

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
21 September 2006 (21.09.2006)

PCT

(10) International Publication Number  
**WO 2006/097299 A1**

(51) International Patent Classification:  
**G10L 11/00** (2006.01)

(21) International Application Number:  
PCT/EP2006/002398

(22) International Filing Date: 15 March 2006 (15.03.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
05005994.8 18 March 2005 (18.03.2005) EP

(71) Applicant (for all designated States except US): **SONY DEUTSCHLAND GMBH** [DE/DE]; Hugo-Eckener-Strasse 20, 50829 Köln (DE).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **KEMP, Thomas** [DE/DE]; Kleine Huelengasse 9, 73733 Esslingen (DE). **LAM, Yin, Hay** [GB/DE]; Marktstrasse 56/1, 70372 Stuttgart (DE). **RIGUEIRO, Marta, Tolós** [ES/ES]; C/Congres 10, E-08340 Vilassar de Mar (ES).

(74) Agent: **MÜLLER, Frithjof, E.**; Müller . Hoffmann & Partner, Innere Wiener Strasse 17, 81667 München (DE).

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

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

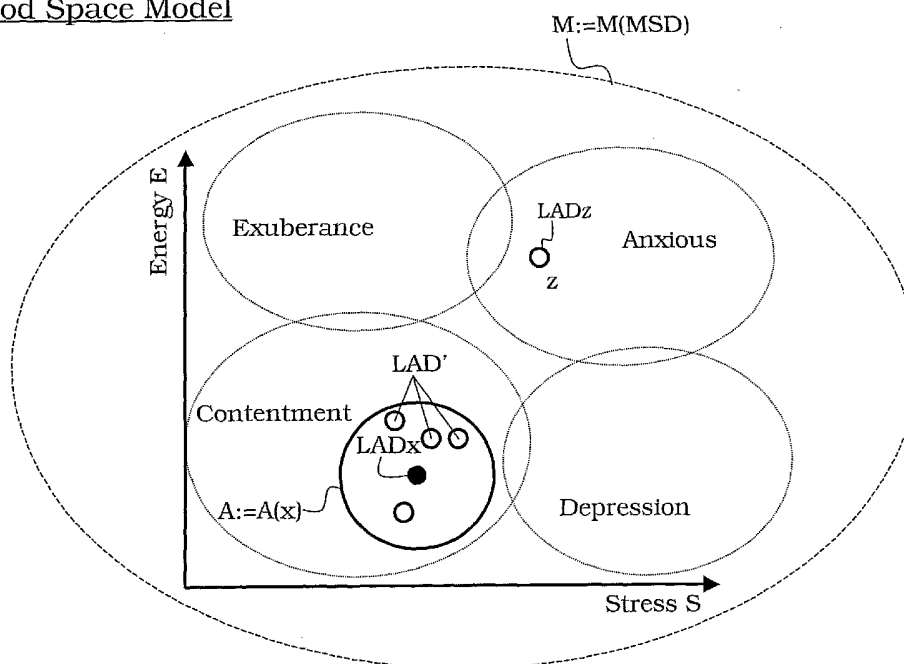
Published:

— with international search 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.

(54) Title: METHOD FOR CLASSIFYING AUDIO DATA

### Mood Space Model



(57) Abstract: A method for classifying audio data (AD) is provided. For a given piece of audio data (AD) a location or position (LAD) for said given audio data (AD) within a mood space (M) is generated and compared to a comparison mood space location (CL). As a result of the comparison comparison data (CD) are generated and provided as a classification result with respect to said given audio data (AD).

WO 2006/097299 A1

### Method for Classifying Audio Data

The present invention relates to a method for classifying audio data. The present invention more particularly relates to a fast music similarity computation method based on e.g. N-dimensional music mood space relationships.

5 Recently, the classification of audio data and in particular of pieces of music becomes more and more important as many electronic devices and in particular customer devices enable a respective user to store and manage a large plurality of music items and titles. In order to enhance the managing mechanism for such music data basis it is necessary to obtain a comparison between  
10 different pieces of audio data or different pieces of music in an easy and fast manner.

Therefore, a variety of mechanisms have been developed in order to extract from an analysis of audio data particular properties and features in order to  
15 compare pieces of music by comparing the respective sets or n-tuples of properties and features. However, many of the known features to be evaluated within such a comparison mechanism are difficult to calculate and the computational burden is in some cases not reasonable.

20 It is an object underlying the present invention to provide a method for classifying audio data which enables a reliable and easy and fast to compute comparison and classification of audio data.

The object is achieved according to the present invention by a method for  
25 classifying audio data with the features of independent claim 1. Preferred embodiments of the invention method for classifying audio data are within the scope of the dependent subclaims. The object underlying the present invention is also achieved by an apparatus for classifying audio data, by a computer program product, as well as by a computer readable storage medium according  
30 to independent claims 18, 19 and 20, respectively.

The method for classifying audio data according to the present invention comprises a step (S1) of providing audio data in particular as input data, a step (S2) of providing mood space data which define and/or which are descriptive or representative for a mood space according to which audio data can be  
35 classified, a step (S3) of generating a mood space location within said mood space for said given audio data, a step (S4) of providing at least one compari-

son mood space location within said mood space, a step (S5) of comparing said mood space location for said given audio data with said at least one comparison mood space location and thereby generating comparison data, and a step (S6) of providing as a classification result said comparison data in particular  
5 as output data which can be used in subsequent classification steps, mainly in detailed comparison steps.

It is therefore a key idea of the present invention to obtain from an analysis of given audio data a position or location within a mood space wherein said mood  
10 space is pre-defined or given by mood space data. Then the given audio data can be classified or compared by comparing the derived mood space location for said given audio data with said at least one comparison mood space location. The thereby generated comparison data or classification data are provided as a classification result or a comparison result. It is therefore essential  
15 to have for a given piece of audio data a position or location, e. g. by means of coordinate n-tuple, which can easily compared with other locations or positions in said mood space, e. g. by simply comparing the respective coordinates of the position or location. Therefore audio data can easily be classified and compared with other audio data.

