WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

DOCUMENT MADE AVAILABLE UNDER THE PATENT COOPERATION TREATY (PCT)

International application number: **PCT/SE2015/050503**

International filing date: **07 May 2015 (07.05.2015)**

Document type: **Certified copy of priority document**

Document details:
Country/Office: **US**
Number: **61/990,354**
Filing date: **08 May 2014 (08.05.2014)**

Date of receipt at the International Bureau: **12 May 2015 (12.05.2015)**

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a),(b) or (b-*bis*)

34, chemin des Colombettes
1211 Geneva 20, Switzerland
www.wipo.int

Digital Access Service
(DAS)

WIPO
WORLD
INTELLECTUAL PROPERTY
ORGANIZATION

CERTIFICATE OF AVAILABILITY OF A CERTIFIED PATENT DOCUMENT IN A DIGITAL LIBRARY

The International Bureau certifies that a copy of the patent application indicated below has been available to the WIPO Digital Access Service since the date of availability indicated, and that the patent application has been available to the indicated Office(s) as of the date specified following the relevant Office code:

Document details: Country/Office: US
Filing date: 08 May 2014 (08.05.2014)
Application number: 61990354

Date of availability of document: 29 May 2014 (29.05.2014)

The following Offices can retrieve this document by using the access code:
AU, CN, ES, FI, GB, IB, JP, KR, SE, US, DK

Date of issue of this certificate: 12 May 2015 (12.05.2015)

34, chemin des Colombettes
1211 Geneva 20, Switzerland
www.wipo.int

1111

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

May 12, 2015

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE UNDER 35 USC 111.

APPLICATION NUMBER: ***61/990,354***
FILING DATE: ***May 08, 2014***

THE COUNTRY CODE AND NUMBER OF YOUR PRIORITY APPLICATION, TO BE USED FOR FILING ABROAD UNDER THE PARIS CONVENTION, IS ***US61/990,354***

Certified by

Michelle K. Lee

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

Electronic Acknowledgement Receipt

EFS ID:	18980267
Application Number:	61990354
International Application Number:	
Confirmation Number:	7828
Title of Invention:	Harmonic-Noise like Audio Signal Classifier (Discriminator)
First Named Inventor/Applicant Name:	Erik Norvell
Customer Number:	27045
Filer:	Steven Ware Smith/Kara Coffman
Filer Authorized By:	Steven Ware Smith
Attorney Docket Number:	P43568-US1
Receipt Date:	08-MAY-2014
Filing Date:	
Time Stamp:	15:42:35
Application Type:	Provisional

Payment information:

Submitted with Payment	yes
Payment Type	Deposit Account
Payment was successfully received in RAM	$260
RAM confirmation Number	1874
Deposit Account	501379
Authorized User	

The Director of the USPTO is hereby authorized to charge indicated fees and credit any overpayment as follows:

Charge any Additional Fees required under 37 C.F.R. Section 1.16 (National application filing, search, and examination fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.17 (Patent application and reexamination processing fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.19 (Document supply fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.20 (Post Issuance fees)

Charge any Additional Fees required under 37 C.F.R. Section 1.21 (Miscellaneous fees and charges)

File Listing:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
1		P43568-US1_2014-05-08_14-3145_Provisional_Transmittal.pdf	699947 30336376c52324921360d4a77b841932160eed61	yes	3

Multipart Description/PDF files in .zip description

Document Description	Start	End
Provisional Cover Sheet (SB16)	1	2
Authorization to access Appl. by Trilateral Office	3	3

Warnings:

Information:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
2	Specification	P43568-US1_2014-05-08_14-3145_Provisional_Application_RW_edited.pdf	310254 ab9f7cf8bbfe46c601fbee7a346e4ddc1beae540	no	11

Warnings:

Information:

Document Number	Document Description	File Name	File Size(Bytes)/ Message Digest	Multi Part /.zip	Pages (if appl.)
3	Fee Worksheet (SB06)	fee-info.pdf	29607 e6761d6533e28818f590cc7464531f6313a43580	no	2

Warnings:

Information:

Total Files Size (in bytes):	1039808

This Acknowledgement Receipt evidences receipt on the noted date by the USPTO of the indicated documents, characterized by the applicant, and including page counts, where applicable. It serves as evidence of receipt similar to a Post Card, as described in MPEP 503.

