



G00001

A METHOD OF LABELLING A SOUND OR A REPRESENTATION THEREOFBACKGROUND AND FIELD OF THE INVENTION

5 This invention relates to a method of labelling a sound.

Multimedia documents are in many cases a combination of separately produced multimedia events, such as visual, video events and corresponding audio events. It is often desired
10 to locate a particular event in a multimedia document, such as video footage and/or the corresponding sound effect, for example of a man walking on sand.

One method of providing this information is for the
15 multimedia document to be reviewed and labels of event, to be manually entered to provide a database of such events. This, however, is a labour intensive and tedious process. Tools have been proposed which allow the existence of events in a multi-media document to be logged as the events are
20 produced and for some events it is possible for such logging to be produced automatically. For example, modern cameras can "stamp" each picture with the date the picture was taken. Many multi-media file types include a header with such information automatically provided. However, present
25 automatic techniques cannot provide content-related text descriptions and thus cannot replicate the manual process described above.



159159

It is an object of the invention to provide a method of labelling a sound event in a content-related manner which provides a measure of automation.

5 SUMMARY OF THE INVENTION

According to the invention in a first aspect, there is provided an apparatus for labelling a sound or a representation thereof, the apparatus comprising a sound
10 generator capable of generating a family of sounds or their representations by selection of values of parameters of a sound model, at least some parameter values being associated with descriptive labels whereby selection of the value automatically selects the corresponding label.

15

According to the invention, in a second aspect, there is provided a method of labelling a sound or a representation thereof comprising the steps of: selecting a sound or representation by selection of values of parameters of a
20 sound model, at least some parameter values being associated with descriptive labels whereby selection of a value automatically selects a corresponding label, generating the sound or representation as a file and associating the file with the label.

25

Preferably, the values of each parameter are divided into a plurality of ranges, a label being associated with each range and the value labels are preferably combined with a model

label in a grammatical structure whereby the value label(s) qualify the model label description, as adjectives or adverbs, for example.

- 5 The sound or representation thereof may be in the form of a digital audio file, analog audio file, control codes for a synthesizer or in the form of the selected parameters values for the model, for example.
- 10 Further features of the invention may be found in the appendant claims.

BRIEF DESCRIPTION OF THE DRAWINGS

- 15 An embodiment of the invention will now be described, by way of example with reference to the accompanying drawings, in which:

FIG. 1 is block diagram illustrating an interactive sound
20 effect system with which the present invention may be used.

FIG. 2 is a functional block diagram illustrating the logical structure of the system Fig. 1.

- 25 FIG. 3 illustrates a graphic user interface showing user parameters for the sound effect "footsteps" where the user parameters are represented in the form of graphical sliders.

FIG. 4 illustrates the parameter structure employed by the system of Fig. 1.

FIGS. 5-13 are tables charting the parameters used for actual
5 sound models.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT OF THE INVENTION

10 The present invention to be described is concerned with the labelling of sounds produced by a sound model. By way of background explanation, a sound model-based system for generating sounds from sound models will first be described, although it will be appreciated by those skilled in the art
15 that the present invention is applicable to any parameter-adjustable model system.

SYSTEM CONFIGURATION

20 The overall system can be conceptualized as several layers, the top layer being applications that will use the sounds. A graphical user interface is just one such application (though a rather special one). The next layer is the collection of algorithmic sound models. These are objects
25 that encapsulate data and describe the sound behaviour. The models provide the interface which applications use to control the sounds by playing, stopping, and sending messages or by updating parameters.

Sound models generate commands and pass them at the proper time to the synthesizer. The synthesizer is the back-end where the sample by sample audio waveforms are produced, mixed, and sent to the audio output device.

5

FIG. 1, is a block diagram illustrating the main elements of the system. A central processing unit 1 is connected to random access memory (RAM) 2, one or more input devices such as keyboard 3, mouse 4, joystick 5, MIDI controller 6; a
10 visual display device 7; a sound synthesizer 8; and audio output system including digital-to-analog converter (DAC) 9, amplifier 10, loudspeaker or headphone 11 (or alternatively, a soundcard integrating some of these individual devices can be used); and a nonvolatile storage device such a hard disk
15 12. These components play a supporting role to the central element of the system, the interactive sound effects computer program (SFX program) which consists of a sound effects software engine (SFX engine) 13 and optionally a graphical user interface program (GUI) 14. In one mode of operation
20 the GUI can be replaced by a controlling program 15. Also present is an operating system program 16, such as the standard operating system of any personal computer system.

The CPU 1 executes the stored instructions of the programs
25 in memory, sharing its processing power between the SFX engine 13, the GUI 14 or controlling program 15, the operating system 16 and possibly other programs according to a multi-tasking scheme such as the well known timeslicing

technique. Under command of the SFX engine, the CPU delivers a stream of commands to the sound synthesizer 8, which produces digital audio in response to these commands. The output of the synthesizer is a digital audio signal which is converted to an analogue form by the digital to analogue converter (DAC) 9, then amplified and delivered to the user by means of the amplifier 10 and loudspeaker or headphones 11. Optionally the digital audio signal may also be delivered back to the CPU allowing it to be further processed or stored as a sound file for later retrieval. The hard disk or other nonvolatile storage 12 provides means to store indefinitely the following items:

1. The SFX program itself including the data and instructions representing multiple sound effects models.
2. Settings of parameters and other variable elements of the SFX program.
3. Optionally, sound files comprising digital audio output from the synthesizer 8 under control of the SFX program.

The SFX engine 13 is controlled directly by means of the GUI 14, or from an external controlling program such as a computer game 15 or, rarely, by both at the same time. When under control of the GUI, the user effectively interacts directly with the SFX program, controlling the program by means of one or more input devices such as an alphanumeric

computer keyboard 3, a pointing device 4, a joystick 5, a specialized controller such as a slider bank or music keyboard connected by means such as the Musical Instrument Digital Interface (MIDI) standard 6, or other physical
5 controlling means. In this mode of use, the GUI program 14 uses the display device 7 to provide to the user with visual information on the status of the SFX engine 13 including which sound effects models are currently invoked, the structure of these sound models, the settings of their
10 parameters, and other information.