20 According to a preferred embodiment of the method for classifying audio data according to the present invention said mood space may be or may be modelled by at least one of an Euclidean space model, a Gaussian mixture model, a neural network model, and a decision tree model.

25 Additionally or alternatively, according to a further preferred embodiment of the method for classifying audio data according to the present invention said mood space may be or may be modelled by an N-dimensional space or manifold and N may be a given and fixed integer.

30 Further additionally or alternatively, said comparison data may be alternatively or additionally at least one of being descriptive for, being representative for and comprising at least one of a topology, a metric, a norm, a distance defined in or on said mood space according to a another embodiment of the  
35 method for classifying audio data according to the present invention.

Additionally or alternatively, said comparison data and in particular said topology, metric, norm, and said distance may be obtained based on at least one of said Euclidean space model, said Gaussian mixture model, said neural

network model, and said decision tree model according to an advantageous embodiment of the method for classifying audio data according to the present invention.

5 Said comparison data may be derived based on said mood space location within said mood space for said given audio data and they may be based on said comparison mood space location within said mood space according to an additional or alternative embodiment of the method for classifying audio data according to the present invention.

10

Said mood space and/or the model thereof may be defined based on Thayer's music mood model according to an additional or alternative embodiment of the method for classifying audio data according to the present invention.

15 According to a further preferred embodiment of the method for classifying audio data according to the present invention said mood space and/or the model thereof may be at least two-dimensional and may be defined based on the measured or measurable entities stress  $S()$  describing positive, e.g. happy, and negative, e.g. anxious moods and energy  $E()$  describing calm and energetic moods as emotional or mood parameters or attributes.

20

Further additionally or alternatively, according to a still further preferred embodiment of the method for classifying audio data according to the present invention said mood space and/or the model thereof are at least three-dimensional and are defined based on the measured or measurable entities for happiness, passion, and excitement.

25

Said step (S4) of providing said at least one comparison mood space location may additionally or alternatively comprise a step of providing at least one additional audio data in particular as additional input data and a step of generating a respective additional mood space location for said additional audio data, and wherein said respective additional mood space location for said additional audio data is used for said at least one comparison mood space location according to an additional or alternative embodiment of the method for classifying audio data according to the present invention.

30

35

At least two samples of audio data may be compared with respect to each other – one of said samples of audio data being assigned to said derived mood space location and the other one of said of audio data being assigned to said

additional mood space location or said comparison mood space location – in particular by comparing said derived mood space location and said additional mood space location or said comparison mood space location.

- 5 Further additionally or alternatively, according to a still further preferred embodiment of the method for classifying audio data according to the present invention said at least two samples of audio data to be compared with respect to each other may be compared with respect to each other based on said comparison data in a pre-selection process or comparing pre-process and then  
10 based on additional features, e. g. based on features more complicated to calculate and/or based on frequency domain related features, in a more detailed comparing process.

- 15 In this case said at least two samples of audio data to be compared with respect to each other may be compared with respect to each other in said more detailed comparing process based on said additional features, if said comparison data obtained from said pre-selection process or comparing pre-process are indicative for a sufficient neighbourhood of said at least two samples of audio data.

- 20 Alternatively, a plurality of more than two samples of audio data may be compared with respect to each other.

- 25 Alternatively or additionally, said given audio data may be compared to a plurality of additional samples of audio data.

- 30 In these cases from said comparison a comparison list and in particular a play list may be generated which is descriptive for additional samples of audio data of said plurality of additional samples of audio data which are similar to said given audio data.

- 35 According to a further preferred and advantageous embodiment of the method for classifying audio data according to the present invention music pieces are used as samples of audio data

According to a further aspect of the present invention, an apparatus for classifying audio data is provided which is adapted and which comprises means for carrying out a method for classifying audio data according to the present invention and the steps thereof.

According to a further aspect of the present invention a computer program product is provided comprising computer program means, which is adapted to realize the method for classifying audio data according to the present invention and the steps thereof, when it is executed on a computer or a digital  
5 signal processing means.

Additionally a computer readable storage medium is provided which comprises a computer program product according to the present invention.

10 These and further aspects of the present invention will be further discussed in the following:

#### Concept

15 The present invention inter alia relates to a fast music similarity computation method which is in particular based on a N-dimensional music mood space.

It is proposed that a N-dimensional music mood space can be used to limit the number of candidates and hence reduce the computation in similarity list  
20 generation. For each of the music piece in a huge database, its location in a N-dimensional music mood space is first determined and only music pieces which are close to the music in the mood space are selected and the similarity are computed between the given music and the pre-selected music pieces.

#### 25 Background

Music similarity is a relatively new topic, and at this moment, the interest into it is quite academic. Systems have been developed that compare music pieces with one another using statistics over what is called 'timbre' – a mixture of a  
30 variety of low-level features. Various distance measures have been proposed including expensive methods like Monte-Carlo-simulation of samples of a distribution and probability estimation of the artificial samples using the statistics from the other music piece. See e.g. [3] for details.

35 The state of the art in emotion recognition in music is a rather new topic. While a huge amount of papers have been written about music processing in general, few papers have been published regarding emotion in music. State of the art system used for emotion classification in music classifiers include Gaussian mixtures models, support vector machines, neural networks etc.

There are also studies about perception of emotion in music, but the results are still very preliminary. Reference [1] and [2] provides information about the state-of-the art mood detection techniques.