New Applications Under 35 U.S.C. 111

If a new application is being filed and the application includes the necessary components for a filing date (see 37 CFR 1.53(b)-(d) and MPEP 506), a Filing Receipt (37 CFR 1.54) will be issued in due course and the date shown on this Acknowledgement Receipt will establish the filing date of the application.

National Stage of an International Application under 35 U.S.C. 371

If a timely submission to enter the national stage of an international application is compliant with the conditions of 35 U.S.C. 371 and other applicable requirements a Form PCT/DO/EO/903 indicating acceptance of the application as a national stage submission under 35 U.S.C. 371 will be issued in addition to the Filing Receipt, in due course.

New International Application Filed with the USPTO as a Receiving Office

If a new international application is being filed and the international application includes the necessary components for an international filing date (see PCT Article 11 and MPEP 1810), a Notification of the International Application Number and of the International Filing Date (Form PCT/RO/105) will be issued in due course, subject to prescriptions concerning national security, and the date shown on this Acknowledgement Receipt will establish the international filing date of the application.

Harmonic-Noise like Audio Signal Classifier (Discriminator)

1 BACKGROUND

1.1 Technical background / Existing technology

Modern audio codecs consists of multiple compression schemes optimized for signals with different properties. With practically no exception, speech-like signals are processed with time-domain codec, while music signals are processed with transform-domain codec. Coding schemes that are supposed to handle both speech and music signals require a mechanism to recognize input signal (speech-music classifier) and switch between the appropriate codec modes. An overview illustration of a multimode audio codec using mode decision logic based on the input signal is shown in Figure 1.

In a similar manner among the class of music signals one can discriminate more noise like music signals and harmonic music signals, and build classifier and optimal coding scheme for each of these groups. This abstraction of creating a classifier to determine the class of a signal which then controls the mode decision is illustrated in Figure 2. These implementations deal with finding a better classifier for discrimination of harmonic and noise like music signals.

1.2 Problems with existing solutions

Variety of speech-music classifiers are used in the field of audio coding. However, these classifiers cannot discriminate between different classes in the space of music signals. Many classifiers do not provide enough resolution to discriminate between classes which are needed in a complex multimode codec.

2 BRIEF SUMMARY OF THE PROPOSED SOLUTION / ABSTRACT

The problem of harmonic and noise-like music segments discrimination is solved by a novel metric, calculated directly on the frequency-domain coefficients. The metric is based on the distribution of pre-selected spectral peaks candidates and the average peak-to-noise floor ratio.

3 ADVANTAGES OF THE PROPOSED SOLUTION

The proposed solution allows harmonic and noise-like music segments to be identified, which in turn allows for optimal coding of these signal types as illustrated in figure 3. This coding concept provides a superior quality over the conventional coding schemes.

4 DETAILED DESCRIPTION OF THE PROPOSED SOLUTION AND FIGURES

The exemplary audio codec, and therefore the presented audio discriminator/classifier, operate on short blocks (e.g. 20ms) of the input waveform. The presented embodiment show exemplary numerical values which are preferred for the embodiment at hand. It should be understood that these numerical values are given only as examples and may be adapted to the audio codec at hand.

Before performing spectral analysis, the input waveform may optionally be pre-emphasized with a first-order high-pass filter $H(z) = 1 - 0.68z^{-1}$

The discrete Fourier transform (DFT) is used to convert the signal into the frequency domain. The spectral analysis is performed once per frame using 256-point fast Fourier transform (FFT).

An FFT is performed on the windowed input signal to obtain one set of spectral parameters:

$$X(k) = \sum_{n=0}^{255} x(n) e^{-j2\pi\frac{kn}{256}}$$

Where $k = 0, \ldots, 255$, is an index of frequency coefficients, and n is an index of waveform samples. It should be noted that any length N of the transform may be used.

