(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
3 February 2005 (03.02.2005)

PCT

(10) International Publication Number
WO 2005/010865 A2

(51) International Patent Classification[7]: G10L

(21) International Application Number: PCT/IN2003/000259

(22) International Filing Date: 31 July 2003 (31.07.2003)

(25) Filing Language: English

(26) Publication Language: English

(71) Applicant *(for all designated States except US):* **THE REGISTRAR, INDIAN INSTITUTE OF SCIENCE** [IN/IN]; Sir C.V. Raman Avenue, Karnataka State, 560012. Bangalore (IN).

(72) Inventors; and

(75) Inventors/Applicants *(for US only):* **VENKATASUBRAMANIAN, Viraraghavan** [IN/IN]; c/o Indian Institute of Science, Sir C.V. Raman Avenue, Karnataka State, 560 012. Bangalore (IN). **RAMAKRISHNAN, Ramaswami, Kalpathi** [IN/IN]; c/o Indian Institute of Science, Sir C.V. Raman Avenue, Karnataka State, 560 012. Bangalore (IN).

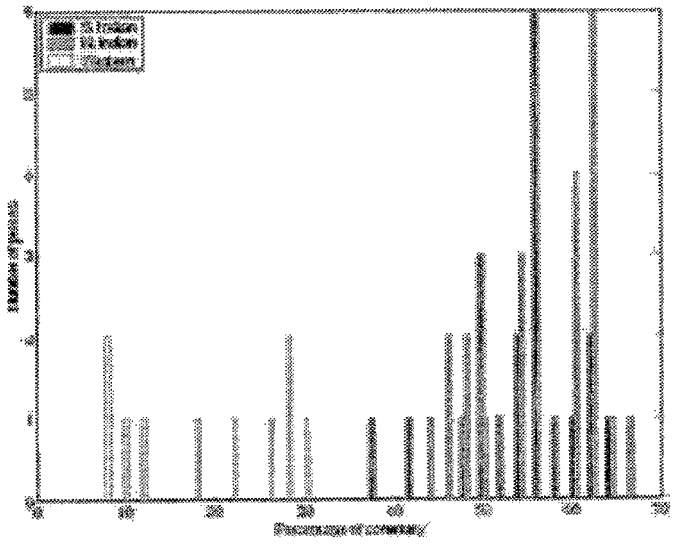
(74) Agent: **VAIDYANATHAN, Alamelu;** 451, 2nd Cross,, 3rd Block, 3rd Stage, Basaveshwaranagar,Karnataka State, 560 079 Bangalore (IN).

(81) Designated States *(national):* AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, *HR*, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, **NI,** NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional):* ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, **NL,** PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, *MR*, **NE,** SN, TD, TG).

[Continued on next page]

(54) Title: METHOD OF MUSIC INFORMATION RETRIEVAL AND CLASSIFICATION USING CONTINUITY INFORMATION



(57) Abstract: This invention relates to a method of semi-automatic or automatic music information retrieval and classification using continuity information and a method for obtaining the continuity information by finding the discontinuous transitions in the pitch curve of a musical piece and using energy considerations. The music information retrieval comprises retrieving desired musical pieces from a database of musical pieces which includes, but not limited to, private databases, archives, internet, etc., consisting of musical data stored as audio, audio tracks in multimedia data, musical notation, and other representations of music, including encoded representations; or embedded within other kinds of audio data including, but not limited to, audio effects, conversational speech, lectures, background music etc.

WO 2005/010865 A2

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, UZ, VC, VN, YU, ZA, ZM, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BE BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)*

— *of inventorship (Rule 4.17(iv)) for US only*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

METHOD OF MUSIC INFORMATION RETRIEVAL AND CLASSIFICATION USING CONTINUITY INFORMATION.

This invention relates to a method of semi-automatic or automatic music information retrieval and classification using Continuity information and a method for obtaining the continuity information by finding the discontinuous transitions in the pitch curve of a musical piece and using energy considerations. Though the present invention suggests the use of continuity information in music information retrieval and Classification, the said information could be used in music analysis, for e.g. raga identification, improved music notation and transmission etc.

Present State of the Art

a Music Signal Processing

Analysis of (usually digitized) audio signals using digital signal processing techniques is a popular technique in music analysis. An important problem in Musical Signal Processing is the Pitch Tracking *Problem*, a problem whose complexities increase manifold when complex (but red life) signals me considered. For example, the background music of a film or an orchestra can consist of several instruments and voices simultaneously. In general, pitch tracking techniques work in the time domain (autocorrelation, zero crossings, etc.) or frequency domain (Fourier transform, Constant-Q Transform etc.) or both (e.g. wavelet based). The pitch contour, usually obtained by identifying the pitch in each *frame* of a signal segmented into frames, is used for further musical analysis.

The other closely associated problem is the Beat Tracking problem. One approach in beat tracking is to find the note onsets accurately and use these for further processing.

b. Music Information Retrieval and Classification

Music information retrieval and classification are important in managing large databases of musical pieces. Prior art in music information retrieval is varied and draws from the fields of digital signal processing (DSP), musicology, artificial

intelligence (AI) and pattem recognition (PR) etc. Popular techniques of music information retrieval use some musical attributes like, rhythm and pitch, along with computational techniques like AI, PR, **DSP** etc.

c. Music **Information Retrieval and Classification Using Pitch Information**

Many music information retrieval systems use the pitch variations occurring in the music **as** a retrieval **key.** The *melody contour,* which records at least the directions of note movements, is widely used **as** a feature for matching tunes. The most common approach is to use the *approximate string-matching* algorithm to decide how dose two melody contours are. This is really the basis of most content-based music retrieval today, e.g. Ref. 14.

The sources of the melodies can be various. For example, manual inputs of musical scores can be used to analyze the nature of music. Alternates include automatic processing of musical signals and analyses of the corresponding music by accounting for processing errors or even ignoring them.

MPEG-7 **(Ref.** 1) has standardized a melody descriptor, based on the melody contour, for the Query by humming (QBH) application. This application is aimed at retrieving a ranked list of tunes similar to a query. The resolution of the melody contour does play an important part in accurate tune matching, but the first step is to have only **a** three level quantization - (Up, Down and Same), usually denoted by $\{U, D, S\}$. Therefore, the pitch of a musical piece is tracked and the melody contour is found by noting the direction of pitch movement from one note to the other. Experiments with human volunteers suggest that this is the way that humans match similar tunes. Let $\{U, D, S\}$ be the alphabet of the melody contour, where each element (i.e. U, D or S) is a *direction* (denoted by d) of the melody contour. Thus, the melody contour would be a set of directions with each direction possibly having information about the time of its note transition.

In melody retrieval, both pitch and rhythm information are important **as** shown in Ref. **13.**

d. Related Work specific to Indian Music

Very recently (Refs. **3** and 4), continuous pitch variation in Indian music in the context of studying the musical scale used in South India was reported independently. Notes (as defined by South Indian music grammar) have been extracted manually. The aim of this study is to find the nature of the musical scale used in India, while accounting for the continuous pitch variation. It highlights the importance of continuous pitch variation in Indian music.

A raga is a musical attribute central to Indian classical music and music derived from it. It roughly corresponds to a scale in Western classical music. Raga identification using Hidden Markov Models is under research. In this research, a threshold on the change of slope of the pitch curve marks notes, which are quantized to the nearest scale note. **A** special case of this change of slope is an extremum, where the slope changes sign as well.

The first attempt of QBH on Indian film songs, *not* using MIDI databases, was done at Indian Institute of Technology (IIT), Mumbai. This approach employs a slightly modified method of using melody contours. The details are available in Ref. 7.

Recently, a query by example (QBE) retrieval based on timbral features (energy distribution in audio bands) dealing with a database of Indian classical music was developed. The system, which is used to retrieve the same instruments (voice included) as the query, is described in Ref. 8.

Modern Western composers are looking at using exactly modelled ornamentations of Indian classical music in Western compositions. For example, Ref. **9** models the ornamentation using Bezier curves, and almost parallels the work in the MuM project (Ref. 10). In both cases, with the aim of exact reproduction of ornamentations in Indian classical music, curve fitting is used.

Even though Western classical music does not depend on ornamentation, the small-range pitch variations (vibrato) play a very important part. Work has been done on estimation, extraction and parametrization of vibratos (e.g. Ref. 15).

e. Other Related Work

Among QBH systems, one that deserves special mention in the context of the present invention is the work on continuous *melody contours* (Ref. 12). The title is confusing, but there is a significant difference. The continuous melody contour is constructed from a set of constant-pitch segments, mainly for resilience towards errors in note segmentation. Thus, it could be called 'time-continuous' melody contour. Having constant-pitch segments leads to loss of ornamentation information. On the other hand, the term 'continuous' is employed in the present invention to suggest smooth *pitch variation* It is possible to apply the time-continuous melody contour technique over the invented descriptor for the same reason, viz., segmentation error resilience.

In the accompanying drawings:

FIG. 1 illustrates a constant-pitch note;

FIG. 2 illustrates a note with continuous pitch variation;

FIG. 3 illustrates constant-pitch notes having the same melody contour as in FIG.2;

FIG. 4 illustrates the preferred descriptor extraction algorithm;

FIG. 5 illustrates a query by example system;

FIG. 6 illustrates a qualitative algorithm to find pitch discontinuities;

FIG. 7 illustrates a quantitative algorithm to qualify abrupt jumps in pitch as pitch discontinuities;

FIG. 8 illustrates an algorithm to qualify local insignificant minima in energy as note onsets;

FIG. 9 illustrates a method of music-genre classification;

FIG. 10 illustrates the average number of retrievals vs. query length for different retrieval methods using the preferred implementation;

FIG. 11 is a histogram showing number of pieces having a range of continuity percentage, calculated using the preferred implementation.

FIG. 12 illustrates the average number of retrievals vs. query length for different retrieval methods using an alternative embodiment of the invention; and

FIG. 13 is a histogram showing number of pieces having a range of continuity percentage, calculated using an alternative embodiment of the invention

Limitations of Prior Art:

The major limitation of current retrieval and classification systems is that notes are assumed to have constant pitches. **FIG.** 2 shows that this is not the case. FIG. 1 shows a note that has constant pitch (i.e. constant fundamental frequency) and **FIG.** *2* shows a note that has varying pitch. In fact, with this kind of pitch variation, the definition of 'note' needs to be clearer.

Such continuous pitch variation is found in different kinds of music. For example, in South Indian classical and North Indian classical music, as well as in folk and film music of India. Such pitch variations can be oscillatory or glides from one pitch level to another. Specific, well-defined techniques of executing these pitch changes exist (e.g. Ref. **11).** Continuous pitch variations are also possible in Jazz music (e.g. Ref. *5).* Moreover, with possibilities of blending different genres of music, continuous pitch variation must be treated as an independent dimension of music.