## 5 Problem

For applications which involved music retrieval or music suggestion, a music play list is usually displayed and songs in the play list are usually based on the similarity between the query music and the rest of the music in the data-  
10 base. Nowadays, typical commercial music database consists of hundreds of thousands of music. For each of the music in the database, state-of-the-art system usually compute its similarity to all the other music pieces in the database to generate a similarity list. Based on the applications, a play list is then generated from the similarity list. The computation required in similarity  
15 generation involved about  $N*N/2$  similarity measure computation, where  $N$  is the number of songs in the database. For example, if the number of songs in the database is 500,000, then the computation will be  $500,000*500,000/2$ , which is not practical for real applications.

20 In this proposal, a fast music similarity list generation method based on mood space are proposed. The emotion expressed in different music are usually different. Some music are perceived as happy by the listeners, but the other songs might be perceived as sad. On the other hand, among songs with similar mood or emotion, listeners generally can distinguish the difference in the  
25 degree of emotion expression. For example, one music is happier than the other one, etc. In additional, music with different mood usually are considered as dissimilar. The music similarity list generation approach described in this invention proposal exploits such emotion perception as described above.

30 In this proposal, we first proposed that the emotion of music can be described by a  $N$ -dimensional mood space. Each dimension describes the extent of a particular emotion attribute. For each of the music in the database, the value of each emotion attribute are first generated.. According to the coordinates of a particular music in this  $N$ -dimensional space, music that are located in the  
35 proximity of the given music are first selected. After the pre-selection stage, instead of computing the similarity of the given music to the rest of the database, only the similarity between the given music and the pre-selected music are computed.

Any music emotion/mood model proposed in the literature can be used to construct the N-dimensional mood space. For example, the two-dimensional model proposed by Thayer [1]. The model adopts the theory that the mood is entrained from two factors : stress (positive/negative) and energy (calm/energetic). According to Thayer's mood model, any music can be described by a stress value and an energy value and such values give the coordinates of a given music and hence determine the location of the emotion in the mood space. In Figure 1a, the stress value and energy value of music x is  $S(x)$  and  $E(x)$  respectively and the mood of x is a function of the emotion attribute, i.e.  $mood(x) = f(E(x), S(x))$ , where f can be any function. As mentioned above, two music that are close to each other in the mood space, such as music x and music y, are considered to be similar as they are both considered as "contentment". On the other hand, an "Anxious" music such as z is far away from x in the mood space and anxious music such as z are generally not perceived as similar to a "contentment" music such as x. The similar concept is not limited to Thayer model, it can be extended to any N-dimensional model. For example, in Figure 1b, a three dimensional mood space is depicted. Its coordinates describes the degree of happiness, passion and excitement respectively.

The coordinates of a music in the mood space is proposed to be generated from any machine learning algorithms such as Neural Network, Decision Tree and Gaussian Mixture Models etc. For example, taking Fig. 1b as an example, Gaussian Mixture Models, i.e., passion model, happiness model and excitement model can be used to model each mood dimension. Such mood models are trained beforehand. For a given music, each model will generate a score and such score can be used as the coordinates value in the mood space.

After the location of the music in the mood space are determined, music pieces that are close to a given music in the mood space are identified by using simple distance measure such as Euclidean distance, Mahalanobis distance or Cosine angles etc.

For example, in Fig. 2, only music pieces that fall within the proximity area, e.g. circle A, are considered as close to music x in the mood space and music z is considered as too far away and hence dissimilar to music x. According to the distance, the system can either select N music pieces that are close to the given music or a distance threshold can be set and only music distance smaller than the threshold will be selected.



To generate a similarity list for music x, a similarity measure is introduced to compute the similarity between music x and the pre-selected music piece. The similarity measure can be any known similarity measure algorithms, e.g., each music is modelled by Gaussian Mixture Model. Any model distance criterion  
5 (see e.g. [3]) can then be used to measure the distance between the two Gaussian Models.

#### Advantages

10 The main advantage is the significant reduction in computation to generate music similarity lists for a large database without affecting the similarity ranking performance from the perceptual point of view.

The invention will now be explained based on preferred embodiments thereof  
15 and by taking reference to the accompanying and schematical figures.

**Fig. 1A** is a schematical diagram of a mood space model which can be involved in an embodiment of the inventive method for classifying audio data.

20

**Fig. 1B** is a schematical diagram of a mood space model which can be involved in another embodiment of the inventive method for classifying audio data.

25 **Fig. 2** elucidates by means of a schematical diagram a proximity concept which can be involved in the embodiment for the inventive method for classifying audio data as illustrated in Fig. 1A.

30 **Fig. 3** is a schematical diagram which elucidates basic aspects of the inventive method for analyzing audio data according to a preferred embodiment by means of a flow chart.

In the following functional and structural similar or equivalent element structures will be denoted with the same reference symbols. Not in each case of  
35 their occurrence a detailed description will be repeated.

Fig. 1A demonstrates by means of a graphical representation in a schematical manner a model for a mood space M which can be involved for carrying out the

method for classifying audio data according to a preferred embodiment of the prevent invention.

The mood space M shown in Fig. 1A is based, defined and constructed according to so-called mood space data MSD. Locations or positions within said mood space M and in order to navigate within said mood space M are the entities stress S and energy E. Therefore, the model shown in Fig. 1A is a two-dimensional mood space model for said mood space M. In the coordinate system defined by the two axes for stress S and energy E, three locations for three different sets of audio data AD, AD' are indicated. The respective sets of audio data AD, AD' are called x, y, and z, respectively. In the embodiment shown in Fig. 1A the first set of audio data AD which is called x serves as given audio data x. Based on the evaluation of the entities stress S and energy E for said first set of audio data x respective parameter values S(x) and E(x) are generated. Therefore, the respective location LADx for said first set or sample of audio data x is a function of said measured values S(x), E(x). In the simplest case of a representation the location LADx for audio data x is simply the pair of values S(x), E(x), i.e.

$$LADx := LAD(S(x), E(x)) = \langle S(x), E(x) \rangle.$$

The same may hold for second and third audio data y and z with measurement values S(y), E(y) and S(z), E(z), respectively. According to the general properties for the locations or positions LADy and LADz in said mood space M the following expressions are given:

$$LADy := LAD(S(y), E(y)) = \langle S(y), E(y) \rangle$$

and

$$LADz := LAD(S(z), E(z)) = \langle S(z), E(z) \rangle.$$

As can be seen from the representation of Fig. 1A, under the assumption that a distance function is valid in the Euclidean manner, audio data x and y are close together with respect to each other, whereas audio data z are at a distal position with respect to said first and second audio data x and y, respectively.