The discriminator requires knowledge of spectral peaks location. Spectral peaks are defined as coefficients with absolute value above an adaptive threshold, which is based on the ratio of peak and noise-floor envelopes.

The noise-floor estimation algorithm operates on the absolute values of transform coefficients $|X(k)|$. Instantaneous noise-floor energies $E_{NF}(k)$ are estimated according to the recursion:

$$E_{NF}(k) = \alpha E_{NF}(k) + (1 - \alpha) |X(k)|^2$$

$$\alpha = \begin{cases} 0.9578 & if \quad |X(k)|^2 > E_{NF}(k-1) \\ 0.6472 & if \quad |X(k)|^2 \leq E_{NF}(k-1) \end{cases}$$

The particular form of the weighting factor α minimizes the effect of high-energy transform coefficients and emphasizes the contribution of low-energy coefficients. Finally the noise-floor level E_{NF} is estimated by simply averaging the instantaneous energies E_{NF}.

$$E_{NF} = \sum_{k=0}^{255} E_{NF}(k)$$

The peak-picking algorithm requires knowledge of noise-floor level and average level of spectral peaks. The peak energy estimation algorithm is similar to the noise-floor estimation algorithm, but instead of low-energy, it tracks high-spectral energies.

$$E_p(k)=\beta E_p(k)+(1-\beta)|X(k)|^2$$

$$\beta=\begin{cases}0.4223 & if \quad |X(k)|^2 > E_p(k-1)\\ 0.8029 & if \quad |X(k)|^2 \le E_p(k-1)\end{cases}$$

In this case the weighting factor β minimizes the effect of low-energy transform coefficients and emphasizes the contribution of high-energy coefficients. The overall peak energy $\bar{E}_p$ is estimated by simply averaging the instantaneous energies.

$$\bar{E}_p=\sum_{k=1}^{N} E_p(k)$$

When the peak and noise-floor levels are calculated, a threshold level τ is formed,

$$\tau=\left(\frac{\bar{E}_p}{\bar{E}_{nf}}\right)^{\gamma}\bar{E}_{nf}$$

with $\gamma = 0.88379$. Transform coefficients are compared to the threshold, and the ones with amplitude above it, form a vector of peak candidates.

An alternative threshold value, which may require less computational complexity, is to use the instantaneous peak envelope level with a fixed scaling factor

$$\theta(k) = E_p(k)\cdot 0.64$$

The peak candidates will then be all the coefficients with the squared amplitude above the instantaneous threshold level

$$\begin{cases}|X(k)|^2 > \theta(k), k \in P\\ |X(k)|^2 \le \theta(k), k \notin P\end{cases}$$

where P denotes the frequency ordered set of positions of index candidates. Considering the FFT spectrum, some peaks will be broad and consist of several transform coefficients, while others are narrow and are represented by a single coefficient. To get a representation of individual coefficients, consecutive peak candidates are assumed to be part of a broader peak. By finding the maximum squared amplitude $|X(k)|^2$ of the transform coefficients in a range of consecutive peak positions $\ldots, k-1, k, k+1, \ldots$, a refined set $\hat{P}$ is created where the broad peaks are represented by the maximum position in each range. Figure 4 illustrates the derivation of the peak envelope and noise floor, and the peak selection algorithm.

The above calculations serve to generate features that are used for forming the classifier decision: an estimate of the peak scarcity S and a peak-to-noise floor ratio PNR. The peak scarcity may be represented using the average distance d_i between peaks

$$S = \begin{cases} \frac{1}{N_S} \sum_{i=1}^{N_S} d_i, & N_S > 1 \\ 0, & N_S \leq 1 \end{cases}$$

where N_S is the number of refined peaks in the set P. The PNR may be calculated as

$$PNR = \frac{E_P}{E_{NP}}$$

The classifier decision may be formed using these features in combination with a decision threshold. We can name these decisions

$$issparse = S > S_{THR}$$

$$isclean = PNR > PNR_{THR}$$

The outcome of these decisions may be used to form different classes of signals. An illustration of these classes is shown in Figure 6. Since the classification is based on two binary decisions, the total number of classes may be at most 4. As a next step, the codec decision can be formed using the class information, which is illustrated in Table 1.