When such pitch variations occur in music, (most) people don't perceive a glide from one constant-pitch level to another, as many "notes". Rather, they perceive them as single entities with continuous pitch variation. It is expected that a variation such as the one shown in **FIG, 3** would be perceived very differently from the one shown in FIG. *2.* Note that the melody follows the same 'contour' in both cases. It starts from 490 Hz, moves up to **640** Hz, then down to *550* Hz and back to 640 Hz, but in the former, trmsitions are continuous while in the latter, they are clearly abrupt. The use of the nature of pitch transitions is lacking in the state of the art in music retrieval and classification.

a Unsuitability of conventional MIDI databases for testing

Most researchers have used MIDI files for purposes of experimentation However, with continuous pitch variation, MIDI files, especially those that are freely available, cannot be used as they do not retain accurate pitch variation information, though MIDI does allow for continuous pitch variation in the form of pitch-bend.

Proposed Solution (With examples):

It is an object of the present invention to provide a method of semi-automatic or automatic information retrieval using melody information and continuity information.

It is a further object of the present invention to provide an improved method of music classification using the percentage of continuous boundary transitions derived from the melody information and the said continuity information wherein the genres of music include, but are not limited to, North Indian classical, South Indian classical, Indian folk, Jazz, Western classical, folk, country, blues, rock etc.

Summary of The Invention:

The present invention is a method to use continuity information in automatic retrieval and classification of music. The scope of the invention is **not** limited to Indian music only. It can be used for all kinds of music that employ continuous pitch variation, e.g. in some parts of Jazz music and in multi-genre corpora etc.

Given that continuous pitch variation is an independent dimension of music, the existing retrieval schemes cannot be applied directly. In prior art, whenever the pitch changed from a constant-pitch note to another constant-pitch note, the direction was easily identified. Invariably, a note was considered a note if it had a constant pitch (quantized to the nearest musical scale note) for a minimum period of time. With continuous pitch variations, the *boundaries* of the continuous pitch variations are used to find the directions. These boundaries can be visualized from FIG. 2. We call the directions thus derived as the *modified melody contour*, while stressing that melody information, in general, can include more information, e.g. note duration etc. To use continuity information, at least the nature of the variation of the pitch from one boundary to the next must be retained. There are several ways of doing this and some of these are explained later.

Additionally, these continuous pitch variations are *not* to be confused with the nature of note production, for example, by the human voice. It is known that even trained singers only converge to a constant-pitch note, rather than sing the accurate pitch

straightaway. One could distinguish this form of continuous variation of pitch from the deliberate, aesthetic variation **of** pitch. The preferred implementation does not make this distinction.

Also, it is useful if pitch is deemed to have varied discontinuously across note onsets, since a new note is perceived at each onset. Any convenient method can be used for identifying note onsets.

a. Preferred Implementation of Descriptor Extraction (100 in FIG. 4) and Music Retrieval (200 in FIG. 5)

The simplest way of using continuity information is to identify the continuity and the lack of it in the tracked pitch of a musical piece. Thus, only one bit is required for storing the continuity information. This bit, appropriately called the *continuity flag* (denoted by c), is defined to correspond to the nature of the pitch change.

The continuity flag associated with a direction is 1 if the corresponding pitch transition is continuous and 0 otherwise.

Energy-based segmentation forces $c = 0$ at every note onset as well.

Each direction (i.e. *U, D or S*) of the modified melody contour has an associated continuity flag. A special case of melody information and continuity information together is the modified melody contour with the continuity flag, which we call as *descriptor* in the rest of the document.

In the above definition, the term 'continuous' needs clarification. While any implementation of retrieval with the descriptor must use a precise definition, it could be application-dependent, tolerance-dependent etc. Several definitions that have been employed lead to similar results. The preferred implementation is shown in **FIG.** 4 and is described below. It is worth noting that the continuity flag is simple and requires only one additional bit. Thus, it could be easily added in existing schemes especially considering the improved search accuracy it provides, for instance, in the query by example scheme that we used **(FIG.** 5).

Pitch Tracking **(110** in FIG. 4)

Any suitable pitch-tracking scheme, including signal processing techniques, sensor techniques etc. may be used to obtain the pitch contour from which the melody and continuity information can be derived. The digital signal processing based method that we used is described below for the sake of completion.

The input audio samples are segmented into overlapping frames **and** windowed with the Hamming function. The frame size depends on the minimum fundamental frequency occurring in the music **(Ref. 2). Our** values were **16** ms for the flute ($f_0 >$ **300** Hz) and **32** ms for the violin ($f_0 > 150$ Hz). The choice of frame-shift (hop-size) is dictated by the rate of change of pitch. It was fixed empirically at **8** ms. Of the numerous ways of tracking pitch, the *autocorrelation* followed by the *Short Term (Discrete Time) Fourier Transform* (STFT) was used. The autocorrelation function gave the approximate pitch period ($1/f_0$). For better resolution in frequency, the STET around the approximate fundamental frequency **was** evaluated every 2 Hz (other resolutions are possible) in the range

$$\left\lfloor \frac{f_s}{n+2} \right\rfloor \le f \le \left\lceil \frac{f_s}{n-2} \right\rceil, \; n = \left[\frac{f_s}{f_0} \right]$$

where $\lfloor x \rfloor$ denotes the greatest integer lesser than x, $\lceil x \rceil$ denotes the least integer greater than x and [x] indicates the nearest integer to X.

Harmonic Jumps Correction: (**120** in **FIG.** 4)

Any pitch-tracking scheme would invariably bring about harmonic errors – choosing some harmonic component rather than the fundamental. Besides, when there is no signal the recorded position of the maximum is arbitrary. To counter *this,* if the pitch at some frame is found to be close (nominally within a semi tone) to double the pitch of an adjacent frame, then the pitch is halved. This is done iteratively till there is no such doubling.

Even after taking these "corrective" measures, some pitch tracking errors do remain with current techniques. The preferred descriptor extraction algorithm accounts for some of these errors.

From Pitch to Musical scale (Frequency to relative notes): **(130** in FIG. **4)**

To simplify threshold selection, the dimension of Hz is done away with by a simple transformation, which leads us to the musical scale. The frequency values are converted to note values **as** follows. Each note value, $n[i]$, at the i^{th} frame is calculated as

$$n[i] = 12\log_2\left(\frac{P[i]}{P_0}\right), \quad i \in \{0,1,\ldots,N\}$$

where $P[i]$, in **Hz**, is the pitch at the i^{th} frame, P_0, also in **Hz**, is the user-supplied tonic and N is the number of frames. Thus, absolute frequencies have been converted to relative notes, **All** fiirther operations are performed on these note values.

Importantly, the value of the tonic does not affect the extraction procedure itself. For, if the note values, $n_1[i]$, are found based on **an** arbitrary tonic P_1, then

$$n_1[i] = 12\log_2\left(\frac{P[i]}{P_1}\right) = n[i] + 12\log_2\left(\frac{P_0}{P_1}\right)$$

which translates $n[i]$, but the extraction algorithm uses **only** differences **in** $n[i]$.

*The step of converting pitch values to note values **is** a matter **of** convenience. The following algorithm could be used with appropriate changes and the thresholds and other parameters can be evaluated mathematically, when working with absolute pitch values **or** in any other transformed domain.*

Extraction of Melody and Continuity: (140 in FIG. 4)

Extraction of melody and continuity comprises the steps of, in no particular order,

1. Marking pitch discontinuities **(141** in FIG. **4)**
2. Marking note onsets **(142 in** FIG. **4).** Continuity information is derived from the pitch discontinuities and note onsets.
3. Melody extraction **(143 in** FIG. **4)**

The above steps are described below.

Marking Pitch Discontinuities: **(141** in **FIG. 4)**

A qualitative explanation of the algorithm is given in **FIG. 6.** The left branch of the flow chart is for testing each point for **an** abrupt jump. That is, if one note differs from the previous one by more **than** a threshold **A,** it is marked **as an** abrupt jump. From the set of abrupt jumps, the ones that qualify **as** discontinuities are chosen by applying the algorithm given in the right branch of the flowchart. For the abrupt jump at k, four neighbouring sets of note values (henceforth called *windows)* are considered – two to the left and two to the right of the abrupt jump. These Windows, shown at the top of **FIG. 6,** are not to be confused **with** the windowing function. Each of them is a set of note values of consecutive frames. Note that none of these windows includes $n[k]$ so that single-point errors in pitch tracking are not marked **as** discontinuities. Thus, the qualitative algorithm simply states that the pitch discontinuities correspond to the absence of convergence (in a threshold sense) of left and right limits at an abrupt jump. The methods of measuring the said convergence are many – only the preferred one that accounts for some errors in pitch tracking is described below in detail.

Consider a set of note values $n[i]$ at the i^{th} frame, where $i \in \{0,1,...,N\text{-}1\}$ and N is the number of frames. Points of pitch discontinuity are found **as** follows.

- **An** abrupt jump occurs at the k^{th} frame if successive note values differ by more than a threshold. That is, if

$$|n[k] - n[k\text{-}1]| > \Delta, k \in \{1,2,...,N\text{-}1\}$$

For genres of music such as Indian classical, Δ=0.75 is preferred. However, $A=l$ would work for the universal equal tempered musical scale. Other musical genre-dependent thresholds may be used if required.

If there is no abrupt jump at k, then we have a continuity at k.

An equivalent criterion to find abrupt jumps based on absolute pitch values is **as** follows. If the ratios of successive absolute pitch values is either less then **2-1/16** or is greater **than** $2^{1/16}$, then the location of the later pitch value is marked **as an** abrupt jump. Similarly, the following steps could be modified suitably when using absolute pitch values.

- Define four sets of note values, viz.

$$L = \{n[i]\}\ |_{k-W}^{k-1}$$
$$l = \{n[i]\}\ |_{k-w}^{k-1}$$
$$R = \{n[i]\}\ |_{k+1}^{k+W}$$
$$r = \{n[i]\}\ |_{k+1}^{k+w}$$

In the above sets, W is the number of frames used to verify if a note is of constant pitch. Similarly w is the number of frames used for curve fitting. Empirically, it was found that $WT_w \approx 100$ ms and $wT_w \approx 30$ ms were useful, though other values are possible. (Here, T_w is the frame-shift.) Thus, w and W determine the size of the neighbourhood around k.

- Find the means of L and l as $\hat{L}$ and $\hat{l}$ respectively. Similarly, find $\hat{R}$ and $\hat{r}$ as the means of R and r. The mean of a set is defined as the arithmetic mean of all its note values. For example,

$$\hat{L} = \frac{1}{W}\sum_{i=W}^{1} n[k-i]$$

The definitions of $\hat{R}, \hat{r}$ and $\hat{l}$ are similar

- Define v_L and v_R as

$$v_L = |\max(L) - \min(L)|$$
$$v_R = |\max(R) - \min(R)|$$

Define m as

$$m = \frac{n[k-1] + n[k+1]}{2}$$

Define O as **TRUE** if

$$|\hat{l} - \hat{r}| < \Delta \ \&\ |\hat{L} - \hat{R}| < \Delta \ \&$$

$$(|m - n[k]| < \Delta)\ \&\ \left(\left|\frac{m - n[k]}{n[k+1] - n[k-1]}\right| < 0.25\right)$$

and FALSE otherwise.