Additionally certain regions of the complete mood space M can be assigned to certain characteristics moods such as contentment, depression, exuberance, and anxiousness.

5 Fig. 1B demonstrates by means of a graphic representation in a schematic way that also more than two dimensions in said mood space M are possible. In the case of Fig. 1B one has three dimensions with the entities happiness, passion and excitement defining the respective three coordinates within said mood space M.

10

Fig. 2 demonstrates in more detail the notion and the concept of neighbourhood and vicinity for the embodiment already demonstrated in Fig. 1A. Here one has the original audio data x with a respective location or position LADx in said mood space M. With respect to a given concept of distance or metric one can generate or receive a threshold value which might be used in order to realize or define neighbourhoods A(x) for said audio data x within said mood space M. The shown neighbourhood A(x) for said audio data x is a circle with the position LADx for said first audio data x in its centre and having a radius with respect to the distance or metric underlying the neighbourhood concept discussed here which is equal to the chosen threshold value. Any additional audio data AD within said neighbourhood circle A(x) are assumed to be comparable and similar enough when compared to said first and given audio data x. In contrast, additional audio data z is too far away with respect to the underlying distance or metric so that z can be classified as being not comparable to said given and first audio data x. Such a concept of vicinity or neighbourhood can be used in order to compare a given sample of audio data x with a data base of audio samples, for instance in order to reduce computational burden when comparing audio data samples with respect to each other. In the case shown in Fig. 2 a pre-selection process is carried out based on the concept of distance and metric in order to select a much more refined subset from the whole data base containing only a very few samples of audio data which have to be compared with respect to each other or with respect to a given piece of audio data x.

35 Fig. 3 is a schematical block diagram containing a flow chart for the most prominent method steps in order to realize an embodiment of the method for classifying audio data AD according to the present invention.

After initialization step START a sample of audio data AD is received as an input I in a first method step S1.

5 Then, in a following step S2 information is provided with respect to a mood space underlying the inventive method. Therefore in step S2 respective mode space data MSD are provided which define and/or which are descriptive or representative for said mood space M according to which audio data AD, AD' can be classified and compared.

10 A step S3 follows wherein a mood space location LAD for said given audio data AD within said mood space M is generated. Contained is a substep S3a for analyzing said audio data AD, e.g. with respect to a given feature set FS which might be obtained from a respective data base. In the following substep S3b the mood space location LAD for said audio data AD is generated as a function  
15 of said audio data AD:

$$LAD := LAD(AD).$$

In the following step S4 a comparison mood space location CL is received, for  
20 instance also from a data base. Said comparison mood space location CL might be dependent on one or a plurality of additional audio data AD' to which the given audio data AD shall be compared to. Additionally in this case the comparison mood space location CL might also be dependent on the feature set FS underlying the present classification scheme.

25

In the following step S5 the locations LAD for the given sample of audio data AD and the comparison location are compared in order to generate respective comparison data CD. Said comparison data CD might also be realized by indicating a distance between said locations LAD and CL.

30

In the following step S6 the comparison data CD are given as an output O.

Finally, the process demonstrated in Fig. 3 is terminated either with a process step END-1 if a quick and sub-optimal classification is sufficient or with –  
35 after a detailed and expensive classification S7 is needed – with an alternative process step END-2.

**Cited Literature**

- 5 [1] Dan Liu, Li Lu & Hong-Jiang Zhang, "Automatic mood detection from acoustic music data", Proceedings of the Fourth International Conference on Music Information Retrieval (ISMIR) 2003.
- 10 [2] Tao Li & Mitsunori Ogihara, "Detecting emotion in music", Proceedings of the Fourth International Conference on Music Information Retrieval (ISMIR) 2003.
- [3] J.J. Aucouturier & F. Pachet, "Finding songs that sound the same", in Proc. Of the IEEE Benelux Workshop on model based processing and coding of audio, Nov 2002.

**Reference Symbols**

|                              |                                                                                                |
|------------------------------|------------------------------------------------------------------------------------------------|
| <b>A, A(x)</b>               | neighbourhood, vicinity, neighbourhood or vicinity w.r.t. mood space location for audio data x |
| <b>AD</b>                    | audio data, audio data sample                                                                  |
| <b>AD'</b>                   | audio data, audio data sample, additional audio data                                           |
| <b>CD</b>                    | comparison data                                                                                |
| <b>CL</b>                    | comparison mood space location                                                                 |
| <b>E, E()</b>                | energy                                                                                         |
| <b>FS</b>                    | feature set                                                                                    |
| <b>I</b>                     | input, input data                                                                              |
| <b>LAD, LADx, LADy, LADz</b> | mood space location for received audio data AD, x, y, z, respectively                          |
| <b>LAD'</b>                  | additional mood space location for received additional audio data AD'                          |
| <b>M</b>                     | mood space                                                                                     |
| <b>MSD</b>                   | mood space data                                                                                |
| <b>O</b>                     | output, output data                                                                            |
| <b>S, S()</b>                | stress                                                                                         |
| <b>x</b>                     | audio data, audio data sample                                                                  |
| <b>y</b>                     | audio data, audio data sample                                                                  |
| <b>z</b>                     | audio data, audio data sample                                                                  |

**Claims**

1. Method for classifying audio data (AD),

comprising:

- a step (S1) of providing audio data (AD) in particular as input data (I),
- a step (S2) of providing mood space data (MSD) which define and/or  
5 which are descriptive or representative for a mood space (M) according  
to which audio data (AD, AD') can be classified,
- a step (S3) of generating a mood space location (LAD) within said mood  
space (M) for said given audio data (AD),
- a step (S4) of providing at least one comparison mood space location  
10 (CL) within said mood space (M),
- a step (S5) of comparing said mood space location (LAD) for said given  
audio data (AD) with said at least one comparison mood space location  
(CL) and thereby generating comparison data (CD), and
- a step (S6) of providing as a classification result said comparison data  
15 (CD) in particular as output data (O).