When the control is by means of a pointing device, the display device 7 also provides feedback to the user on the logical position of the pointing device in the usual manner.
15 By observing the display 7 and/or listening to the audio output while manipulating the input devices 3 through 6, the user is able to alter sound effects until satisfied with the results. This mode of operation is designed to allow the user to create specific sound effects according to his/her
20 needs from the generic sound effects models of the SFX system, by selecting sound effects models, initiating or triggering them to produce audio output from the system, adjusting the parameters of the models, selecting elements of models, and other actions.

25

In the alternative mode of operation, the SFX engine 13 is under the control of an external controlling program 15, such as a computer game, the program of a network resident

information site (website), a virtual reality program, a video editing program, a multimedia authoring tool, or any other program which requires sound effects. In this mode the user interacts with the controlling program 15 by means of the input devices 3 through 6 and the display device 7. The SFX engine 13 acts as a slave to the controlling program 15, producing sound effects under its control. This is achieved by allowing the controlling program to send data to the SFX engine 13, this data being interpreted by the SFX engine as controlling messages. In this mode of operation, the SFX engine will typically not be visible to the user on the display 7, and will be controllable by the user only indirectly via aspects of the controlling program which influence the SFX engine. The manner and degree of control which the user has over the SFX engine is entirely a function of the controlling program and is decided by the designer of the controlling program.

LOGICAL STRUCTURE

The logical structure of the present system is shown in Fig 2. The main elements are the SFX engine 1 which, as described above, may be under control of the GUI 14 or, in the alternative mode of operation, under control of an external controlling program 15. Also shown is the synthesizer 8 which leads to the audio output system. These elements are the same as the corresponding elements of Fig 1, but are here shown in a way which highlights their logical

interrelationships. In the mode of operation where the invention is being used directly by a user, the user controls the system by means of the GUI 14, which acts to accept user input (such as keystrokes of the computer keyboard or movements of a pointing device) and to inform the user both of the status of the system and of the effect of his/her actions. User actions which affect the production of sound effects generate control messages which are sent from the GUI to the SFX Engine 13 in order to initiate, terminate, and control sound effects. These messages are in a format determined by the SFX Engine and known to the GUI. In response to these messages, the SFX engine 13 models the behaviour of the currently active sound effects and generates a stream of events or commands which are sent to the synthesizer 4, which in turn generates the audio output. Certain information affecting the manner of display to be used by the GUI 14 is contained within the SFX engine 13 for example the manner in which the control parameters of a sound effects model should be displayed varies from one model to another, and the information about the currently active sound effects models is held by the SFX engine. Thus there is a need for information to be returned from the SFX engine to the GUI, and this is achieved by allowing the SFX engine to send display information to the GUI or allowing the GUI to elicit display information from the SFX engine.

In the alternative mode of operation where the SFX engine is being controlled from a program external to the invention,

the user interacts with the external controlling program 15 in a manner which is completely independent of the invention. The controlling program 15 sends control messages to the SFX engine 13 in order to initiate, terminate, and control sound effects. These messages are in a format determined by the SFX Engine and known to the controlling program, and typically are similar to, or a subset of those used by the GUI in the first mode of operation described above. In response to these messages, the main purpose of the SFX engine 13 is to model the behaviour of the currently active sound effects and generate a stream of events or commands which are sent to the synthesizer 8, which in turn generates the audio output.

15 The main internal elements of the SFX engine 13 are a set of interactive sound effects models (SFX models) 20; an Application Programmer Interface (API) 17; a Message Processor 18; a Parameter Linker/Mapper 19 and a Timing and Synthesizer Command Processor (TSCP)21.

20

In the library or set of interactive sound effects models (SFX models) 20, each model consists of data and programmed instructions representing the sound characteristics and behaviour of a sound effect, or a class of sound effects. These models may be invoked sequentially or simultaneously, so that the system is capable of producing sound effects in isolation or in combination, typically after an imperceptible or near imperceptible delay (in so called "real time"). Each

SFX model is provided with one or more control parameters which may be used to alter the sound produced by the SFX model, and these control parameters may also be modified in real time to produce audible changes in the output while the system is producing sound effects. In certain cases compound sound effects models may be made up of other sound effects models arranged in a hierarchy consisting of any number of levels, thus enabling arbitrarily complex models to be built from a number of simpler models.

10

The Application Programmer Interface (API) 17 receives data which is interpreted by the SFX engine as controlling messages, these messages arriving from either the GUI 14 or the external controlling program 15. The API decodes the messages in order to establish which type of message has been sent, and forwards the messages to the Message Processor 8.

15

The Message Processor 8 performs actions as directed by the controlling messages, including starting and stopping particular sound effects, loading and unloading sound effects models from RAM, applying the effect of modifications of control parameters to the SFX models, modifying settings of the SFX engine which influence its overall behaviour, and otherwise controlling the SFX Engine.

20

25

A Parameter Linker/Mapper 19 provides a means of endowing SFX models with one or more alternative sets of control parameters or metaparameters, where these metaparameters are

linked to the original control parameter set of the SFX model or to other metaparameters in a hierarchy of parameters. The Linker/Mapper 19 also provides means of applying mathematical transformations to the values of control parameters and metaparameters. The Parameter Linker/Mapper 19 is useful because the original control parameters of a particular SFX model are not necessarily the most appropriate or useful in every case, for example when the SFX engine is being controlled by an external controlling program 15 which has its own design constraints, or when the SFX model forms part of a compound SFX model as described above.