The above definition of O ensures that some errors in the pitch tracking are taken care of. The types of errors covered are spurious jumps in the pitch (first two conditions) and slight errors in smoothly varying pitch (to the

extent of 25% in the last two conditions). Moreover, note that O is tested **only** if at least one of L and R is not a constant-pitch note.

Any method of pitch tracking could require its own set of corrective measures, possibly different from the above.

- Define a *prediction* function $f(x)=ax^2+bx+c$, where a, b and c are to be determined. f is estimated by minimizing the squared error **(Ref. 6):**

$$\sum_{i=0}^{w-1}(f(i)-n[i+k+1])^2$$

Evaluate $\tilde{r}_{-1}=f(-1)$ and $\tilde{r}_{-2}=f(-2)$. Similarly define another prediction function g, of the same (quadratic) form. Estimate g by minimizing the squared error:

$$\sum_{i=0}^{w-1}(g(i)-n[k-w+i])^2$$

Evaluate $\tilde{l}_1=g(w+1)$ and $\tilde{l}_2=g(w+2)$.

- With the above definitions the algorithm shown in **FIG.** 7 is applied. In this figure,
  - The conditions $v_R < \delta$ and $v_L < \delta$, are used to decide if the window is **of** constant pitch.
    - The convergence is tested using mean values for constant-pitch windows, and predicted values for continuous pitch windows, the prediction being done at the closest pitch value to the abrupt jump if the other window is of constant-pitch, and at the location of abrupt jump if the other window is not of constant-pitch
  - The prediction is usually based on a window smaller than the one used to test for constant-pitch.

Marking Note Onsets: **(142** in **FIG.** 4)

Consider a set of energy values $E[i]$ of the signal in the sub-frame of the i^{th} frame, where $i \in \{0,1,...,N_f\text{-}1\}$ and N_f is the number of frames within two pitch discontinuities. The sub-frame of a frame is the first $M = \lceil \frac{\text{frame-shift}}{\text{sampling period}} \rceil$ samples of the frame. **We** find locations of note onsets **as** follows.

- Smoothen the energy curve using an n-point moving average filter, with n such that $nT_w \approx 250ms$.
- Locate the positions of peaks (i.e. local maxima) and troughs (i.e. local minima) in the energy curve. Let $\tilde{T}$ and $\tilde{P}$ be the set of positions of troughs and peaks respectively. If $B \in \tilde{P}$, then $E(P)$ denotes the energy of that peak and similarly $E(T), T \in \tilde{T}$ denotes the energy of the trough **T.**
- Define a threshold t to decide how small a minimum should be compared to the next best peak. We chose t=0.05, i.e. 13dB below the next maximum.
- From the set $\tilde{P}$, find the positions of *significant maxima,* i.e. maxima that are greater than $\min(\frac{\max(E(\tilde{P}))}{5t})$ and collect them in set $\mathcal{P}$. Similarly find the positions of *insignificant minima,* i.e. minima that are lesser than 5*max($E(\tilde{P})$*t) and collect them in set $\mathcal{T}$.
- For each the trough, $T \in \mathcal{T}$, find **the** newest trough $\hat{T} \in \mathcal{T}$, $\hat{T} > T$. Between the two troughs, find the maximum peak $\hat{P} \in \mathcal{P}$ and apply the algorithm given in FIG. **8.**

Melody Extraction: **(143** in FIG. 4)

All pitches within two discontinuities are deemed to have been varied deliberately continuously (except when there is no signal). Within such *segments,* the continuity ***flag*** is 0 for the first direction and *1* for the remaining directions that are found from the boundaries of the pitch contour within the segment.

The algorithm to find the boundaries of the pitch contour within each segment is as follows,

- Based on a histogram, eliminate notes whose occurrences are less than 5% of the maximum occurrence. (This again, may be used to take care of errors in pitch tracking.)
- Filter the segment using a moving average filter of length W_n, such that $W_n T_w$ nominally equals 50ms.
- Identify constant-pitch sections (usually at least $2W_n$ long) of the pitch curve. These are marked as notes with the mean note value. A constant-pitch section is one whose maximum and minimum differ by less than a threshold δ (usually 0.5).
- Identify *curved* sections between constant-pitch regions or between the beginning/end of the segment and the nearest constant-pitch region. Ignore the curved section if it is shorter than the threshold, W_n. Otherwise,
  (a) Find the local extrema of the curved section
  (b) Retain only those extrema that cause a change in the direction of the modified melody contour
  (c) Replace two consecutive extrema that differ by less than the threshold, δ, by their arithmetic mean.
  (d) Iterate (b) and (c) till the set of extrema converges to some finite set or the null set
- Cluster the remaining extrema that are separated by less than δ.

Further, one could store the locations and durations of the identified boundaries to help in playing back accurately the music that was found to match a query. The modified melody contour is found by marking the differences between adjacent boundaries as 'Up', 'Down' and 'Same' (U, D, S). More detailed melody information could also be used.

Retrieval algorithm using Continuity:

Notation: The ordered set of directions $\{D,S,U\}$ is represented by the ordered set of integers $\{-1,0,1\}$. The continuity flag can be either 0 or 1. That is,

$$d[n] \in \{-1,0,1\}, \text{ and}$$

$$c[n] \in \{0,1\}, \text{where } n \text{ is the time (or frame) index}$$

Here $d[n]$ is the n^{th} direction of some tune and $c[n]$ is the corresponding continuity flag. Further, let the actual corresponding pitch changes in the modified melody contour be $P[n]$. Thus,

$$d[n]=\text{sgn}(P[n])$$

where sgn (x) is defined as

$$\text{sgn}(x) = \begin{cases} +1, x > 0 \\ 0, x = 0 \\ -1, x < 0 \end{cases}$$

The basic alphabet of the descriptor $m[n]$ consists of complex numbers such that

$$m[n]=d[n]+jKc[n]$$

where $j = \sqrt{-1}$ and K is a constant that reflects the importance of continuity information. We chose $K = 1$, but other values are possible, and it could be tune-dependent, for e.g. based on range of $P[n]$ etc.

Also, define $M[n]$ as

$$M[n]=P[n]+jKc[n]$$

Clearly,

$$m[n] \in \{-1, 0, 1, -1+jK, jK, 1+jK\}$$

Three approaches are analyzed below (other possible approaches exist).

Approach 0: This is the existing method, where $d[n]$ is used for string matching and $P[n]$ for Euclidean distance calculation. Euclidean distance is used to rank all results that have the same modified melody contour. Euclidean distances are calculated as given in Approach 1 with the condition that $c[n]=0 \; \forall \, n$.

Approach 1: $d[n]$ is used for the purposes of approximate string matching but the results are ranked based on the continuity *flags* $c[n]$ as well. That is, given a set of results $\mathcal{R}$ that have the same edit distance, which is well known to those familiar with prior art, they are ranked by the Euclidean distance between each result and the query. The ranking scheme is as follows.

Let $m_i[n]$, $n \in \{0...L_i\text{-}1\}$ be the descriptor of the i^{th} retrieval $R_i \in \mathcal{R}$ where L_i is the length of R_i. Let A_i is the position of the match in R_i Then,

$$m_i[n]=d_i[n]+jKc_i[n],\ n \in (0,1,...,L_i\text{-}1\}$$

For the query, the descriptor $m_q[n]$ is given by

$$m_q[n]=d_q[n]+jKc_q[n],\ n \in \{0,1,...,L_q\text{-}1\}$$

where L_q is the length of the query. Note that the subscript q is used for the query. The Euclidean distance, E_i, between the Q and R_i is defined as

$$E_i = \sum_{n=0}^{L_q-1} | M_i[n+A_i] - M_q[n] |^2$$

and the queries are ranked in ascending order of E_i's. Without continuity information, E_i is calculated as:

$$E_i = \sum_{n=0}^{L_q-1} | P_i[n+A_i] - P_q[n] |^2$$

Approach 2: In this approach, $m[n]$ rather than $d[n]$, is used for the string matching and $M[n]$ for Euclidean distance calculations. It provides for very strict matching.

Any of the above approaches (1 or 2) can be used in retrieval, or any other improved retrieval scheme that uses the continuity information also.

b. Preferred Implementation of Music Classification using continuity (300 in FIG. 9)

As is well known to those skilled in music, an important difference between Indian and Western forms of (classical) music is the amount of continuous pitch variation – Indian music being replete with it. Based on this, a classification algorithm **(300)** is

shown in FIG. 9. The feature used in classifying music genres broadly is the percentage of continuous note transitions, which is extracted (310 in FIG. 9), and an appropriate threshold can be used for classification (320 in FIG. 9). The *percentage of continuity*, α_p, is defined as:

$$\alpha_p = \frac{\text{Number of continuous extrema transitions,}}{\text{Number of extrematransitions}} 100$$

e. Results

Experimental Set-*up for Music retrieval:*

The experimental set-up consisted of the query by example system (200) shown in FIG. 5. The database (230 in FIG. 5) consisted of samples of unaccompanied classical instrumental recorded (onto magnetic tapes) in normal residential conditions and digitized using commonly available sound cards (that come with PC's). Single-channel (mono) digitization was done at a sampling rate of 16000 Hz using an audio jack connected from a commercial tape recorder. The database consisted of natural samples of South Indian classical, North Indian classical and Western classical genres of music.

For each file (about a minute long), the above melody and continuity were extracted (100 in FIG. 5) and stored as metadata. In a query by example set up, the user could choose the following parameters, through an interface (210 in FIG. 5), before the system picked up a random query (211 in FIG. 5) to be used for retrieval of similar pieces:

- The genre of the music to pick the query from – Required
- The maximum length of the query – optional. The default was set to 10 transitions.
- The specific piece to pick the query from – Optional. If not specified, a random file of the given genre was selected to pick a query.

The query is processed by searching (220 in FIG. 5) the database. The accompanying results (240 in FIG. 5) pertain to such randomly picked queries. Five distinct queries from each genre were chosen for averaging the results. Since the exact and approximate string matching methods were tested separately, the number of distinct

queries is at least 15 and at most **30.** Multiple matches within a tune are possible, but one 'retrieval' corresponds to a single tune.

Retrieval Results:

Using the three different approaches, viz., Approach 0, Approach 1 and Approach 2, give search results as predicted. Approaches 0 and 1 give the maximum number of retrievals while Approach 2 reduces the number of retrievals. These observations can be seen in **FIG. 10** (a) to (f). The reduction in the number of retrievals should help the user to locate faster, the exact piece he/she is looking for.

ClassificationResults:

Typical continuity percentages for the three classical genres of music are given below.

Music Genre ↓	Percentage of Continuity		
	Min	Mean	Max
South Indian	36	56	67
North Indian	51	66	7s
Western	10	21	34

It is clear from the above that the feature 'percentage of continuity', can be used to classify music into Indian and Western – if the continuity percentage is less than 35% (threshold), it is Western music, and Indian music otherwise. A glance at the histogram in FIG. 11 would bear this out. *Depending on the continuity detection algorithm, the threshold could change.*

d. Alternative Embodiments of Invention

Manual Extraction: For example, one could define discontinuities based on (subjective) visual criteria. If the subject, who has a graph of pitch vs. time, feels that there is an abrupt jump in the pitch, then that point is marked as a pitch discontinuity. Similarly, a note onset could be found by listening to a piece. When the subject perceives attacks, the corresponding dips in the energy are marked as note onsets. The

boundaries can be found as in the preferred implementation. The descriptors were extracted in this way from a very small database. The results are similar to those shown in FIG. 10 and FIG. 11.

