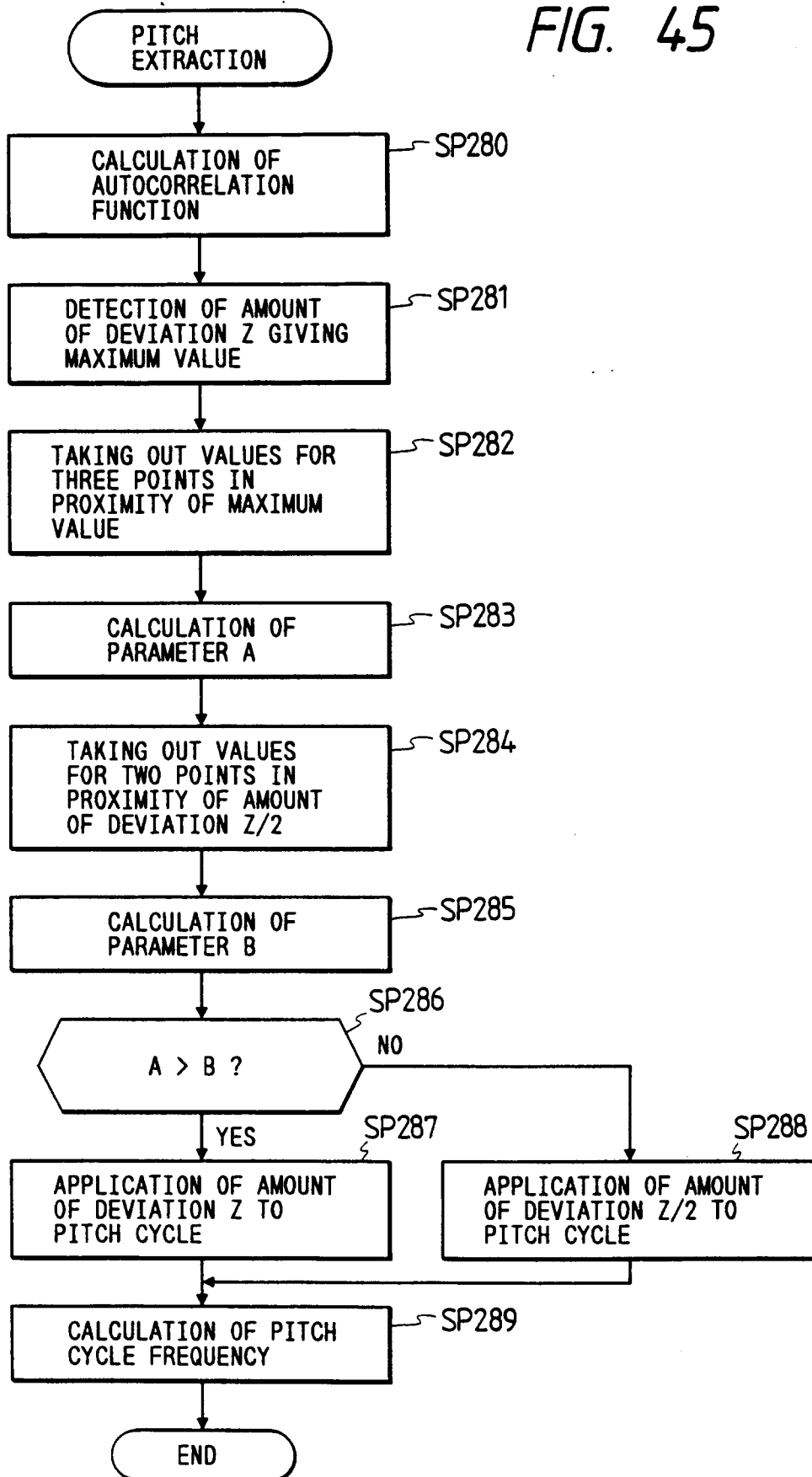
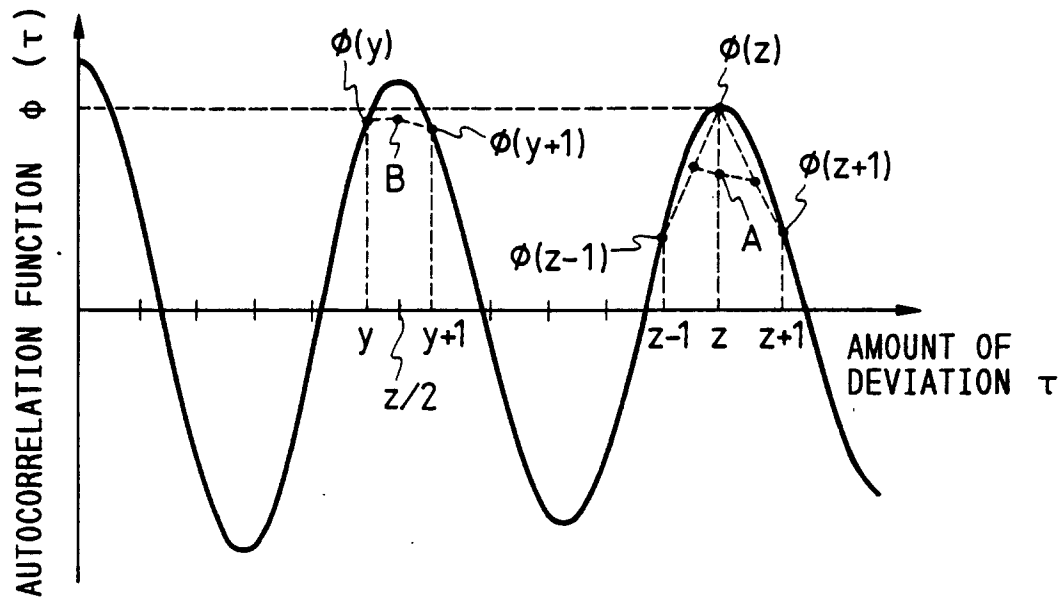