Table 1: Possible classes formed using two feature decisions.

	isclean	***Issparse***
Class A	false	False
Class B	true	False
Class C	true	True
Class D	false	true

In the following step in the audio codec, the decision is to be made which processing steps to apply to which class. This mapping will depend on the characteristics and capabilities of the different coding modes or processing steps available. As an example, perhaps Codec mode 1 would handle Class A and Class C, while Codec mode 2 would handle Class B and Class D. The coding mode decision can be the final output of the classifier to guide the encoding process. The coding mode decision would typically be transferred in the bitstream together with the codec parameters from the chosen coding mode.

It should be understood that the above classes may be further combined with other classifier decisions. The combination may result in a larger number of classes, or they may be combined using a priority order such that the presented classifier may be overruled by another classifier, or vice versa that the presented classifier may overrule another classifier.

The described implementation is a high-resolution music type discriminator with application audio coding. The decision logic of the discriminator is based on statistics of positional distribution of frequency coefficients with prominent energy.

5 GENERAL

An overview of the discriminator can be seen in Figures 3 and 7.

Thus, Figure 3 is a schematic block diagram of a discriminator according to embodiments. The discriminator comprises:

- an input unit (not shown) configured to receive an input signal representing an audio signal to be handled,

- a Framing unit for producing frames of the input signal,

- an optional Pre-emphasis unit for filtering a frame of the input signal with a high-pass filter,

- a Frequency transforming unit for performing a Fourier transform on the output of the Pre-emphasis unit (or the output of the Framing unit if the Pre-emphasis unit is omitted) to obtain frequency coefficients,

- a Peak/Noise envelope analysis unit for performing an envelope analysis on the frequency coefficients to calculate a threshold value,

- a Peak candidate selection unit for selecting peak candidates from among the frequency coefficients based on the threshold value,

- a Peak candidate refinement unit for refining the selection of peak candidates to account for broad peaks in the Fourier transform spectrum to produce a set of selected peaks,

- a Feature calculation unit for generating features (e.g., peak scarcity and peak-to-noise floor ratio PNR) based on the set of selected peaks,

- a Class decision unit for making a classification decision based on the features,

- a Coding mode decision unit for determining a coding mode based on the classification decision,

- a Multi-mode encoder unit for encoding frames of the audio signal based on the determined coding mode,

- a Bit-streaming/Storage unit for streaming and/or storing the encoded frames, and

- an output unit (not shown) for outputting the encoded frames of the audio signal.

All these units could be implemented in hardware. There are numerous variants of circuitry elements that can be used and combined to achieve the functions of the units of the encoder. Such variants are encompassed by the embodiments. Particular examples of hardware implementation of the discriminator are implementation in digital signal processor (DSP) hardware and integrated circuit technology, including both general-purpose electronic circuitry and application-specific circuitry.

The discriminator described herein could alternatively be implemented e.g. by one or more of a processor and adequate software with suitable storage or memory therefore, in order to perform the discriminatory action on an input signal, according to the embodiments described herein, see Fig. 7. The incoming signal is received by an input (IN), to which the processor and the memory are connected, and the discriminatory representation of an audio signal (parameters) obtained from the software is outputted from the output (OUT).

The technology described above may be used e.g. in a sender, which can be used in a mobile device (e.g. mobile phone, laptop) or a stationary device, such as a personal computer.

It is to be understood that the choice of interacting units or modules, as well as the naming of the units are only for exemplary purpose, and may be configured in a plurality of alternative ways in order to be able to execute the disclosed process actions.

It should also be noted that the units or modules described in this disclosure are to be regarded as logical entities and not with necessity as separate physical entities. It will be appreciated that the scope of the technology disclosed herein fully encompasses other embodiments which may become obvious to those skilled in the art, and that the scope of this disclosure is accordingly not to be limited.