Improved continuity extraction algorithm: Another embodiment of the invention employs a modified version of the algorithm shown in FIG. 7. To increase robustness to pitch-tracking errors, checking for convergence is done using predictions from either *filtered* or raw neighbouring note values. The retrieval results using this algorithm are shown in FIG. 12 and the corresponding improved classification results in FIG. 13.

Re-ordering in the extraction algorithm: It is possible to extract continuity information from musical data using slight variations of the method shown in FIG. 4. For example in FIG. 4, one could perform 'marking note onsets' **(142)** before 'marking pitch-discontinuities' (**141**), or one could even move 'marking note onsets' *outside* 'Extraction of melody and continuity **(140)'** and restrict continuity extraction to 'marking pitch continuities' **(141).**

It should be appreciated that similar variations are possible and could be made by those skilled in the art without departing from the scope of the invention as defined in the following claims.

References:

Standards

1, J. M. Martinez, *MPEG-7 Overview,* http://www.chiariglione.org/mpeg/standards/mpeg-7/mpeg-7.htm

Theses

2. Venkatasubramanian **V,** *Music Information Retrieval using Continuity,* M.Sc.(Engg) thesis, Indian Institute of Science, Bangalore. To be submitted.

Publications

3. Arvindh K, "Application of Bitch Tracking to South Indian Classical music," http://www-ccrma.stanford.edu/~arvindh/cmt/ieee-icassp-2003-05-00557.pdf

4. Arvindh K, "Pitch Measurements versus Perception of South Indian Classical Music," http://www-ccrma.stanford.edu/~arvindh/cmt/smac03.pdf

5. Dylan Menzies, *New Electronic Performance For Electroacoustic Music,* Doctoral Dissertation, Department of Electronics, University of York, May 1999.

6. E. Kreyszig, *Advanced Engineering Mathematics,* John Wiley and Sons, seventh ed., 1993.

7. M.A.Raju, B.Sundaram, and P. Rao, "TANSEN: **A Query** by Humming based Music Retrieval System," http://www.tenet.res.in/ncc2003/ncc/C-4.pdf

8. Roongta, S. Sood, S. Chaudhury, and **A.** Kumar, "Content Based Retrieval in a Repository of Indian Music," in *Proceedings of International Conference on Communicatoins, Computers and Devices,* pp. 641 – 644, Austrian Computer Society, 2001.

9. Battey, "Computer modeling of Ornamentation in Hindustani Music," to appear.

10. P. Pandya, http://www.tcs.tifr.res.in/~pandya/music/index.html

11. P. Sambamoorthy, *South Indian Music – Books III to VI.* The Indian Music Publishing House, Madras, 1960 – 1964

12. Y. Zhu and M. Kanakanhalli, "Similarity Matching of Continuous Melody Contours for Humming Querying of Melody Databases." http://www.comp.nus.edu.sg/~cs4241/TR2002.pdf

13. Y. E. Kim, W. Chai, R. Garcia and B. Vercoe, "Analysis of a contour-based representation for melody," http://www.media.mit.edu/~chaiwei/papers/Kim.pdf

14. A Ghias, J Logan, D Chamberlain and B. C Smith, "Query by Humming Musical Information Retrieval in An Audio Database," *ACM Multimedia*, 1995, pp. 231 – 236

15. S Rossingol, P Depalle, J Soumagne, X Rodet and J. L Collette, "Vibrato: Detection, Estimation, Extraction, Modification," *Proceedings of the COST-G6 Workshop on Digital Audio Effects (DAFx-99)*, Dec. 1999

Related research labs

Some research projects around the world involved in music information retrieval are listed below.

Project/Lab	Web site
MELDEX	http://www.dlib.org/dlib/may97/meldex/05witten.html
Leeds	http://www.leeds.ac.uk/music
Media Labs, MIT	http://www.media.mit.edu/research
CCRMA	http://ccrma-www.stmford.edu
IRCAM	http://mediatheque.iamfr/index-e.html
McGill University	http://www.mcgill.ca/music-resources

Claims

1. A method of semi-automatic or automatic music information retrieval using melody information and continuity information, comprising the steps of, in no particular order:

a Obtaining the Melody information, at the desired resolution, comprising at least the extrema of the pitch contour, the said pitch contour being obtained by any convenient means; where the melody contour known to those ordinarily skilled in the art, is a special case of the melody information requiring at least 2 bits for representation and is derived from music usually assumed to consist of series of constant-pitch notes and where continuity information is unused.

b. Obtaining the Continuity information consisting of pitch continuity information, derivable from the said pitch contour; and note onset information, where the pitch continuity information contains information about the traversal of musical pitch between and/or across musical scale notes, the said information being at any desired resolution and the said musical scale being the same as that known to those skilled in the art; or is defined according to genres of music including, but not limited to, North Indian classical, South Indian classical, Indian folk, Jazz, Western classical, folk, country, blues, rock etc.,

c. The retrieval step comprising retrieving musical pieces using at least the said continuity information and melody information, wherein the melody information could be at my desired resolution;

and wherein the music information retrieval comprises retrieving desired musical pieces from a database of musical pieces, browsing such a database etc., which could be done in conjunction with any already existing music retrieval technique known to those familiar with prior art, wherein the said database includes, but is not limited to, private databases, archives, Internet, etc. comprising musical data stored as audio, audio tracks in multimedia data, musical notation, and other representations of music, including encoded representations; or embedded within other kinds of audio data including, but not limited to, audio effects, conversational speech, lectures, background music etc.

2. A method of obtaining the continuity information consisting of the steps of, in no particular order, marking pitch discontinuities and marking note onsets, the said continuity information consisting of at least one bit, where the one-bit continuity information is called the continuity flag, with usually 0 representing either a pitch discontinuity or a note onset and 1 representing pitch continuity and absence of a note onset.

3. A method as claimed in claims 1 and 2, wherein the step of marking the note onsets comprises the steps of

   a. Finding local minima in the smoothened or raw energy curve of a musical piece, and

   b. Qualifying an insignificant local minimum as a note onset if, and only if, the next significant local maximum is above a threshold, nominally 20 times the said minimum and only if no other insignificant local minima lie between the said minimum and the next significant local maximum.

4. A method as claimed in claim 2, wherein the step of marking pitch-discontinuities, consists of the step of partly or fully automatically finding discoiitinuous transitions in the pitch curve of a musical piece.

5. A method as claimed in claim 4, wherein the step of partly or fully automatically finding discontinuous transitions in the pitch curve of a musical piece, preferably after converting the pitch values to logarithmic note values, similar to MIDI note values, using my suitable reference pitch value, comprises the steps of

   a. Marking abrupt pitch transitions, wherein an abrupt pitch transition is defined by the increase or decrease in successive note values by more than a threshold, the said threshold nominally being ¾ of a semitone for classical forms of music, or any other convenient value; and

   b. Qualifying the abrupt pitch transitions as a pitch discontinuity or continuity according as the left and right limits of the note values diverge or converge, wherein the said convergence is usually determined by the threshold of half a semi tone, and the said divergence is the absence of the said convergence; and

corrective measures in determining the limits to account for errors in pitch tracking, if any, may be applied as required.

6. A method as claimed in claim 5, wherein the said left and right limits are defined symmetrically as:

a. The arithmetic mean value of a set of note values of constant pitch neighbouring the abrupt pitch transition, wherein 'neighbouring' is defined empirically based on the musical genre or theoretically for all genres depending on the expected duration of constant-pitch notes or by any other convenient means; and

b. The predicted value of the note values of varying pitch neighbowing the abrupt pitch transition, the prediction being done at the location of the note value closest to the abrupt transition, on the other side of the respective limit provided the neighbouring note values on the other side are of constant-pitch and the prediction being done at the abrupt pitch transition provided the other neighbouring note values are not of constant pitch, the said prediction being done by any convenient means including, but not limited to, curve fitting, subjectively etc., wherein curve fitting, as is well known to those familiar with prior art, could use polynomial curves, trigonometric curves, spline curves or any other useful curve, where the number of neighbowing note values used in prediction is chosen conveniently and could be different from the number of neighbouring note values used in testing for constant-pitch.

7. A method as claimed in claim 1, wherein the said music retrieval comprises matching tunes using the approximate or exact string matching algorithm on a combined alphabet having the same cardinality as the cross-product of the alphabet of the melody information consisting of the possible directions of the boundary transitions, and the alphabet of the continuity flag; wherein the said boundaries comprise at least the extrema of the pitch contour.

8. A method as claimed in claim 1, wherein the said music retrieval comprises matching tunes using the approximate or exact string matching algorithm on the

combined complex alphabet obtained by using the melody information consisting of the possible directions of the boundary transitions suitably mapped to real numbers, as the real part; and the alphabet of the continuity flag suitably mapped to real numbers, as the imaginary part of the said alphabet, or vice-versa.

9. A method as claimed in claim 1 wherein the actual or quantized boundary transition values and the corresponding continuity *flags,* suitably mapped to real numbers, are respectively used as the real parts and imaginary parts, or vice-versa, of complex numbers used in calculating distance between melodies; with or without employing the said combined alphabet or the said combined complex alphabet in approximate or exact string matching.

10. Any method of using continuity information as claimed in claim 1b in semi-automatic or automatic music analysis including?but not limited to, classification of music, identification of musical attributes, etc., possibly in conjunction with any existing techniques.

11. A method of extracting the melody information, iteratively or otherwise, from the local extrema of the pre-processed or raw continuous-pitch sections with directional repetitions removed, the said extrema being clustered according to the optimal scale note resolution for the musical genres, wherein the said optimal scale note resolution is well known to those familiar with the musical theories of the respective genres and wherein any relevant resolution could be used in the absence of such a theory; and wherein the said continuous-pitch sections consist of the note values of a musical piece between successive pitch discontinuities and/or musical note onsets as claimed in claim 6.

12. A method of classification of genres of music such as the ones in step 1b using the percentage of continuous boundary transitions, wherein a continuous boundary transition is a boundary transition within a continuous pitch section and a discontinuous boundary transition is a boundary transition across continuous pitch sections, i.e. boundaries separated by a pitch discontinuity or note onset.