The Timing and Synthesizer Command Processor (TSCP) 21 provides a number of functions related to timing and to the processing of events and other commands to be sent to the synthesizer 4. The invention is not restricted to any particular method of synthesis, and details of this element depend significantly on the type and design of the synthesizer. However two general functions may be identified:

The SFX engine operates by producing a stream of commands such as MIDI commands which are delivered to the synthesizer in order to produce sounds, and typically this process occurs in real time. Most synthesizers operate by producing or modifying the output sound at the moment an event or command is received. A simple implementation of the SFX engine might therefore produce synthesizer commands only at the moment

they are required by the synthesizer, but this is liable to timing disruption because the CPU may be unable to process the complex command stream of multiple SFX models quickly enough to avoid audible disruption of the output sound.

5 Hence a more sophisticated implementation can achieve greater consistency of timing by generating the commands a short interval ahead of the current time, queuing them in a mechanism such as a data buffer, and delivering them to the synthesizer at the appropriate time. The TSCP provides this

10 function in such a way that the interval by which commands are generated ahead of the current time may be adjusted to an optimum value which may also be set differently for different SFX models. The optimum is a compromise between the need to avoid timing disruption and the need to make the

15 system responsive to changes in its control parameters.

If more than one SFX model is active, or if a single SFX model is complex, there is a need to produce multiple command streams which must be delivered to different channels of the

20 synthesizer, where each synthesizer channel is set up to create different sound elements of the sound effects. In typical implementations these channels are a limited resource and must be managed carefully, for example allocated dynamically upon demand. The TSCP acts as a synthesis

25 channel manager.

As stated above, one purpose of the hard disk or other non-volatile storage (12 in Fig 1) is to provide a means to store

indefinitely the settings of parameters and other variable elements of the SFX program. Such parameters and other elements may be saved while the system is in the mode where it is being controlled directly by a user using the GUI, then
5 recalled when the system is in the alternative mode under control of an external controlling program. This allows a user to experiment directly with the parameters of the sound effects using the GUI, save the set of values of the parameters found to be most appropriate to the application,
10 then recall this same set of values while the SFX engine is under control of the external controlling program in order to have the system produce an identical or near identical sound effect. Saving and recalling a sound effect in this way differs from saving and recalling a digital audio signal
15 of the sound effect in that it is entirely based on a model of the sound effect and may therefore be altered after it has been recalled by means of changing its parameters.

The sound models may be modelled closely on the physics or
20 produces realistic sound, and responds in realistic and/or predictable ways to parameter changes. The sound effects models may be assigned a set of control parameters deemed most important or appropriate to the particular sound effect in question, these being closely related to the behaviour
25 characteristics of the sound generating phenomenon being modelled. This set of parameters may include parameters unique to the particular model, parameters that are generic to sets of similar models, and parameters that are generic

to all models of the system. For example a model of human footsteps might have a parameter for walking style which would be unique to this model, another parameter for walking speed which would be common to all human and animal footstep
5 models, and other parameters such as volume or reverberation depth common to all models.

The system can include models which are programmed with realistic simulations of naturally occurring sound producing
10 entities, other sound effects which are exaggerated in character for dramatic effect, and other sound effects of a purely imaginative nature which have no counterpart in the real world. In the case of realistic simulations and exaggerations of real sounds, the sound effects models may
15 be modelled to any chosen degree of precision on the behaviour of their naturally occurring counterparts, so that the sound effects models will automatically provide accurate reproductions of the sounds, sound sequences or other audible characteristics of their naturally occurring counterparts.

20

The system can also support "Compound Sounds": these are sound models consisting of a hierarchy of other sound models with any number of levels in the hierarchy. Typically they may represent an entire scene consisting of many sonic
25 elements. At the top level the user can make changes to the whole scene (e.g., changing the overall volume), but control over individual elements is also possible, and these lower level elements can optionally be isolated (listened to

"solo") when making adjustments to them.

The generator includes generic support for "parameter linking" in which parameters may be linked to combinations
5 of other parameters according to mathematical relationships; this allows, for example, high level parameters to be used to make broad sweeping changes in multiple lower level parameters, or to apply scaling to other parameters, or to make complex sets of changes in several other parameters.

10

To make the sound as realistic as possible, the system can introduce fluctuations (typically of a random or semi-random nature) into the sounds produced in order to avoid exact repetition and achieve a natural effect. Techniques for
15 introducing fluctuations include:

1. Altering the timing of commands or events sent to the synthesizer.
2. Altering the values of parameters in the commands sent
20 to the synthesizer if the synthesizer is one based on replaying samples, randomly selecting samples from a collection of similar but non-identical samples.

The system generates the stream of commands to the
25 synthesizer a short interval ahead of the current time, this interval being set such that it is long enough to overcome potentially audible disruption of the sound output which would occur if time critical commands were generated at the

moment they are required, but short enough that the system responds to changes in its control parameters after an imperceptible or near imperceptible delay.

- 5 The system provides two modes of triggering. In one mode, the sound effects, typically of a continuous, evolving, or repetitive nature, will once started run continuously until explicitly stopped. In the other mode, the sound effects, typically of a short, non-continuous nature, are triggered
10 each time they are required, thus allowing precise synchronization with visual events in a computer game, film, video production, or animation.

The system includes generic sound effects models in which the
15 behaviour of a class of sound effects is encoded, and which provides a method by which a user of the system can create specific sound models by selecting options of the generic models, setting the values of variables of the generic models to specific values, and providing the synthesizer with its
20 own samples.

SOUND EFFECTS SYNTHESIS TECHNIQUE

Sound Representation

25

The sound models consist of the interface functions, the parameters for external control, private data for maintaining state while the process is suspended to share CPU time,

indexes into the bank wave tables or synthesis data the model uses and the event generating code.

The sound models are arranged as an object oriented class hierarchy, with many sound classes being derived directly from the base class. This structure is due to the fact that there are many attributes and methods common to all sounds (e.g. location, volume), while most other attributes are common to one model, or shared with other models that otherwise have little in common (e.g. surface characteristics of footsteps).