Reference to an element in the singular is not intended to mean "one and only one" unless explicitly so stated, but rather "one or more." All structural and functional equivalents to the elements of the above-described embodiments that are known to those of ordinary skill in the art are expressly incorporated herein by reference and are intended to be encompassed hereby. Moreover, it is not necessary for a device or method to address each and every problem sought to be solved by the technology disclosed herein, for it to be encompassed hereby.

In the preceding description, for purposes of explanation and not limitation, specific details are set forth such as particular architectures, interfaces, techniques, etc. in order to provide a thorough understanding of the disclosed technology. However, it will be apparent to those skilled in the art that the disclosed technology may be practiced in other embodiments and/or combinations of embodiments that depart from these specific details. That is, those skilled in the art will be able to devise various arrangements which, although not explicitly described or shown herein, embody the principles of the disclosed technology. In some instances, detailed descriptions of well-known devices, circuits, and methods are omitted so as not to obscure the description of the disclosed technology with unnecessary detail. All statements herein reciting principles, aspects, and embodiments of the disclosed technology, as well as specific examples thereof, are intended to encompass both structural and functional equivalents thereof. Additionally, it is intended that such equivalents include both currently known equivalents as well as equivalents developed in the future, e.g. any elements developed that perform the same function, regardless of structure.

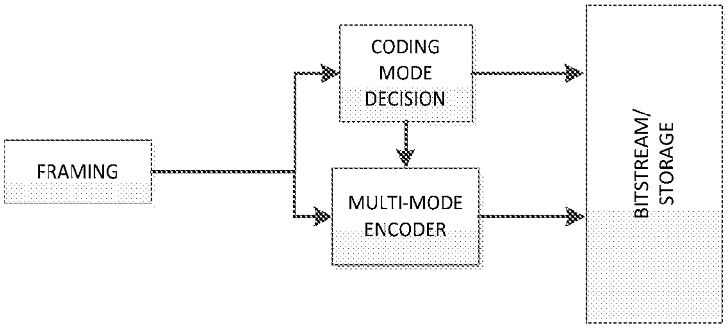
Thus, for example, it will be appreciated by those skilled in the art that the figures herein can represent conceptual views of illustrative circuitry or other functional units embodying the principles of the technology, and/or various processes which may be substantially represented in computer readable medium and executed by a computer or processor, even though such computer or processor may not be explicitly shown in the figures.

The functions of the various elements including functional blocks may be provided through the use of hardware such as circuit hardware and/or hardware capable of executing software in the form of coded instructions stored on computer readable medium. Thus, such functions and illustrated functional blocks are to be understood as being either hardware-implemented and/or computer-implemented, and thus machine-implemented.

The embodiments described above are to be understood as a few illustrative examples of the present invention. It will be understood by those skilled in the art that various modifications, combinations and changes may be made to the embodiments without departing from the scope of the present invention. In particular, different part solutions in the different embodiments can be combined in other configurations, where technically possible.

FIGURE 1

Signal discriminator in an audio codec. Usage of the discriminators output to switch between different coding modes.



PTO/SB/16 (10-07)
Approved for use through 06/30/2010. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

PROVISIONAL APPLICATION FOR PATENT COVER SHEET – Page 1 of 2

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53(c).

Express Mail Label No. Transmitted Via EFS-Web

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Erik	Norvell	Stockholm, Sweden
Volodya	Grancharov	Solna, Sweden

Additional inventors are being named on the ______ separately numbered sheets attached hereto

TITLE OF THE INVENTION (500 characters max):

Harmonic-Noise like Audio Signal Classifier (Discriminator)

Direct all correspondence to: **CORRESPONDENCE ADDRESS**

☑ The address corresponding to Customer Number: 27045

OR

☐ Firm or Individual Name

Address

City	State	Zip
Country	Telephone	Email

ENCLOSED APPLICATION PARTS *(check all that apply)*

☐ Application Data Sheet. See 37 CFR 1.76
☐ CD(s), Number of CDs ______
☐ Drawing(s) *Number of Sheets* ______
☐ Other (specify) ______
☑ Specification (e.g. description of the invention) *Number of Pages* 11

Fees Due: Filing Fee of $250 ($105 for small entity). If the specification and drawings exceed 100 sheets of paper, an application size fee is also due, which is $310 ($130 for small entity) for each additional 50 sheets or fraction thereof. See 35 U.S.C. 41(a)(1)(G) and 37 CFR 1.16(s).