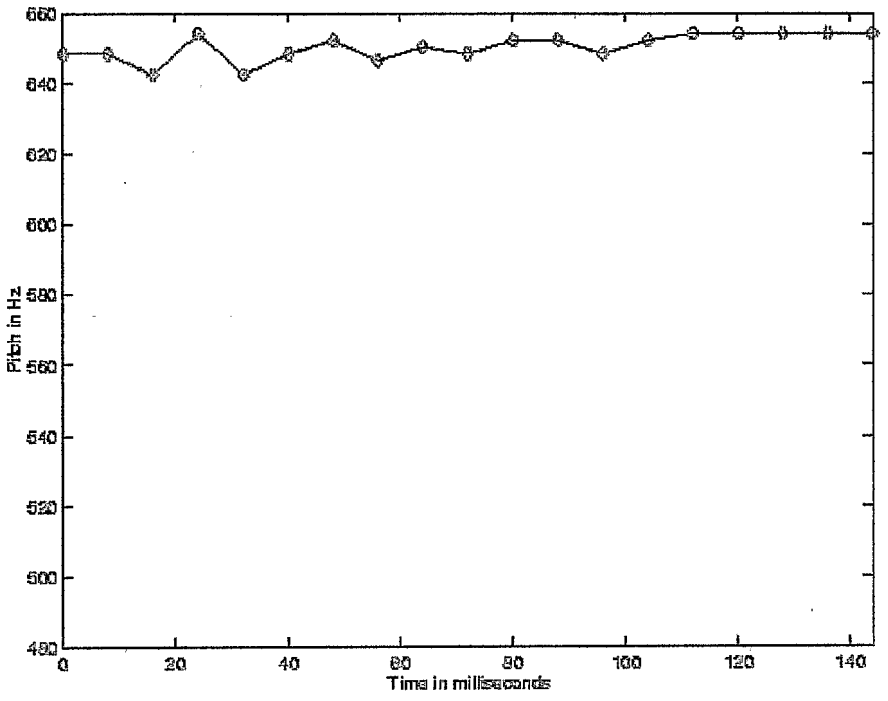

660
640
620
600
580
560
540
520
500
480
Pitch in Hz
0
20
40
60
80
100
120
140
Time in milliseconds

FIG 1.

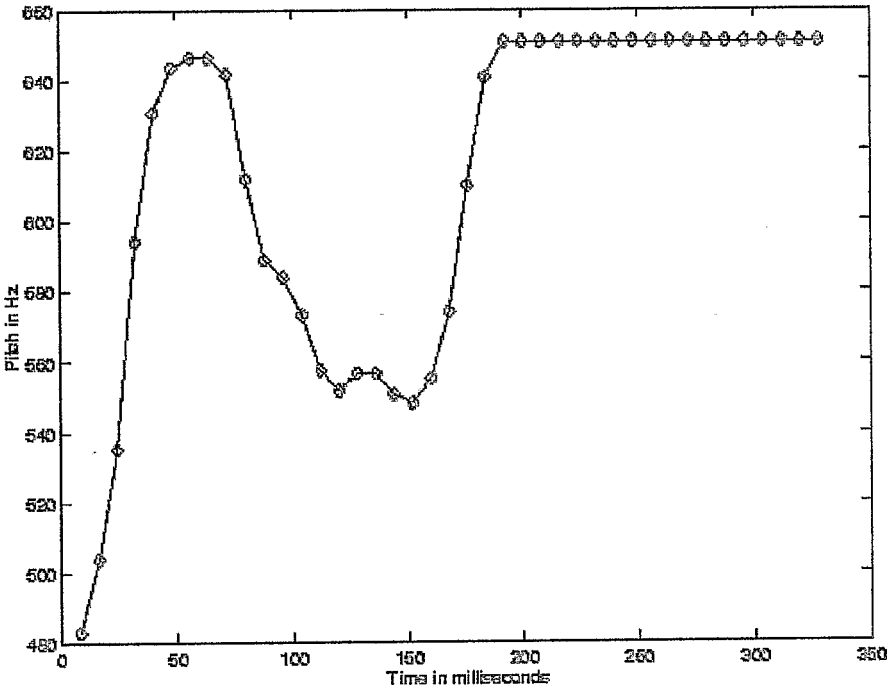

Pitch in Hz
660
640
620
600
580
560
540
520
500
480
0
50
100
150
200
250
300
350
Time in milliseconds

FIG 2.

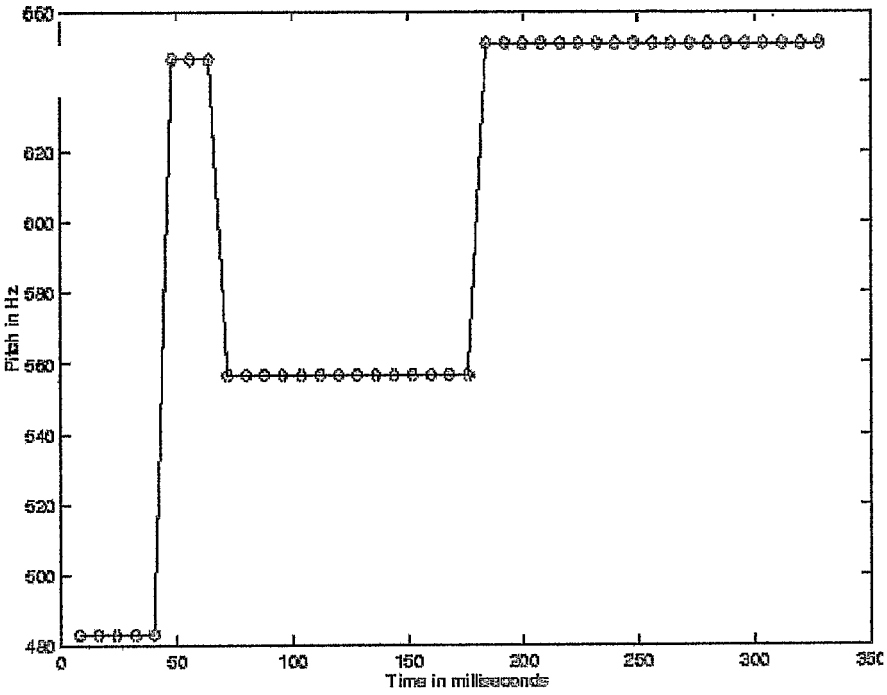

660
620
600
580
560
540
520
500
480
Pitch in Hz
0
50
100
150
200
250
300
350
Time in milliseconds

FIG 3.

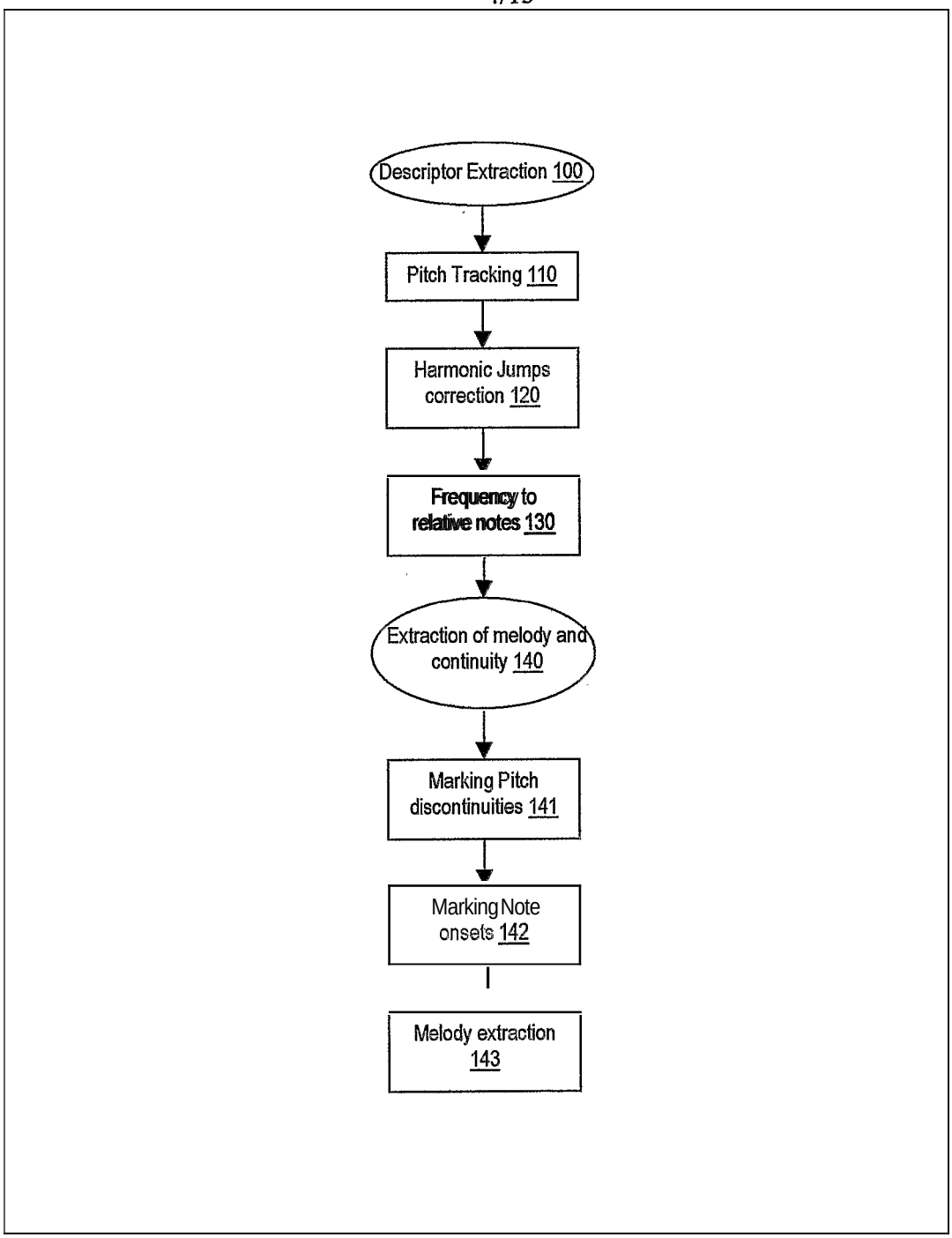

Descriptor Extraction 100
Pitch Tracking 110
Harmonic Jumps correction 120
Frequency to relative notes 130
Extraction of melody and continuity 140
Marking Pitch discontinuities 141
Marking Note onsets 142
Melody extraction 143

FIG 4.