The sound models have a compute ahead window of time which is the mechanism by which the model share the CPU. This window can be different for different sound models, and is usually in the range of 100-300 milliseconds. The sound model process is called back at this rate, and computes all the events up to and slightly beyond the next expected callback time. The events are time-stamped with their desired output times, and sent to the output manager. Several aspects of sound model representation have been developed as a direct result of how multimedia applications developers need to control them. The first is parameter presets. These come from the frequent need to use a model with several different distinct parameterizations. Rather than burden the developer with having to write code explicitly to change each of perhaps many parameters, parameters may be adjusted using the graphical user interface

("GUI") while monitoring their effect on the sound, then stored as presets which can be recalled by the application. Besides developer convenience, the advantage of switching parameter settings in the application this way within a single instance of a model rather than using two different instances of the model, is that the event generating algorithm (consider footsteps) can continue without breaking stride.

Another representation issue is the need for two different and mutually exclusive methods of control over many sounds. Conceptually, there are two different kinds of parameters; those which the application will use to interact with the sound in real time, and those which select a particular parameterization of the sound from the database. These two groups may be different in different contexts for the same sound.

Consider a footsteps model. If a virtual environment application is designed so that the view of the world is through the user's eyes, then one of the natural controls of the footsteps sound would be a rate parameter. The faster the user moves through a space, the faster the rate of the footsteps. If, however, the footstep sound is attached to visible feet, then the individual sounds obviously need to be synchronized to a graphical event.

In the first case, there is no graphical event and it would

be posing a significant burden on the application to have to time and send a message to the sound model for every step. In the second case, a rate parameter is meaningless.

- 5 The present system provides for either or both methods of control. If event by event control is needed, the model's standard play function is not invoked, but the object provides a massaging interface which is used instead. All the other support (e.g. statistical variability of successive
10 sounds, parameter control) is still available. For sound models that use rate to control other attributes, a meaningful rate must be measured from the event triggers.

A more complex issue of control can be illustrated with an
15 applause model which, for example, is to be controlled in realtime using a parameter for the number of clappers. The parameter would typically start at 0, be driven up to a level corresponding to how many people are in the virtual audience, remain at that level for some time, then gradually decay back
20 to zero. However, for certain purposes, an application may not need such intimate control. It may be preferable to simply specify the number of people and an "enthusiasm" level (a "metatime" parameter) that could in turn affect the temporal envelope of the "number of people" parameter. The
25 application would only have to concern itself with the "enthusiasm" parameter when (or before) the applause sound is initiated. The two methods of control are mutually exclusive.

The applause example is different from the footsteps example because with footsteps, both types of control discussed (individual footsteps vs. rate) are realtime. The contrasting methods of control in the applause example are
5 between a metatime specification of a temporal trajectory, and real time control of the trajectory. It is believed that the most useful way to support these control choices is to record parameter trajectories created by the developer using the GUI, and then use the trajectories during playback after
10 a trigger event from the application.

These control issues arise not from the process of modelling the sound itself, but rather from the contexts in which the models are embedded. The same issues must be considered for
15 every sound model, but the implementation of the different control methods depends heavily on the sound representation.

Sound Modelling Process

20 Below is a description which provides the steps for the sound modelling process. These steps are generic in the sense that they can be used to produce a wide range of sound effects, and are not limited to producing only a particular set. However, so that these principles are more easily understood,
25 various examples will be provided and, at times, discussed for illustration purposes.

The present method and system produce sound effects which

simulate sounds associated with a certain phenomenon. Some examples of a phenomenon are footsteps, earthquake, running airconditioner, bouncing ball, moving car, etc. The phenomenon can be virtually anything so long as there are
5 some sounds with which it is associated. Indeed, the phenomenon need not even necessarily be a real life phenomenon in the sense that it does not have to actually exist in the real world. For instance, the phenomenon could be a firing of a futuristic phaser gun. Although such a gun
10 may not currently exist (hence the phenomenon cannot exist), this fact is irrelevant so long as there is some perception about what the sounds associated with the phenomenon might be like or what might be acceptable to the listeners. It is also useful to have some perception about how the sounds
15 would vary depending on various hypothetical factors. For example, one may perceive the sound associated with the firing of the phaser gun to become louder and sharper as the phaser gun becomes more powerful, and for it to be followed by various kinds of "ricochet" depending on what type of object is struck by the
5 gun's beam.

The sound modeling process begins by identifying the behavioural characteristics associated with the particular sound phenomenon which are relevant to the generation of
10 sound. Behavioural characteristics can be defined as the set of properties which a naive listener would perceive as distinguishing the sound effect from other sound effects,

including those which define how it changes or evolves in response to different conditions impinging upon it. In many cases, the characteristics bear a one to one correspondence to the terms a layman would use to describe the sound effect.

5 In the case of sound effects which do not correspond to an object or phenomenon which actually exists in the real world, e.g., phaser gun as mentioned above, the behavioural characteristics are properties which a naive listener might expect such an object or phenomenon to possess if it did

10 exist.

For instance, in the case of footsteps, the behavioural characteristics would include things such as speed, degree of limp and stagger, weight (of the person producing the

15 footsteps), surface type (e.g., cement, grass, mud), location (i.e., position relative to the listener), surrounding acoustic, etc. It can be easily appreciated that these characteristics define the sound for a particular set of conditions. For instance, the sounds produced from footsteps

20 from a mad dash would be different from those produced in a casual stroll; footsteps on hard marble would sound differently than footsteps on wet mud.

For some of these conditions, it is useful to analyze the

25 mechanics of how the sound is generated from the phenomenon. Once again using footsteps as an example, the sound being generated from footsteps results mainly from two separate events, the impact of the heel hitting a surface, then

shortly after, the impact of the toe hitting a surface. For footsteps of a normal person, the totality of the sound generated results from the heel-toe action of one foot followed by the heel-toe action of the other foot, and so on.