METHOD OF PAYMENT OF THE FILING FEE AND APPLICATION SIZE FEE FOR THIS PROVISIONAL APPLICATION FOR PATENT

☐ Applicant claims small entity status. See 37 CFR 1.27.
☐ A check or money order is enclosed to cover the filing fee and application size fee (if applicable).
☐ Payment by credit card. Form PTO-2038 is attached
☑ The Director is hereby authorized to charge the filing fee and application size fee (if applicable) or credit any overpayment to Deposit Account Number: 50-1379. A duplicative copy of this form is enclosed for fee processing.

260.00
TOTAL FEE AMOUNT ($)

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

PROVISIONAL APPLICATION COVER SHEET
Page 2 of 2

PTO/SB/16 (10-07)
Approved for use through 06/30/2010. OMB 0651-0032
U.S. Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☑ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: ____________________

WARNING:

Petitioner/applicant is cautioned to avoid submitting personal information in documents filed in a patent application that may contribute to identity theft. Personal information such as social security numbers, bank account numbers, or credit card numbers (other than a check or credit card authorization form PTO-2038 submitted for payment purposes) is never required by the USPTO to support a petition or an application. If this type of personal information is included in documents submitted to the USPTO, petitioners/applicants should consider redacting such personal information from the documents before submitting them to the USPTO. Petitioner/applicant is advised that the record of a patent application is available to the public after publication of the application (unless a non-publication request in compliance with 37 CFR 1.213(a) is made in the application) or issuance of a patent. Furthermore, the record from an abandoned application may also be available to the public if the application is referenced in a published application or an issued patent (see 37 CFR 1.14). Checks and credit card authorization forms PTO-2038 submitted for payment purposes are not retained in the application file and therefore are not publicly available.

SIGNATURE /Ronald J. Ward, Reg No 54,870/

Date May 8, 2014

TYPED or PRINTED NAME Ronald J. Ward

REGISTRATION NO. 54,870
(if appropriate)

TELEPHONE 972-583-8656

Docket Number: **P43568-US1**

6 ABBREVIATIONS

Abbreviation	Explanation
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
MDCT	Modified Discrete Cosine Transform
ACELP	Code-Excited Linear Prediction
VQ	Vector Quantization
PNR	Peak to Noise floor ratio

FIGURE 2

Signal discriminator in an audio codec, on the path higher level discriminator (speech/music discriminator)

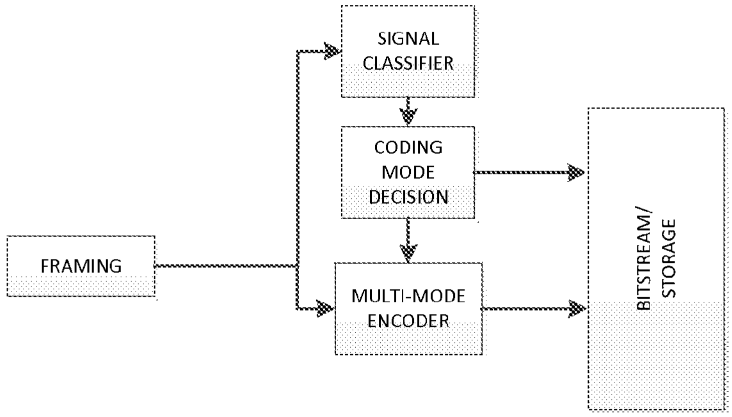


FIGURE 3

Block diagram of the proposed signal discriminator, including the following modules: framing, pre-emphasis, frequency transform, calculating threshold for spectral peaks, calculating metric on the positional distribution of spectral peaks, decision module