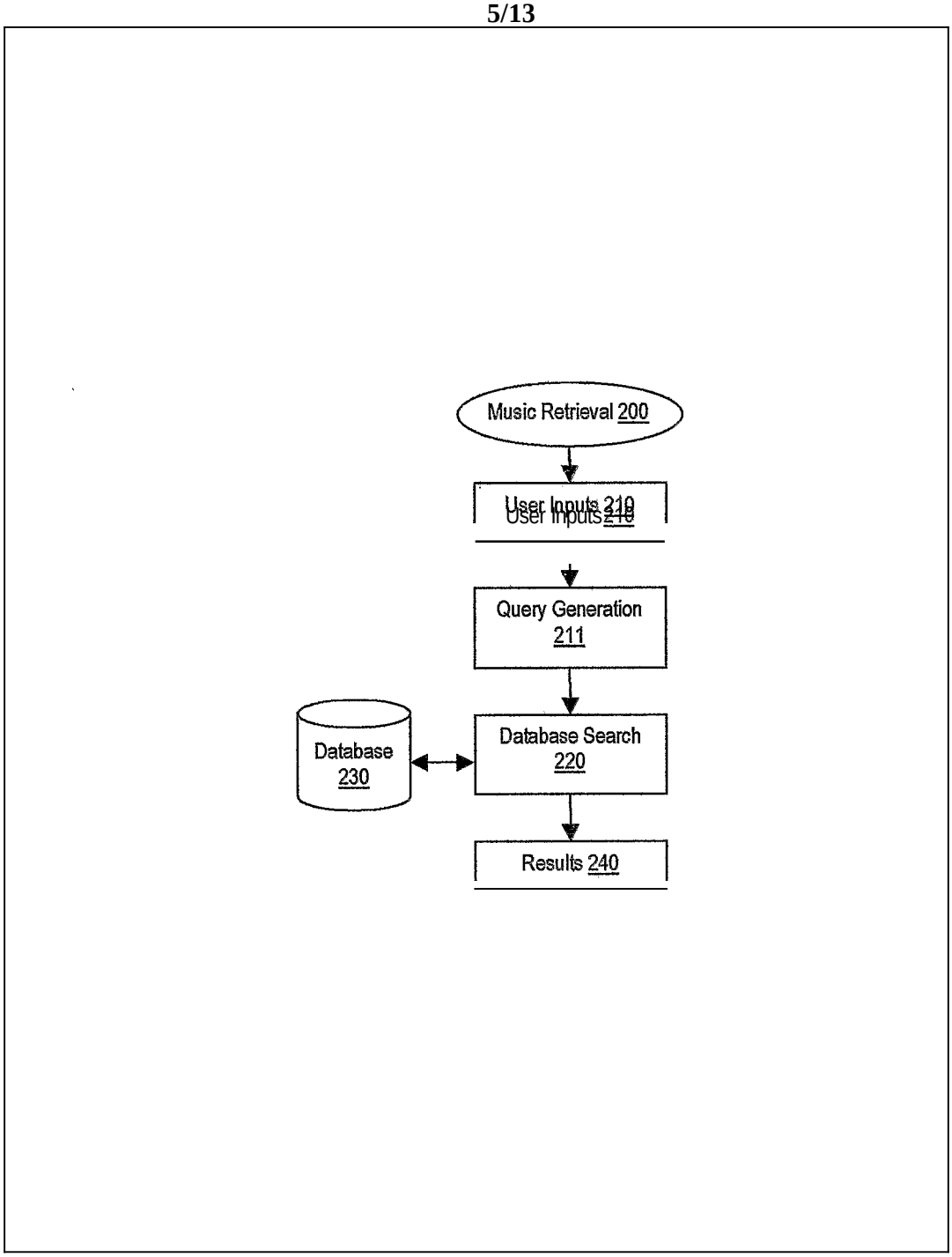

Music Retrieval 200
User Inputs 210
Query Generation 211
Database 230
Database Search 220
Results 240

FIG 5.

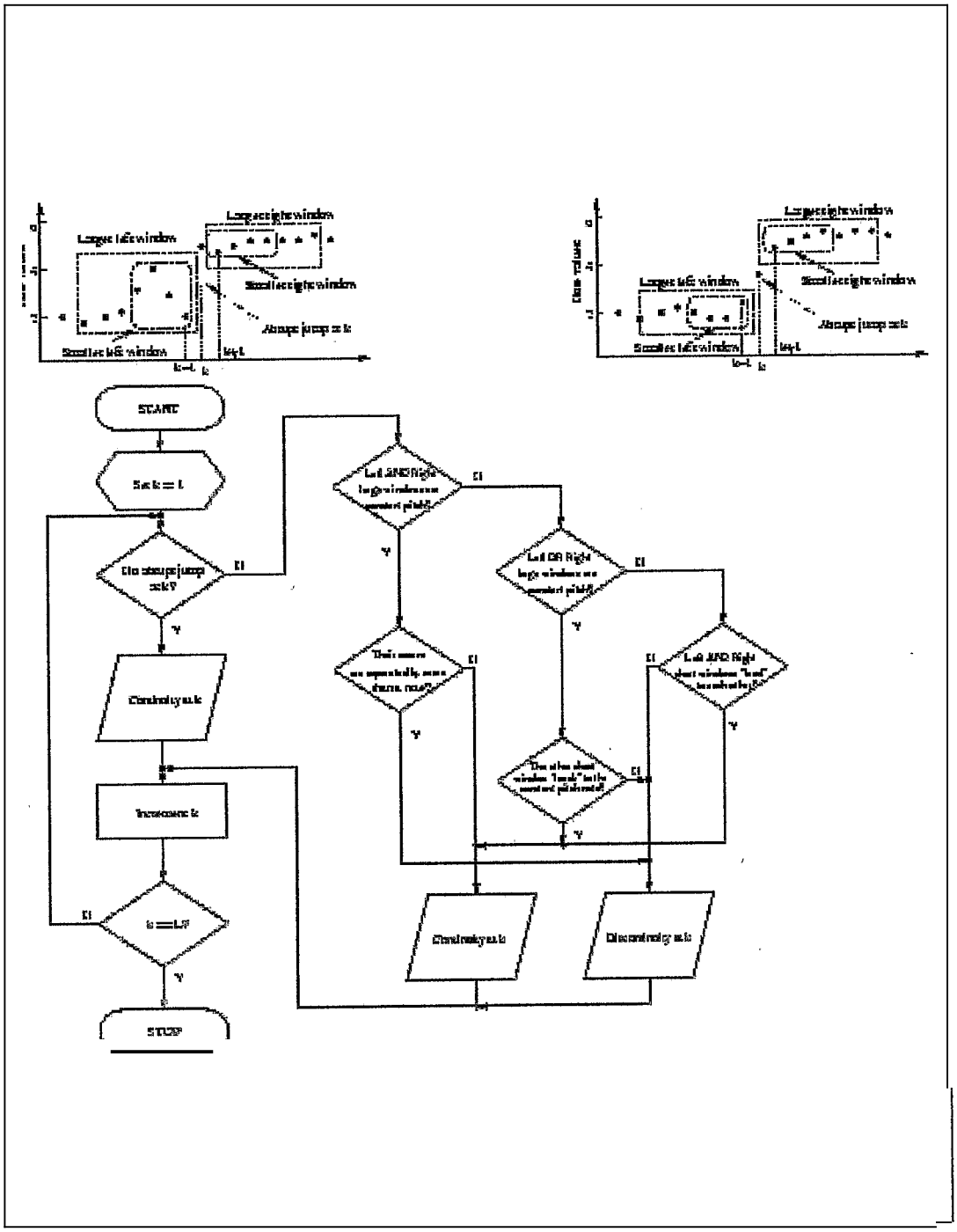
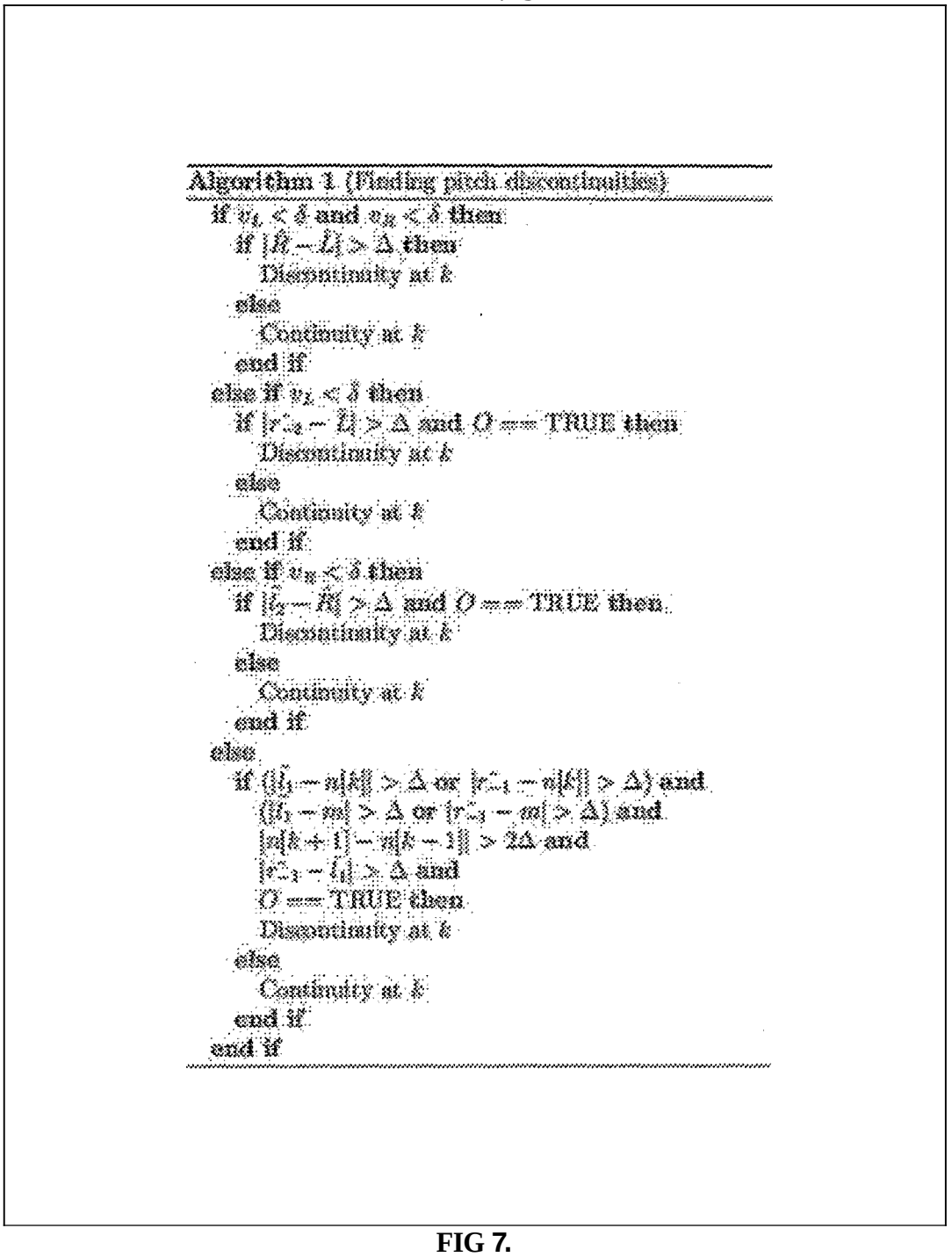

Larger left window
Larger right window
Smaller right window
Abrupt jump at k
Smaller left window
START
Set k = 1
Is abrupt jump at k?
Y
N
Check only at k
Increment k
k == L-1?
STOP

FIG *6.*



Algorithm 1 (Finding pitch discontinuities)

```
if e_L < δ and e_R < δ then
  if |R̂ − L̂| > Δ then
    Discontinuity at k
  else
    Continuity at k
  end if
else if e_L < δ then
  if |r̂_{-2} − L̂| > Δ and O == TRUE then
    Discontinuity at k
  else
    Continuity at k
  end if
else if e_R < δ then
  if |l̂_2 − R̂| > Δ and O == TRUE then
    Discontinuity at k
  else
    Continuity at k
  end if
else
  if (|l̂_1 − a[k]| > Δ or |r̂_{-1} − a[k]| > Δ) and
     (|l̂_1 − m| > Δ or |r̂_{-1} − m| > Δ) and
     |a[k+1] − a[k−1]| > 2Δ and
     |r̂_{-1} − L̂| > Δ and
     O == TRUE then
    Discontinuity at k
  else
    Continuity at k
  end if
end if
```

FIG 7.