2. Method according to claim 1,

wherein said mood space (M) is or is modelled by at least one of a Gaussian  
mixture model, a neural network model, and a decision tree model.

20

3. Method according to any one of the preceding claims,

- wherein said mood space (M) is or is modelled by an N-dimensional  
space or manifold and
- wherein N is a given and fixed integer.

25

4. Method according to any one of the preceding claims,

wherein said comparison data (CD) are at least one of being descriptive for,  
being representative for and comprising at least one of a topology, a metric, a  
norm, a distance defined in or on said mood space (M).

30

5. Method according to any one of the preceding claims,

wherein said comparison data (CD) and in particular said topology, metric,  
norm, and said distance are obtained based on at least one of said Euclidean  
space model, said Gaussian mixture model, said neural network model, and  
35 said decision tree model.

6. Method according to any one of the preceding claims,

wherein said comparison data (CD) are derived based on said mood space location (LAD) within said mood space (M) for said given audio data (AD) and on said comparison mood space location (CL) within said mood space (M).

- 5     **7.**     Method according to any one of the preceding claims,  
wherein said mood space (M) and/or the model thereof are defined based on Thayer's mood model.
- 10     **8.**     Method according to any one of the preceding claims,  
wherein said mood space (M) and/or the model thereof are two-dimensional and are defined based on the measured or measurable entities stress (S()) describing happy and anxious moods and energy (E()) describing calm and energetic moods as emotional or mood parameters or attributes.
- 15     **9.**     Method according to any one of the preceding claims,  
wherein said mood space (M) and/or the model thereof are three-dimensional and are defined based on the measured or measurable entities for happiness, passion, and excitement.
- 20     **10.**    Method according to any one of the preceding claims,  
wherein said step (S4) of providing said at least one comparison mood space location (CL) comprises:  
- a step of providing at least one additional audio data (AD, AD') in particular as additional input data (I) and  
25    - a step of generating a respective additional mood space location (LAD') for said additional audio data (AD'), and  
wherein said respective additional mood space location (LAD') for said additional audio data (AD') is used for said at least one comparison mood space location (CL).
- 30     **11.**    Method according to claim 10,  
wherein at least two samples of audio data (AD, AD') are compared with respect to each other – one (AD) of said samples of audio data (AD, AD') being assigned to said derived mood space location (LAD) and the other one (AD') of  
35    said of audio data (AD, AD') being assigned to said additional mood space location (LAD') or said comparison mood space location (CL) – in particular by comparing said derived mood space location (LAD) and said additional mood space location (LAD') or said comparison mood space location (CL).