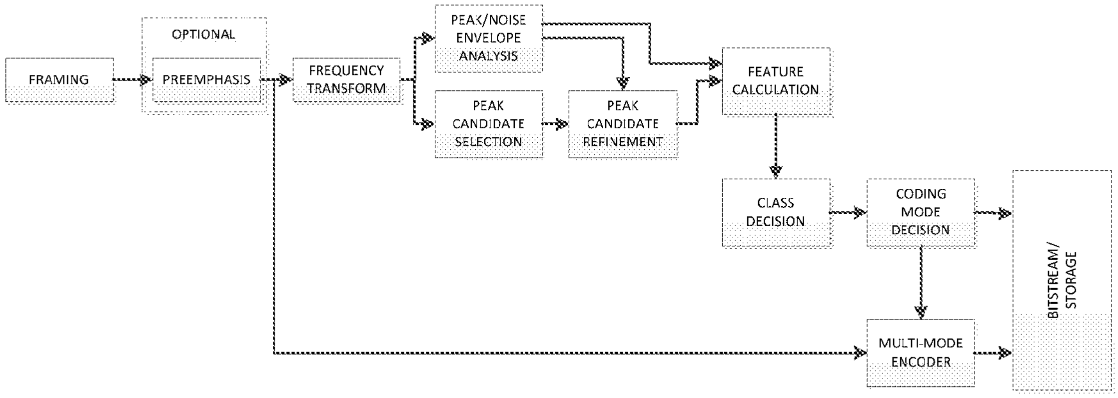


FIGURE 4

Illustration of peak selection algorithm and instantaneous

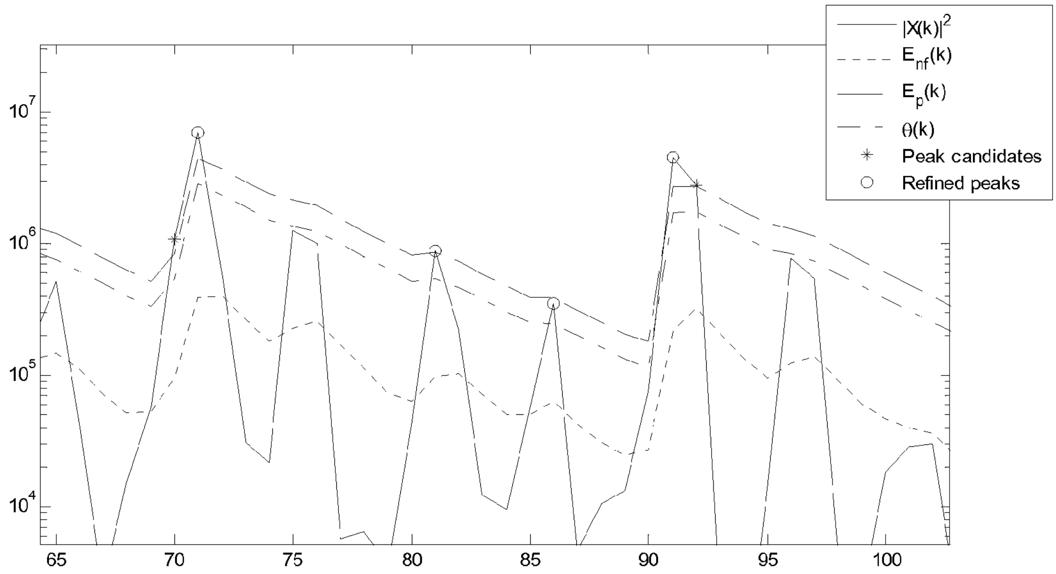


FIGURE 5

Illustration of peak distances *d*d_i

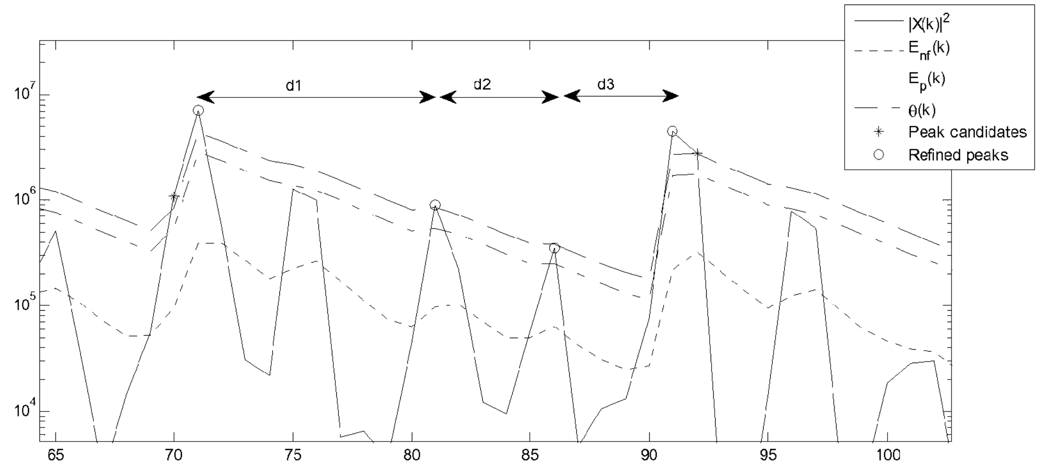