5 As one walks faster, the time interval between the sound produced from the heel-toe action of one foot and heel-toe action of the other foot decreases. However, it is important to also realize that the time interval between the sound produced from a heel and then a toe also decreases in some
10 relationship to the heel-heel time interval. At some point, the sound produced from the heel and the sound produced from the toe actually overlap and it becomes difficult to distinguish the sounds as being separate and distinct.

15 In addition, the heel-to-toe time is affected by another parameter. When marching, the leg falls rapidly and perpendicular to the ground, and thus, the heel-to-toe time is very short. In contrast, a long stride produces a long heel-to-toe time because the heel touches the ground while
20 the leg is far from perpendicular and the toe has a relatively long distance to travel before it touches the ground. Thus, in the present method of modelling, the heel-to-toe time is the net result of both walking speed and "walk style" (march versus long stride). The general principle is
25 that the internal parameters of the model may be influenced by many of the external or "user" parameters in mathematical relationships of arbitrary complexity.

In certain cases, this knowledge about the mechanics of sound generation is important for two main reasons when attempting sound modelling. First, it allows one to vary the correct set of parameters so as to produce the most realistic sound effect. For instance, in the footstep example given above, the resulting sound effect would not sound very realistic had someone varied the time interval between the sound produced from one heel-toe action to another without also proportionately varying the time interval between the heel and the toe. The second main reason for analysing the mechanics is that it allows one some notion of the size and type of sound sample that is needed. For instance, again using the footstep example above, it is important to have independent control of the heel sound and the toe sound individually, and therefore, a separate sample for each of the sounds is needed; it is not enough to have a sample of the heel toe sound as a grouped pair.

Of course, there is a rather large range of behavioural characteristics of any particular phenomenon, and the choice of selection and the extent of the analysis of these behavioural characteristics depend largely upon the potential uses of the sound effects, the nature of the sound producing phenomenon, and degree of realism desired. However, it is generally true that some identification and understanding of the behavioural characteristics of any phenomenon is required to model a sound effect properly.

Once the behavioural characteristics have been identified, some sound samples or other procedurally generated sound is needed which will become the foundation for the many variations. The sample may be obtained either by recording
5 a sample segment of actual sound found in a real world phenomenon (or simply taking a segment from some existing prerecording) or by producing a segment through well known synthesis techniques, whichever is more convenient or desirable given the particular sound effect being modelled.

10 For instance, in a case where a sound of a phaser gun is being modelled, it may be more convenient simply to synthesize a sample, given that no such phaser gun actually exists in the real world. In the case of footsteps, however, it would in most cases be easier simply to record the sound
15 produced from actual footsteps, or to record the individual elements of a footstep (i.e., heel-tap and toe-tap recorded separately), or to record a simulation of these elements (e.g., by tapping together different types of hard objects until the desired sound is achieved). When recording a
20 sample for use in this type of model, it is generally better to isolate the sound, that is, to prevent the inclusion of sounds which are not related to the particular phenomenon at hand.

25 The choice of the length of the sound samples depends on a number of factors. As a general rule, the smaller the sample, the greater the flexibility. On the flip side, the smaller the sample, the greater the labour and the harder it

is to achieve realism. A good rule of thumb is to have a sample which is as long as possible without loss of useful flexibility, that is, where most of the perceptual range of sonic possibilities of the equivalent sound in real life can be achieved by varying the user parameters of the model. For instance, in the case of the footsteps, if one were to want to produce footsteps of different speeds, it would be necessary to obtain a set of samples including heel sounds and toe sounds, for the reasons provided above. However, this does not always mean that one needs to record the two sounds separately since the current editing techniques allow for splicing and other forms of editing to separate a single recording into multiple samples. But the splicing technique may be difficult or impossible for cases where the sounds overlap.

The choice of the sound samples also depends on the behavioural characteristics of the phenomenon to some extent, and also on the limitation of the parameters (parameters are discussed in detail below). Using the footsteps example once again, it should be noted that some sound effects do not require additional samples while some do. For instance, to vary the style of a walk, only the timing needs to be varied, and hence, this can be done with any existing sample. However, to vary the surface on which the footsteps are made, it is easier to simply obtain a sample of footsteps on each of the surfaces rather than attempting to manipulate an existing sample to simulate the effect. For example, it

would not be easy to produce a sound of a footstep on a soft muddy surface using only a sample of footsteps on a concrete surface. How many samples are needed for a given phenomenon, of course, depends on the scope and the range of sound effects one desires, and varies greatly from one sound effect to another.

In many cases, multiple, similar, but non-identical samples are collected. This has two purposes. First, it provides a means of simulating the subtle, everchanging nature of real world sounds. For example, no two footsteps are identical in real life, and a model which produces two identical footsteps in succession is immediately perceived as artificial. With several samples to choose from and a rule which selects randomly from a set of similar samples (typically also excluding the same one being triggered twice in immediate succession), much of the naturalness may be simulated.

The second reason for collecting multiple samples is that a continuous spectrum can often be simulated by collecting points along the spectrum. For example, although there is no known synthesis or sound processing technique for transforming a quiet chuckle into a hearty laugh (or vice versa), a "Strength of Laughter" parameter may be constructed by collecting a set of laugh samples at different "degrees of hilarity", then selecting individual samples according to the setting of the "Strength of Laughter" parameter.

Typically, this technique is combined with the random selection described above.

Sound Modelling Parameter Structure

5

Once the behavioural characteristics have been analyzed and the samples obtained, it is necessary to select the user parameters, i.e., the parameters which are to be made available to a user of the model in order to control the sound effects. The parameters represent the various factors which need to be controlled in order to produce the modelled sounds. Although the parameters can be structured in a number of ways to effect a sound effect, in this system, it is useful to view the parameter structure as illustrated in FIG. 4.

15

In referring to FIG. 4, the top layer consists of the user parameters which are the interface between the user and the sound effects system. The middle layer consists of the parameters employed by the SFX engine, or simply referred to as "engine parameters." The bottom layer consists of the synthesizer parameters which are well known parameters found in any of the current sound or music synthesizers.