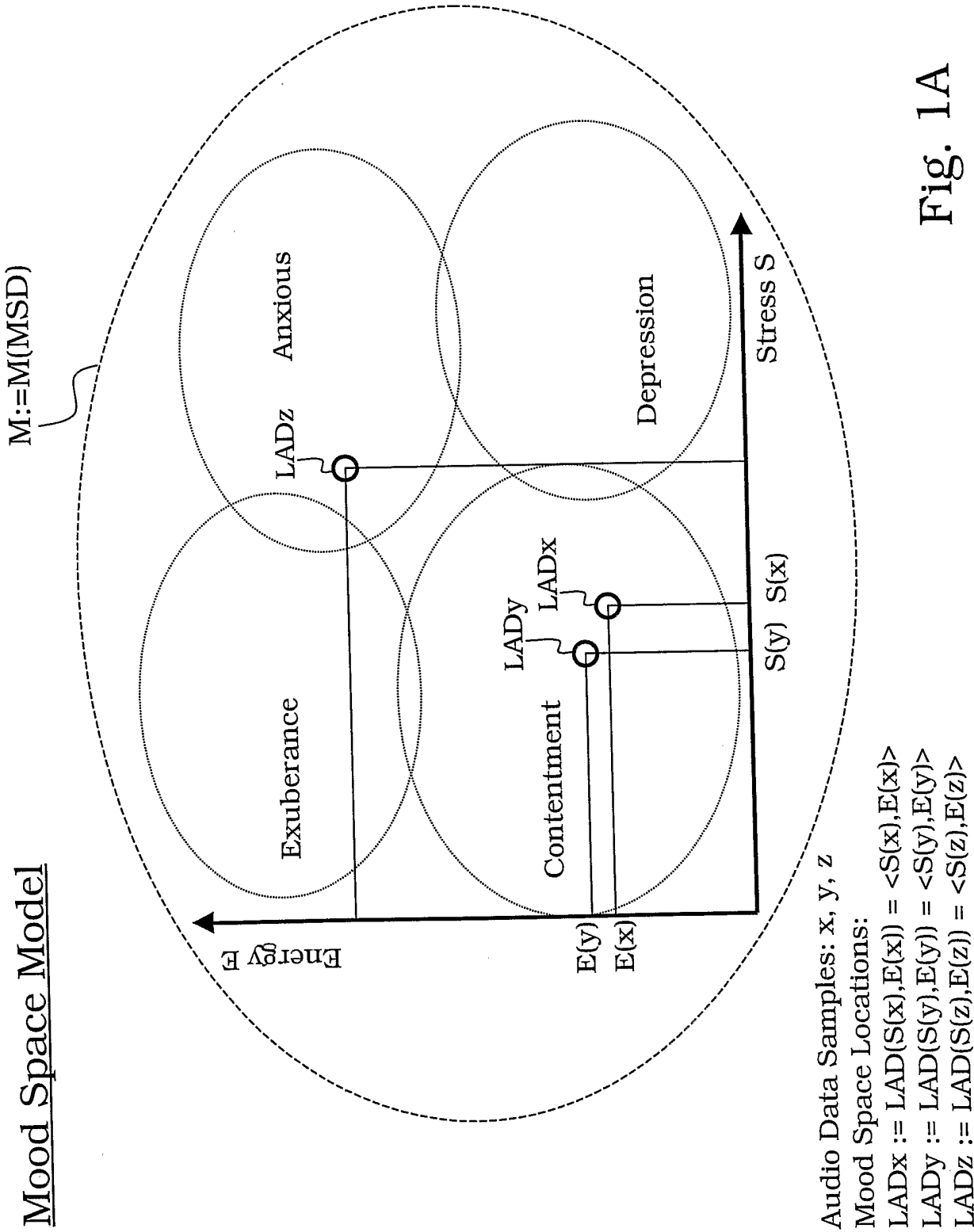
12. Method according to claim 11,  
wherein said at least two samples of audio data (AD, AD') to be compared with  
respect to each other are compared with respect to each other based on said  
comparison data (CD) in a pre-selection process or comparing pre-process and  
5 then based on additional features, e. g. based on features more complicated to  
calculate and/or based on frequency domain related features, in a more  
detailed comparing process.
13. Method according to claim 12,  
10 wherein said at least two samples of audio data (AD, AD') to be compared with  
respect to each other are compared with respect to each other in said more  
detailed comparing process based on said additional features, if said compari-  
son data (CD) obtained from said pre-selection process or comparing pre-  
process are indicative for a sufficient neighbourhood of said at least two  
15 samples of audio data (AD, AD').
14. Method according to any one of the preceding claims,  
wherein a plurality of more than two samples of audio data (AD, AD') are  
compared with respect to each other.  
20
15. Method according to any one of the preceding claims,  
wherein said given audio data (AD) are compared to a plurality of additional  
samples of audio data (AD').
- 25 16. Method according to any one of the preceding claims 14 or 15,  
wherein from said comparison a comparison list and in particular a play list is  
generated which is descriptive for additional samples of audio data (AD') of  
said plurality of additional samples of audio data (AD') which are similar to  
said given audio data (AD).
- 30 17. Method according to any one of the preceding claims,  
wherein music pieces are used as samples of audio data (AD, AD').
18. Apparatus for classifying audio data,  
35 which is adapted and which comprises means for carrying out a method for  
classifying audio data according to any one of the claims 1 to 17 and the steps  
thereof.
19. Computer program product,

comprising computer program means which is adapted to realize a method for classifying audio data according to any one of the claims 1 to 17 and the steps thereof, when it is executed on a computer or a digital signal processing means.