FIG. 46A

WITH DETECTED MAXIMUM VALUE POSITIONED IN
PROXIMITY OF SECOND LOCAL MAXIMUM POINT

**FIG. 46B**

WITH DETECTED MAXIMUM VALUE POSITIONED IN
PROXIMITY OF FIRST LOCAL MAXIMUM POINT

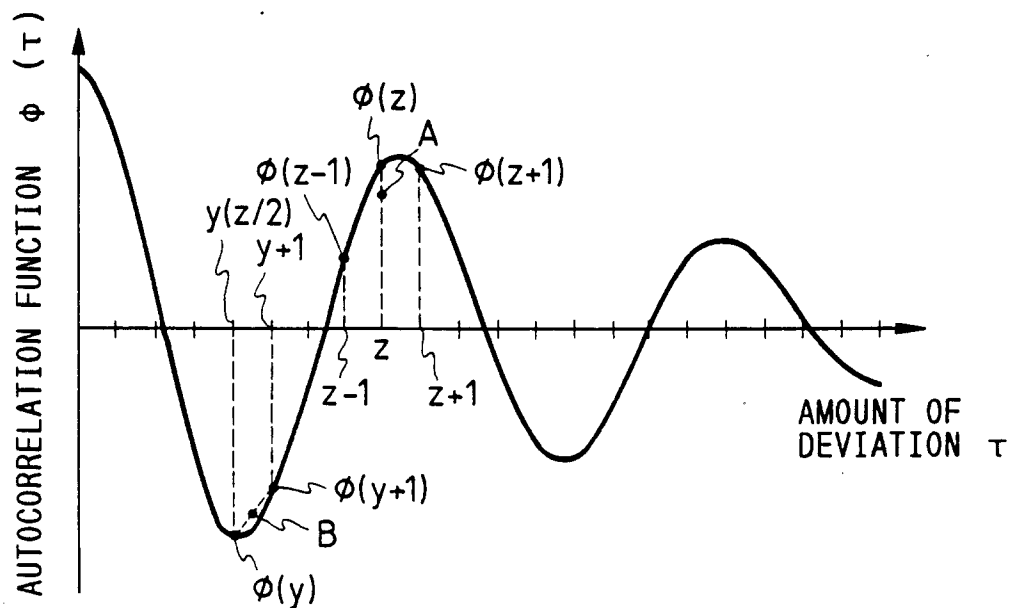


FIG. 47