20

As the arrows indicate, in general, each of the user parameters affects a combination of engine parameters and synthesizer parameters, though, in simpler cases, a user parameter may control only synthesizer parameters or engine

25

parameters. Any combination of engine and synthesizer parameters is theoretically possible; however, the way in which they are combined will depend on how the user parameter is defined in light of behavioural characteristics of a particular phenomenon, as shall be explained in detail below.

The user parameters are defined in terms of the desired sound effect. For instance, in the case of the footsteps, the user parameters can be location, walking speed, walking style, limp, weight, hardness, surface type, etc. Although these parameters can be defined in virtually any manner, it is often most useful if they directly reflect the behavioural characteristics of a phenomenon and the purpose for which the sound effect is being produced. In many cases, they are the obvious, easily understood parameters that a layman might use to describe the sound. For example, while a user parameter such as surface type might be a useful parameter for the phenomenon footsteps, it probably would not be useful for a phenomenon such as earthquake, given that surface type probably has no meaning in the context of an earthquake.

The user parameters can be represented in a number of ways so as to give control access to the user. However, in this system, it is represented in the form of "sliders" on a graphic user interface (GUI), FIG. 3, where the user can slide the slider bar to control the magnitude of the effect. For instance, for the speed slider for the phenomenon footsteps, the walking speed is increased as the slider is

moved to the right. For the limp slider, the amount of limp in the walk is increased as the slider is moved. To combine the effects, several user parameters can be invoked at once. For instance, by invoking both the speed slider and the limp
5 slider, one can achieve any combination of limp and speed. Some combinations are obviously not desirable, though may be possible. For instance, one probably would not combine the surface type "metal" with "marble". In contrast, "leaves" might well be combined with "dirt" to achieve an effect of
10 footsteps on leaves over dirt. The middle layer parameters, or engine parameters, and the bottom layer parameters, or synthesizer parameters, work in combination to produce the sound effects as defined by the user parameters. The bottom layer parameters can include sound manipulation techniques
15 such as volume control, pan, pitch, filter cut off, filter Q, amplitude envelope, and many others which are well known to those skilled in the art.

when and how the middle and bottom layer parameters are
20 combined is controlled by the SFX engine which takes into consideration the behavioural characteristics of a particular phenomenon. Essentially, the middle layer can be viewed as the layer which "models" the sound using the basic sound manipulation parameters provided by the bottom layer.

25

Although the role of the middle layer parameters is complex, the parameters is can broadly be classified as timing, selecting, and patterning. Although these parameters are

defined here as being separate and distinct, it should understood by those skilled in the art that these parameter representations are conceptual tools to illustrate the sound modelling process or techniques employed by the SFX engine
5 and need not necessarily exist as separate and distinct components in the present sound effects system.

Now in describing the role of each class of parameter individually, timing parameters basically control the length
10 of the time intervals between triggering and stopping pieces of sound within a particular sound effect, and time intervals between other commands sent to the synthesizer. The selecting parameters control which sound samples are selected at a given moment, including the order in which samples are
15 selected. The patterning parameters control the relationships between these factors. By appropriately adjusting these three classes of engine parameter in combination with the synthesizer parameters, a large set of sound effects can be produced. The role and the effect of each of these
20 parameters will become clearer in the examples as provided below.

Referring to the footsteps example once again, described above were the behavioural characteristics of footsteps in
25 relation to speed. It was explained that as the speed increases, the time interval between one heel-toe action to another decreases, as well as the time interval between the heel and the toe. Here, the user parameter (top layer) is

- speed. As the user adjusts the speed slider to increase speed, the time parameter is made to decrease the time interval between one heel-toe action to another, as well as the time interval between the heel and the toe. This timing
5 is also affected by the "style" parameter as described above. However, the pattern or behaviour of the footsteps does not change as speed and style are altered. A heel sound is always followed by a toe sound, etc.
- 10 If, however, the sound effect were that of a moving horse, then the behavioural characteristics are more complicated, and hence, additional parameters need to be involved. For example, consider the user parameter "speed" for the sound of horse's hooves. As the speed increases, it is clear that
15 the timing parameter needs to be adjusted such that the mean of the time intervals between the events becomes shorter, reflecting the fact that the time intervals between impacting of the hooves to a given surface become shorter on average. But, in addition, the patterning and ordering of the events
20 change as the horse switches between walking, trotting, cantering and galloping. The exact pattern, of course, needs to be determined empirically using the behavioural characteristics of an actual horse.
- 25 As the last example, if for the footsteps phenomenon, the user parameter were surface type, then the only class of engine parameters affected are those concerned with selection, since the timing and patterning aspects do not

change. Here, depending on what surface or combination of surfaces is chosen different sets of samples will be selected, but the timing and patterning do not change.

- 5 For some of these examples, there may be instances where synthesizer parameters will have to be invoked either in isolation or in combination with engine parameters. For instance, still using the footsteps example, the synthesizer parameters, pitch, volume, etc., need to be controlled in
10 response to the user parameter, weight (of the person making the footsteps), since typically a heavier person would produce footsteps which are deeper in pitch, louder, etc. (though this may not always be true in real life). Although generally, the behaviour characteristics will have some
15 bearing on the choice of the synthesizer parameters to be used, there is no hard and fast rule as to how these parameters should be selected. Because sound is somewhat defined by human perception and also because there are many subtle variations, the study of the behavioural
20 characteristics of a phenomenon may not always reveal sufficient information to determine how synthesizer parameters should be used for a given situation. For instance, it has been found that to produce the best sounding effect for horse's hooves, it is helpful change the tonal
25 quality as the speed increases. The relationship between tone and speed is not necessarily obvious. Hence, some empirical experimentation may need to be performed.