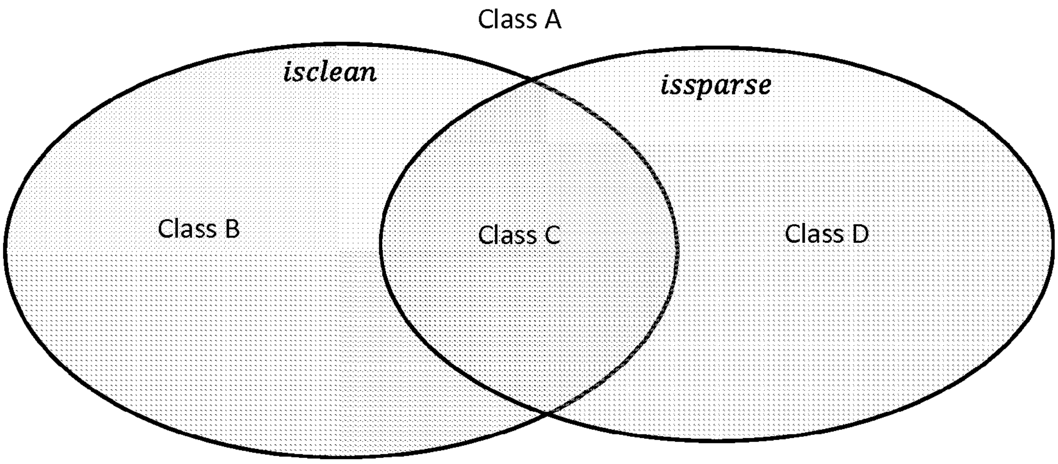


FIGURE 6
Example of class decisions as a Venn diagram

$$issparse = S > S_{THR}$$
$$isclean = PNR > PNR_{THR}$$



Discriminator

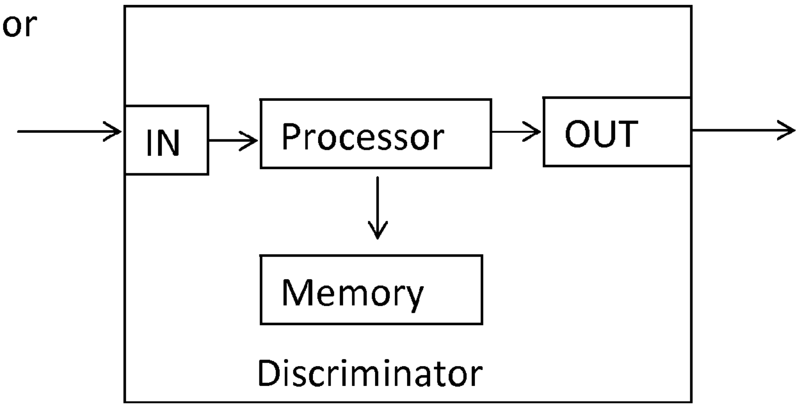


Figure. 7