5

**20.** Computer readable storage medium,  
comprising a computer program product according to claim 19

10



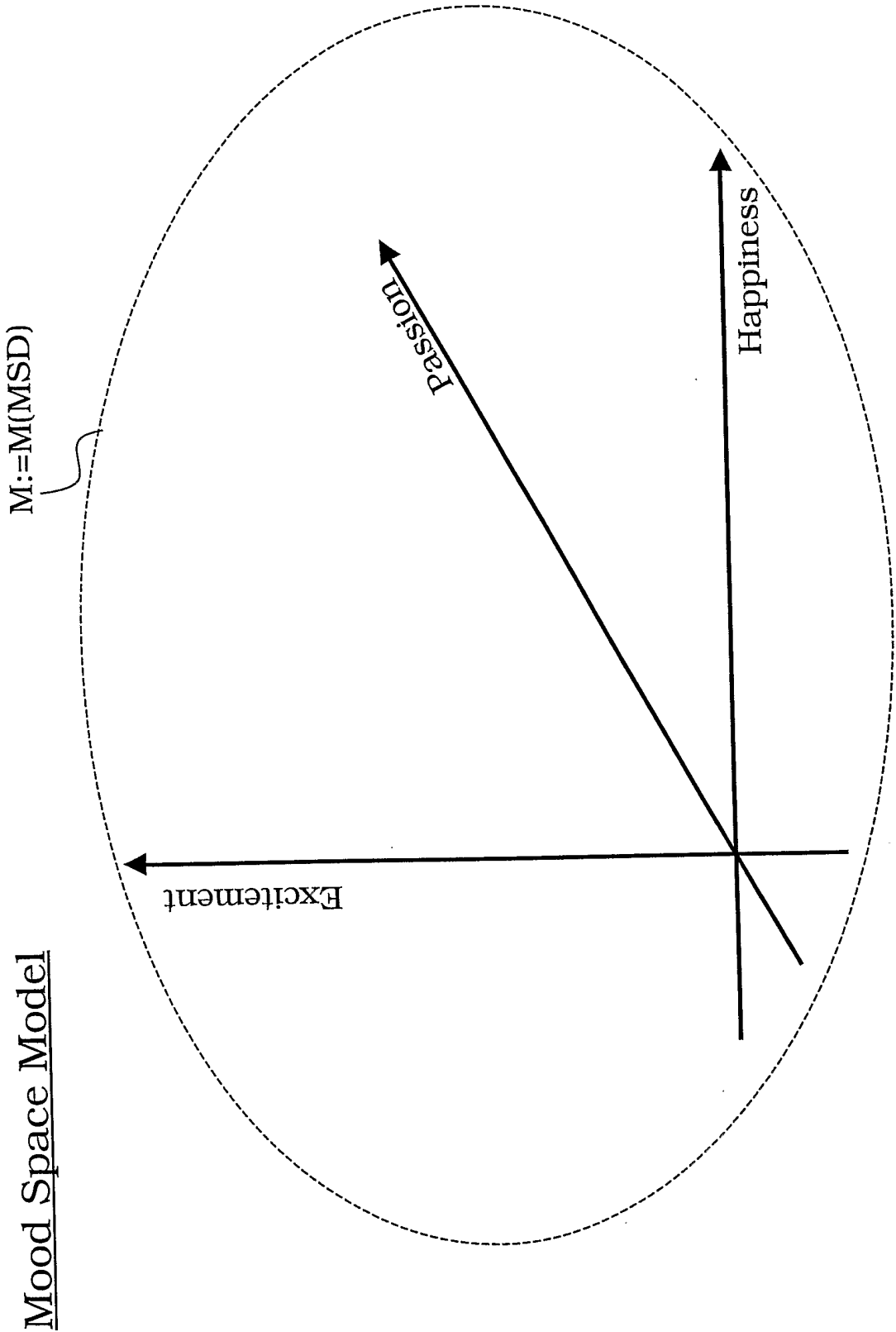


Fig. 1B

Mood Space Model

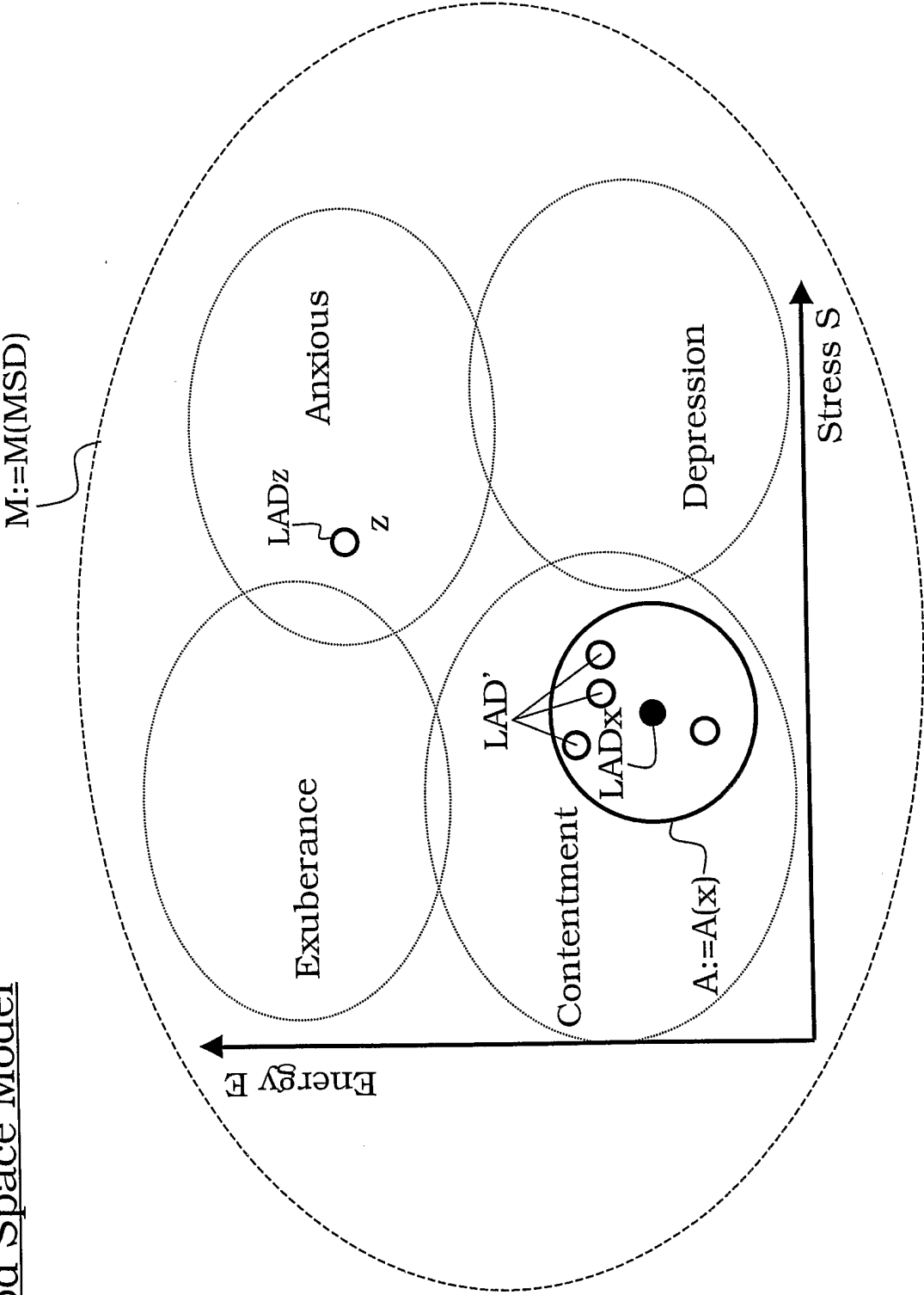


Fig. 2

4/4

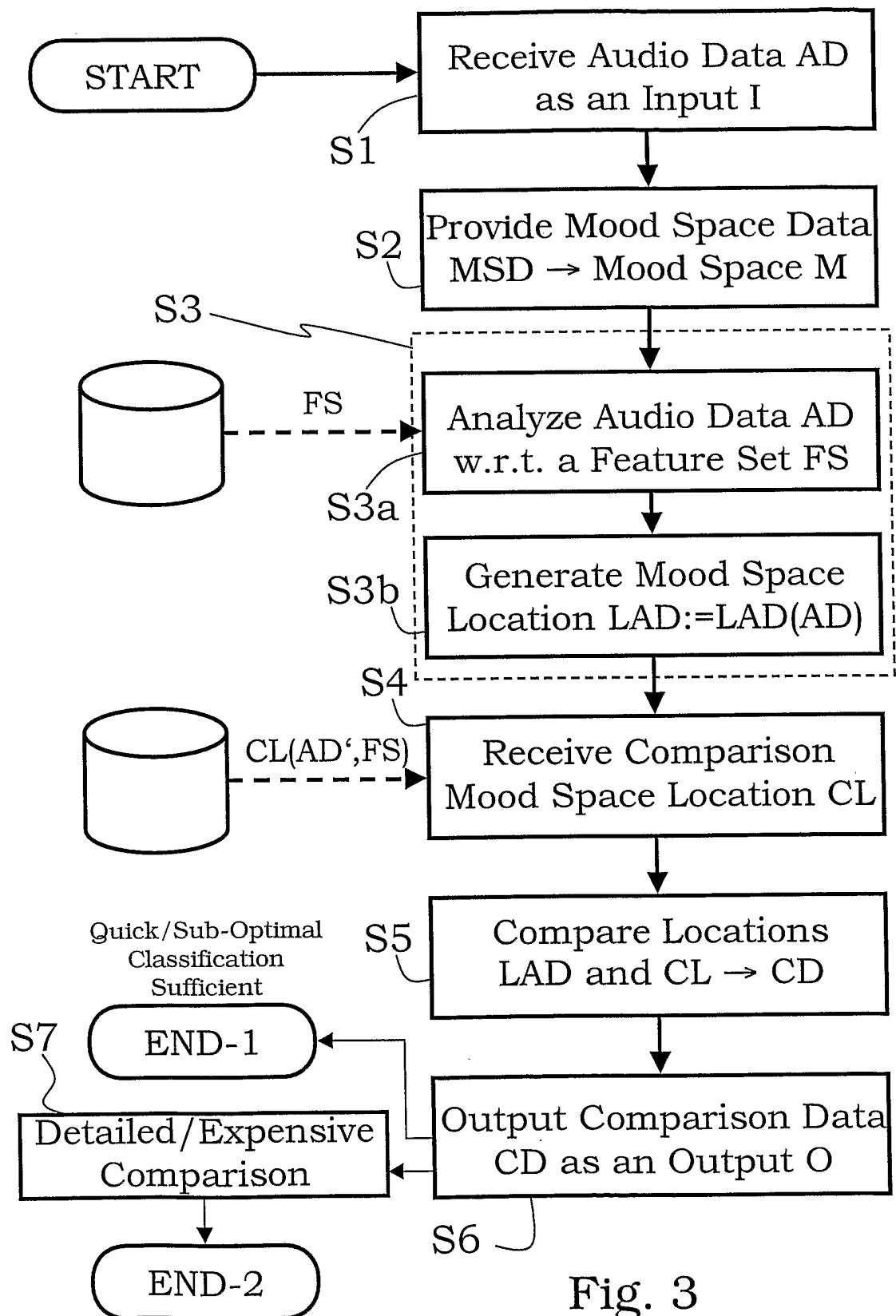


Fig. 3

# INTERNATIONAL SEARCH REPORT

International application No  
PCT/EP2006/002398

**A. CLASSIFICATION OF SUBJECT MATTER**  
INV. G10L11/00

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, INSPEC

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                               | Relevant to claim No. |
|-----------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | TOLOS M ET AL: "Mood-based navigation through large collections of musical data" CONSUMER COMMUNICATIONS AND NETWORKING CONFERENCE, 2005. CCNC. 2005 SECOND IEEE LAS VEGAS, NV, USA 3-6 JAN. 2005, PISCATAWAY, NJ, USA, IEEE, 3 January 2005 (2005-01-03), pages 71-75, XP010787613<br>ISBN: 0-7803-8784-8<br>the whole document | 1-20                  |
| X         | US 2003/045953 A1 (WEARE CHRISTOPHER B)<br>6 March 2003 (2003-03-06)<br>the whole document                                                                                                                                                                                                                                       | 1-6, 8-20             |
|           | -----<br>-/--<br>-----                                                                                                                                                                                                                                                                                                           |                       |

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

\* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

23 May 2006

Date of mailing of the international search report

01/06/2006

Name and mailing address of the ISA/  