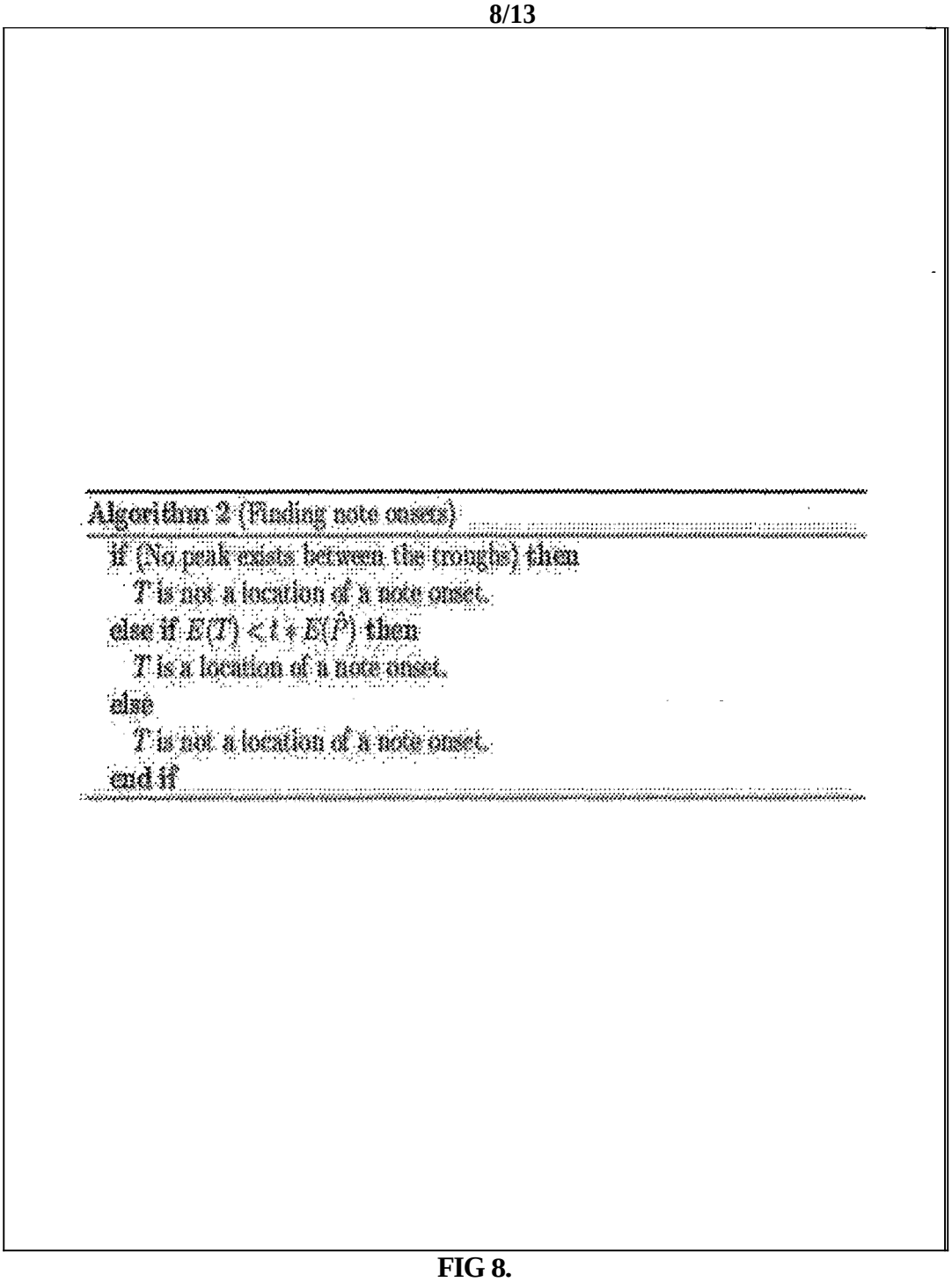

Algorithm 2 (Finding note onsets)
if (No peak exists between the troughs) then
T is not a location of a note onset.
else if E(T) < t * E(P) then
T is a location of a note onset.
else
T is not a location of a note onset.
end if

FIG 8.

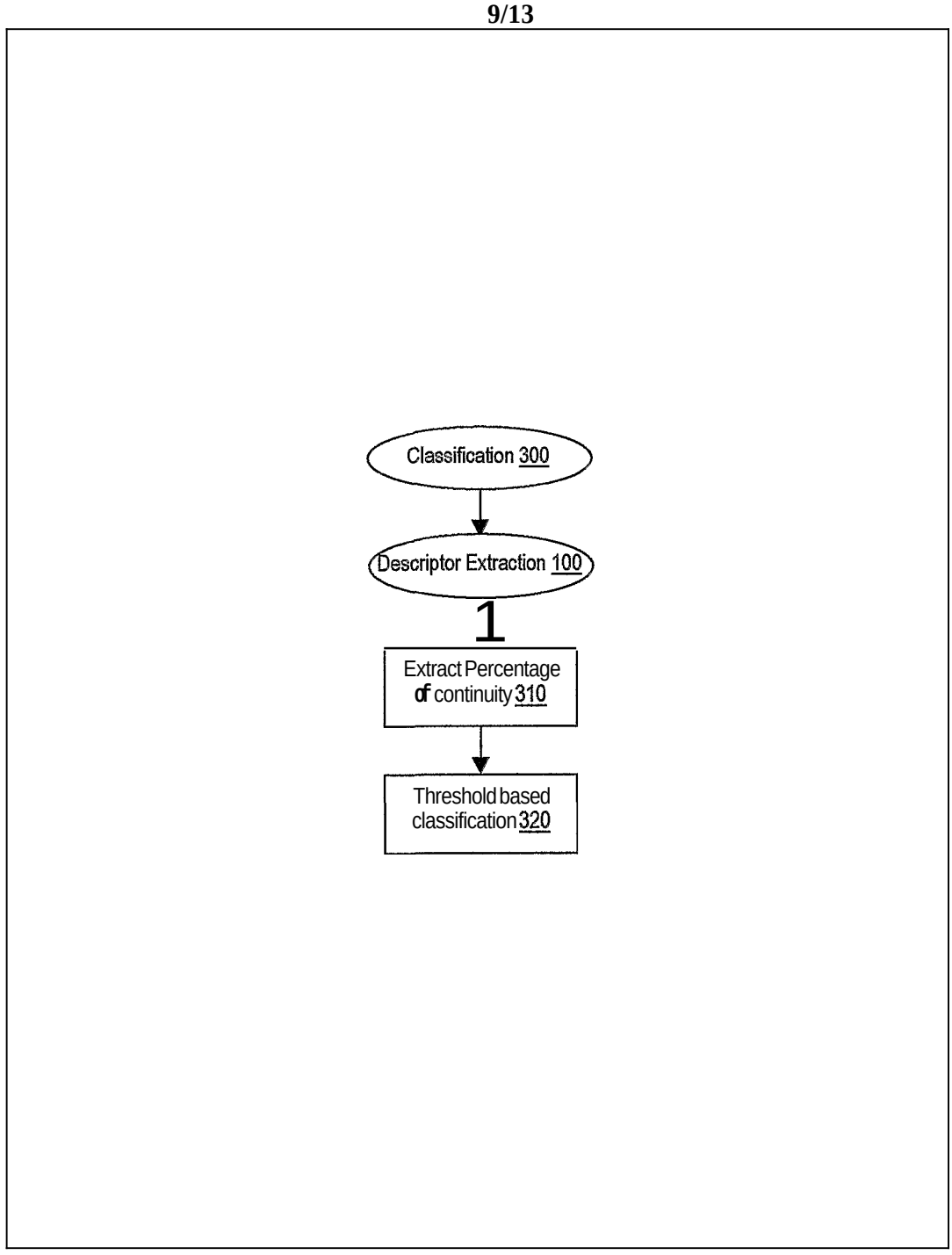

Classification 300
Descriptor Extraction 100
1
Extract Percentage of continuity 310
Threshold based classification 320

FIG 9.

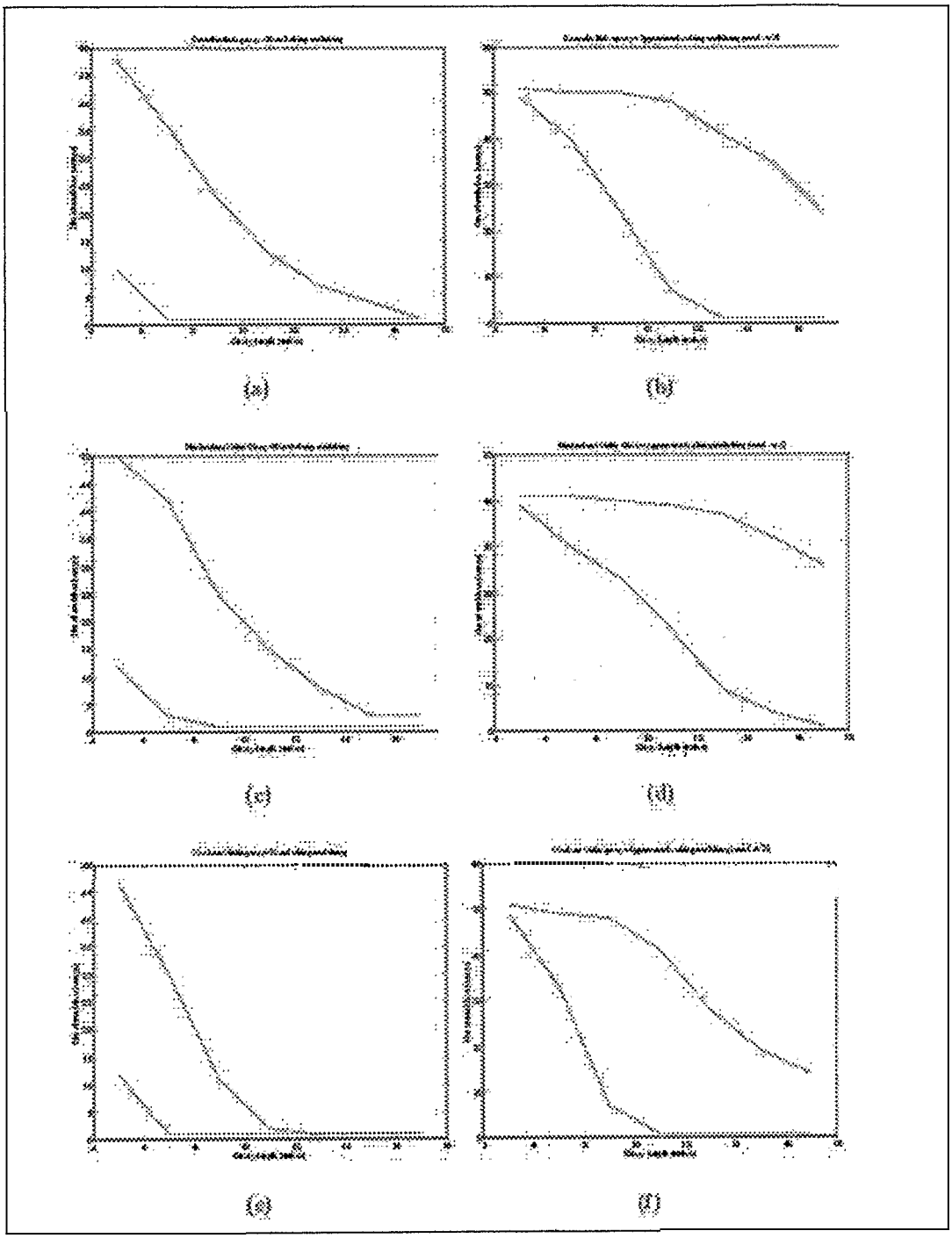


FIG 10.