To illustrate the principles provided above further, the FIGS. 5 through 13 are tables charting the various parameters that are used for some actual sound models. Taking the table in FIG. 5 as an illustrative example, the first column lists the user parameters plus "random fluctuations" (see above for description for "random fluctuations"). The subsequent columns have a heading at the top showing the engine and synthesizer parameters, the engine parameters comprising the first three columns subsequent to the user parameter column. The "X" in a box indicates that the parameter in that column was used for the sound modelling for the user parameter found in that particular row.

These tables show how changes in user parameters affect a) aspects of event patterning, and b) different synthesizer parameters. Also (on the first row of each table) they show which parameters are affected by the small random fluctuations that are introduced to provide naturalness and variety in the sound effect, even where there is no change in a user parameter.

In FIG. 7, the user parameters, Break, Clutch, and Gas Pedal, control two "internal" variables, car speed and engine speed. The two internal variables are governed by a 2D differential equation with the Pedal settings as inputs. The car speed and engine speed in turn control the synthesizer event generation. The engine speed controls the firing of pistons, each firing is a separately triggered event (many hundreds

per second). The car speed controls the "rumble" of the car strolling along the road.

In FIG. 13, the wind sound model consists of a very small
5 number of events. Most of the user parameters (and the random fluctuations) affect the same set of synthesis parameters, i.e., volume, pitch, filter, etc., but they affect them in different ways. For instance, "Strength" controls the mean value of the parameters (stronger wind has
10 higher volume, pitch, filter Q, etc). The "width of variation" controls the deviation from the mean (of the same parameters) and "Gustiness" controls the rate of change of the parameters. "wind Strength" also controls the number of layers (e.g., number of "whistles") in the sound.

15

LABEL GENERATION

From the above, it will be apparent that the system is capable of generating a plurality of different sounds from
20 a generic model by selection of parameter values of that model. In order to provide concurrent automatic generation of content-related information describing a generated sound, text label elements are associated with different ranges of values of the parameters so that selection of a parameter
25 value will automatically select an appropriate descriptive label element. The label elements are then combined with a model label element to form the complete sound label.

The sound label may be associated with the sound it describes in accordance with any suitable means but preferably in accordance with the MPEG 7 or MPEG 9 standards and the label may be attached to a specific time location in any media, for example a movie, where the sound that the label describes is used. The label may be associated with any representation of the sound. For example, the label may be associated with the actual sound either in digital file form or analog file form. Alternatively, the label may be associated with a file of the control codes provided by the model control system or controlling application to the synthesizer. There may also be circumstances where the actual model used to generate the control codes for the synthesizer would be available where the sound is to be reproduced. For example, a multimedia document may have access to a database of sound models (or references thereto), so that the sound can be specified simply by the particular selected parameter values of that model, with the sound label being associated with those selected parameter values and used to search the database for the required model.

The structure of a model can be viewed as:

1. An object and/or event represented by the model name; and
2. A list of attribute/value pairs represented by the parameter name and setting.

The English grammatical structure of a label similarly consists of a subject (object or event) and attributes of the subject. In the labels generated in this embodiment, the "root" object or event is specified by the model label
5 element with a specification of attributes of the root specified by the value label element(s).

As an example, consider the label "small dog barking loudly". The grammatical structure of the label of the sound and the
10 structure of the sound model can be seen to be similar. The model-related object/event (dog barking) is easily identifiable, while the description's adjective and adverb (attribute specifications) typically encapsulate a model structure's attribute/value pair ("loudly" might be
15 expressed in the model structure as volume=1 (on a scale of 0-1), and "small" (on the same scale) might be expressed as size=.2).

The adjectives and adverbs that constitute the attribute
20 specifications in a label can be expressed using a structure that is closer to the model structure attribute/value pairs. The example noted above could be restated as: "Small sized dog barking with high volume." while lacking eloquence, the model structure can be viewed more clearly (size = small,
25 volume=high) since the attribute and values are both specified in the text description. The degree of explanation/eloquence in a label is a matter of design choice.

To translate a numerical value from the model into a label element, the allowable range of the parameter is divided into segments or ranges, and each range is given a name. The value label element is then produced by combining the range name
5 with the parameter name, using the range as a modifier for the parameter.

Consider as an example a model of footsteps with speed as a parameter. The allowable range for speed, say 0 - 12 km/hr,
10 is divided into three equal parts the first being labelled "slow", the second "medium" and the third "fast". The footsteps model with the speed parameter set to .5 can now generate a description of the sound it generates as "Footsteps at medium speed".

15

For some label elements, it is necessary only to specify the range name as the label element. Using again the footsteps model, if a surface is specified, for example "concrete", there may be only two ranges, specifying no concrete surface
20 or the sound of walking on a concrete surface. When the latter range is chosen, it is sufficient for the label element simply to be referred to by the range name, i.e. "concrete".

25 Another component of the label translation allows certain parameter value settings to prevent the parameter from participating in the sound label. Consider a footsteps model that has a "limp" parameter. One might have labels for

certain values of the attribute (e.g. slight, severe), but if the value were such that there were no perceivable limp to the footsteps generated, there would be no reason to include the parameter in the description.

5

The model description elements are also customizable. There are many parameters that are common across a wide range of sounds (eg "speed"), and the range segments might be given useful default text labels. However, new models often
10 require unique parameters, or else a standard parameter name like "speed" might require model-specific labels (speed for walking, trains and cars might use "slow", "medium", "fast", while speed for a wind might more usefully be translated into "gentle" and "strong"). Customization of specific labels
15 for specific parameters is preferably provided, therefore, by user-defining label elements and the corresponding ranges.

Referring to the footsteps model, the graphical interface for which is shown in Fig. 3, an example of construction of a
20 label from the parameter settings of the model will now be described. Only some of the parameters are used in this example and the parameters used to form the label are a matter of choice depending of the information required in the label.

25

The root (model label element) is "footsteps". The parameter values are divided up into ranges and a text label element is associated with each range depending upon the information

desired.