European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Quélavoine, R

## INTERNATIONAL SEARCH REPORT

International application No

PCT/EP2006/002398

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

| Category* | Citation of document, with indication, where appropriate, of the relevant passages                                                                                                                                                                                                                                                                                           | Relevant to claim No. |
|-----------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------|
| X         | KODAMA Y ET AL: "A music recommendation system"<br>CONSUMER ELECTRONICS, 2005. ICCE. 2005<br>DIGEST OF TECHNICAL PAPERS. INTERNATIONAL<br>CONFERENCE ON LAS VEGAS, NV, USA JAN.<br>8-12, 2005, PISCATAWAY, NJ, USA,IEEE,<br>8 January 2005 (2005-01-08), pages<br>219-220, XP010796610<br>ISBN: 0-7803-8838-0<br>the whole document<br>-----                                 | 1-6,8-20              |
| X         | YAZHONG FENG ET AL: "Music information<br>retrieval by detecting mood via<br>computational media aesthetics"<br>WEB INTELLIGENCE, 2003. WI 2003.<br>PROCEEDINGS. IEEE/WIC INTERNATIONAL<br>CONFERENCE ON OCT. 13-17, 2003,<br>PISCATAWAY, NJ, USA,IEEE,<br>13 October 2003 (2003-10-13), pages<br>235-241, XP010662937<br>ISBN: 0-7695-1932-6<br>the whole document<br>----- | 1-6,8,9,<br>17-20     |



## INTERNATIONAL SEARCH REPORT

### Information on patent family members

International application No

PCT/EP2006/002398

| Patent document<br>cited in search report | Publication<br>date | Patent family<br>member(s) | Publication<br>date |
|-------------------------------------------|---------------------|----------------------------|---------------------|
| US 2003045953                             | A1                  | 06-03-2003                 | NONE                |