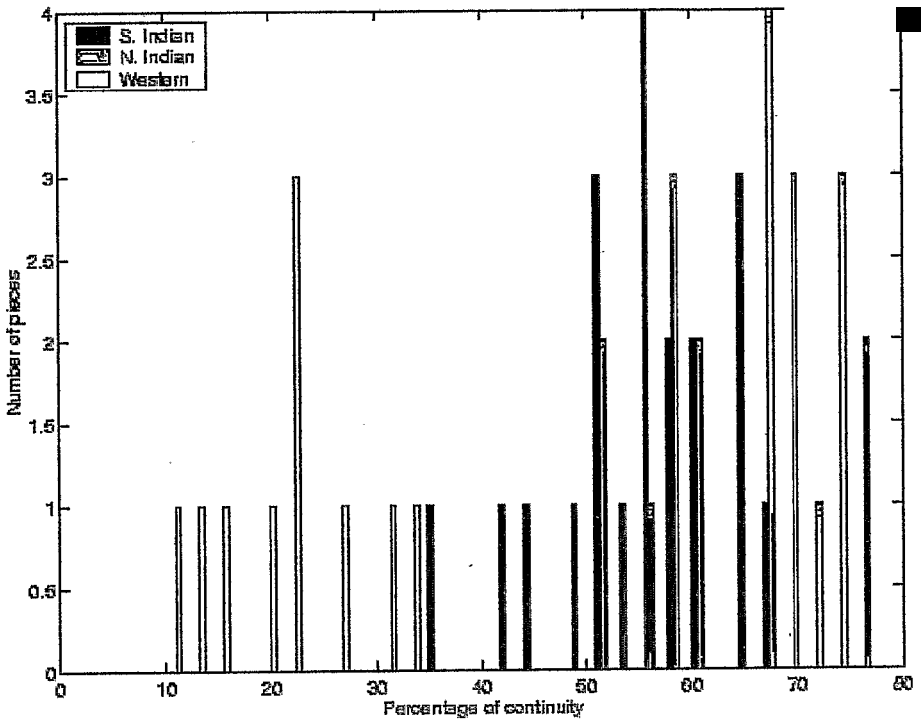

S. Indian
N. Indian
Western
Number of pieces
Percentage of continuity
0
0.5
1
1.5
2
2.5
3
3.5
4
0
10
20
30
40
50
60
70
80

FIG 11.

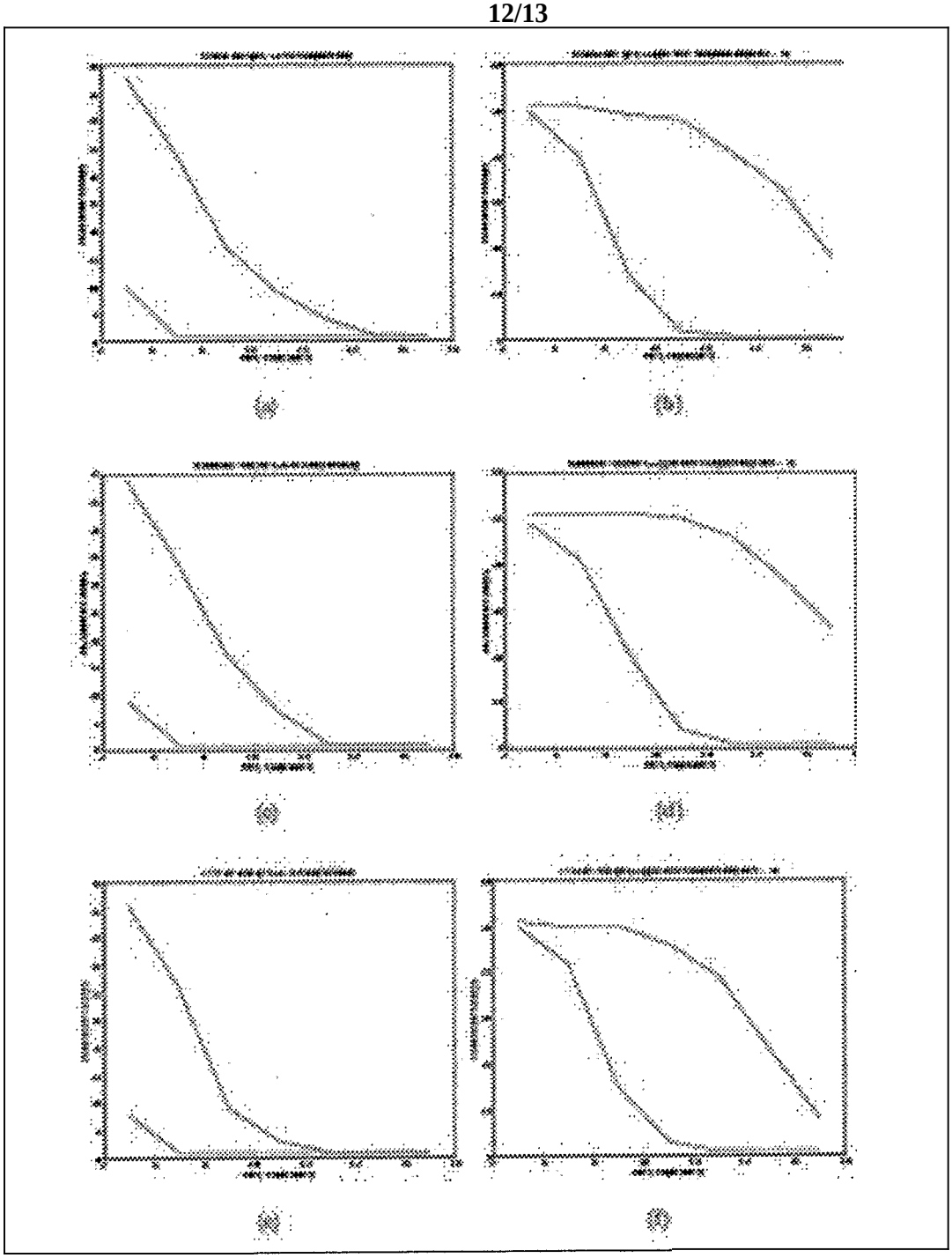


FIG 12.

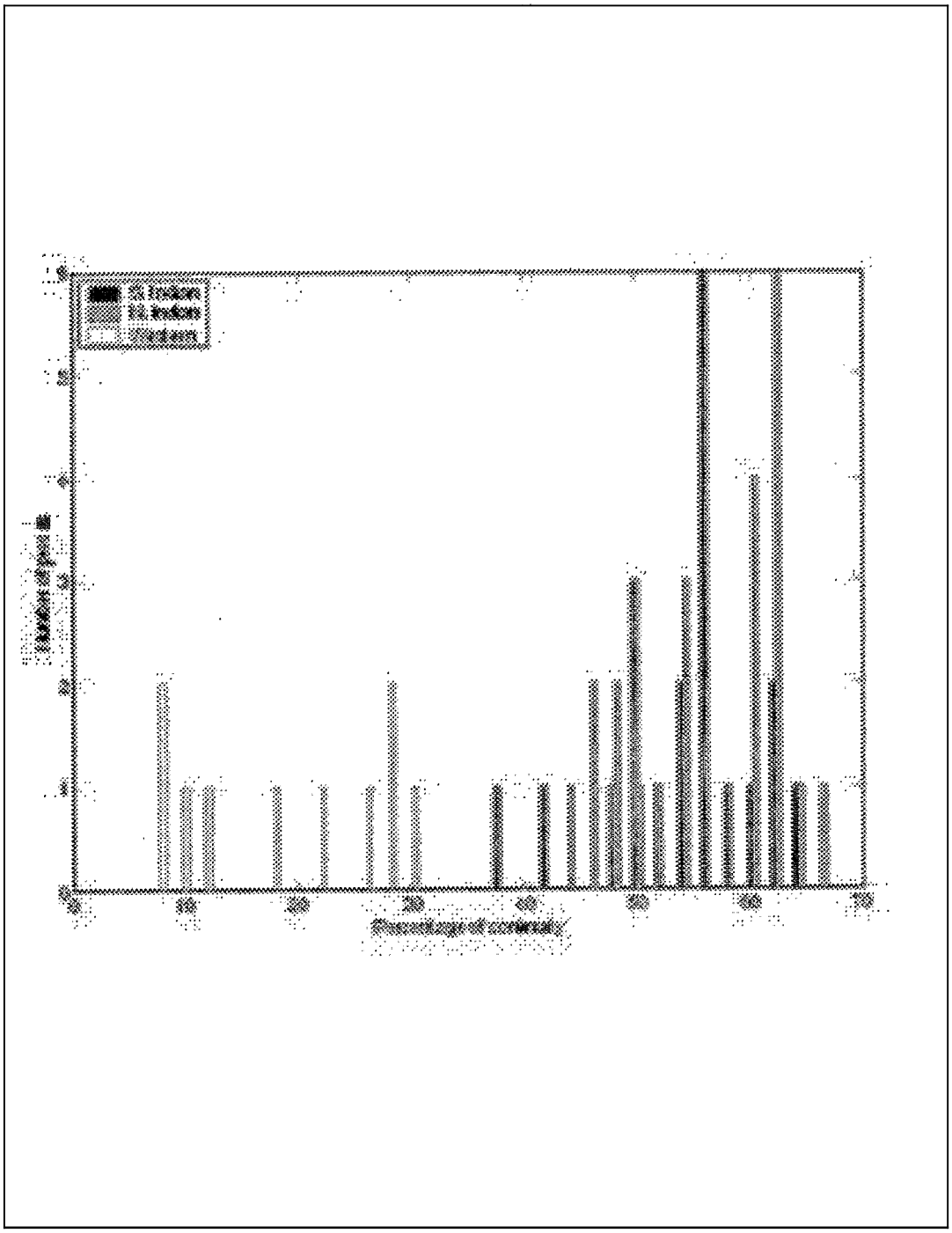


FIG 13.