As noted above, it is desirable to suppress reference to a parameter if that parameter falls within a certain range, for example if the effect is not or is only slightly invoked in the sound produced by the model and a value represented by <N.A> below is associated with such a range. The parameter values lie between 0-1, except for "weight" and "limp" which extend from -1 to 1. Each label is used for a range commencing with the adjacent value up to the next value/label pair, or the top of the range, if appropriate.

Root: Footsteps

15 Param = Speed

0	<N.A>
0.25	Slow
0.6	Fast
1	Running

20

Param = Style

0	<N.A>
0.3	Heel to toe
0.6	Normal style
1	March

25

42

Param = Limp

-1	Limping left leg
-0.4..0.4	<N.A>
1	Limping right leg

5

Param = Stagger

0	<N.A>
.5	Unsteady
.8	Drunken

10

Param = Weight

-1	<N.A>
-0.5	Light weight person
0.5	Average weight person
1	Heavy weight person

15

Param = Concrete

0	<N.A>
1	concrete

20

Param = Creaky floor

0	<N.A>
1	creaky floor

25 Param = Deck

0	<N.A>
1	deck

43

Param = Dirt

0	<N.A>
1	dirt

5 Param = Grass

0	<N.A>
1	grass

Param = Gravel

10	0	<N.A>
	1	gravel

Param = Leaves

	0	<N.A>
15	1	leaves

Param = Metal

	0	<N.A>
	1	metal

20

Param = Mud

0	<N.A>
.6	muddy

25 Param = Sand

0	<N.A>
1	sand

44

Param = Snow

0 <N.A>

1 snow

5 Param = Tile

0 <N.A>

1 tile

Param = wood

10 0 <N.A>

1 wood

The text elements are put together using a model-specific "template" description which is a list of control codes, character strings and parameter names that define how the text components from the model and its parameters are strung together into a sentence-like description of the sound.

The template contains a place-holder for the text corresponding to the root and each model parameter, as well as quoted strings that "glue" the text chunks into a pseudo-sentence. The heart of the footsteps template is:

<<speed>> <<style>> <<stagger>> <<weight>> <<root>> ||{limp}
 25 "with" <<limp>>|| "on" <<mud>> <<concrete>> <deck>> <<wood>>
 <<creakyfloor>> ... other surfaces
 where:

<<...>> indicates the position for text for the parameter named between the angle brackets

"..." indicates a literal string

|| {X}|| is a conditional statement indicating that if
5 the parameter named between the curly brackets is to appear
in the description (in turn determined by the parameter
value), then the rest of the data enclosed in || || is
expanded into the description. Here this applies to the
<<limp>>, with the conjunction "with" being conditionally
10 added if <<limp>> is to be included. It is assumed that a
surface will always be specified, so "on" is not conditional
in this example.

Thus, if the footsteps model were generating sound with the
15 following parameter settings:

Speed = .3	=	Slow
Stagger = .9	=	drunken
Mud = 1	=	muddy
Concrete = 1	=	concrete

20 [weight = -1 and the rest = 0, hence N.A. = blank]

Then the following label of the sound would be generated:

"Slow drunken footsteps on muddy concrete".

25

In practice, the parameter ranges and respective labels, as
shown above, for the "footsteps" model, are stored in a file.
The same file also stores the template. To generate a label,

the system calls a function that compares the current model parameter settings with the file and constructs the label using the template accordingly.

- 5 The embodiment described is not to be considered limitative. For example, although the label has been shown constructed in an intuitive, grammatical way, this is not essential. For example, the label may simply comprise label elements combined in a semi-grammatical way or even as a selection of
- 10 grammatically separate descriptive elements which together form the defined label.

The present invention may be embodied in other specific forms without departing from the scope thereof. The present

15 disclosed embodiments are, therefore, to be considered in all respects as illustrative and not restrictive, the scope of the invention being indicated by the appended claims.

CLAIMS

1. Apparatus for labelling a sound or a representation thereof, the apparatus comprising a sound generator capable
5 of generating a family of sounds by selection of values of parameters of a sound model, at least some parameter values being associated with descriptive labels whereby selection of the value automatically selects the corresponding label.
- 10 2. Apparatus as claimed in Claim 1 wherein the values of each parameter are divided into a plurality of ranges, the labels being associated with respective ranges.
3. Apparatus as claimed in Claim 1 or Claim 2 wherein the
15 value labels are combined with a model label indicating the identity of the model.
4. Apparatus as claimed in Claim 3 wherein the value and model labels are combined in a grammatical or semi-
20 grammatical structure.
5. Apparatus as claimed in Claim 4 wherein the value labels qualify the model label.
- 25 6. Apparatus as claimed in any one of Claims 3 to 5 wherein the value and model labels are combined using a template defining how the labels are combined.

7. Apparatus as claimed in Claim 6 wherein the template specifies the relative position of each label.

8. Apparatus as claimed in Claim 6 or Claim 7 wherein the
5 template specifies text to be used between labels.

9. Apparatus as claimed in any one of Claims 6 to 8 wherein the template includes conditional statements for inclusion of a label and/or text.

10

10. Apparatus as claimed in any one of the preceding Claims wherein the parameters include values not associated with any label.

15 11. Apparatus as claimed in Claim 10 wherein said values not associated with any label include values for which the parameter has little or no effect on the generated sound.

12. Apparatus as claimed in any one of the preceding claims
20 wherein the sound or representation thereof is in the form of a digital audio file.

13. Apparatus as claimed in any one of claims 1 to 11 wherein the sound or representation thereof is in the form
25 of an analog audio file.

14. Apparatus as claimed in any one of claims 1 to 11 wherein the sound or representation thereof in the form of

control codes for a synthesizer.

15. Apparatus as claimed in any one of claims 1 to 11
wherein the sound or representation thereof is in the form
5 of the selected parameter values for the model.

16. A method of labelling a sound or a representation
thereof comprising the steps of: selecting a sound by
selection of values of parameters of a sound model, at least
10 some parameter values being associated with descriptive
labels whereby selection of a value automatically selects a
corresponding label, generating the sound or representation
as a file and associating the file with the label.