THE CREATION OF MOVEMENT AND SPATIAL DIMENSION IN STEREO RECORDING

A RESEARCH REPORT SUBMITTED

IN PARTIAL FULFILMENT OF THE

REQUIREMENTS FOR THE DEGREE

MASTER OF MUSIC (MUSIC TECHNOLOGY)

SCHOOL OF GRADUATE STUDIES

MUSIC

FACULTY OF HUMAN SCIENCES

UNIVERSITY OF NATAL

DURBAN

BY

MICHAEL JOHN WILKINSON

AUGUST 1999

TABLE OF CONTENTS

List of Figures. . . . .

Introduction. 1

2 Theoretical Considerations. 3

2.1 Localization. 3

2.2 Spatial imaging perception. 4

2.3 Simulation of horizontal cues. 6

2.4 Distance perception 8

2.5 Environmental Perception. 9

2.6 Simulation of distance perception. 9

2.7 Simulation of vertical cues. 12

2.8 Simulation of environmental cues. 14

3. Implementation in *Valley of dry bones* 16

4. Conclusion 21

Bibliography. 25

Appendices. 26

Compact disc *Valley of dry bones*

List of Figures

1. Binaural distance perception 3

2. Crosstalk images 5

3. Normal stereo listening position. 6

4. Delay and Level differences between two speakers. 7

5. Environments. 14

Introduction

"The 'art' in recording is the sensitive application of the recording process to shape or create sound as, or in support of, an artistic, creative message" [Moylan.1997:12].
In order for audio producers to be in control of their material, they must understand the substance and behaviour of the material. They must also understand how sound changes with alterations in structure and through influences of changing environments.

The perception of space in audio recording is not the same as the perception of space around an acoustic source in a physical environment. In an acoustic space, listeners perceive the location of sound and certain spacial cues about the apparent acoustic environment in relation to the three-dimensional space around them [Moore.1989:559]. In audio recording, illusions of space are created where sound sources are assigned spatial information through the recording process and/or through signal processing. The spatial information is intended to simulate particular known physical environments, such as concert halls or it is intended to provide spatial cues that have no relation to our reality, such as small rooms with long reverberation times. Stereo recordings can portray the spatial characteristics of recorded sounds independent of the listening conditions. Humans can discern properties of one environment through loudspeakers while we move about in another, however, this does cause some difficulties with spatial interpretation as the original environment of the source is distorted by the listening environment.

To accurately perceive the spatial information of an audio recording, the listening environment ideally needs to be acoustically neutral. The listener must be carefully positioned both within the environment and in relation to the loudspeakers so as to prevent further distortion of the original signal in the playback environment.

Ideally, spatial placement of sounds would allow us to have complete control over the acoustic environment heard through loudspeakers. Each location within the heard environment could have a specified size, direction, distance and motion. Computers and their relevant peripherals and software allows us to have this control over the spatial characteristics of sounds. Much research and development is taking place in recording and playback systems, in the attempt to simulate the dimension of space as in natural hearing [Moylan.1997:12].

This research report will endeavour to illustrate through design in my composition *Valley of dry bones*, the creation of spatial movement and dimension in stereo recording. This dimension and movement goes beyond the usually restricted horizontal stereo image to an area that stretches both horizontally (width), vertically (height) and in distance (breadth), beyond the arc of the two speakers.

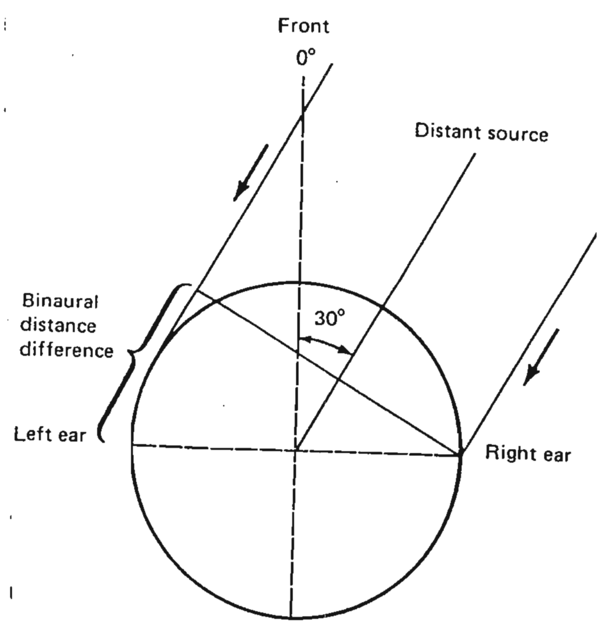
2. Theoretical Considerations

2.1 Localization

In order to understand the brain's interpretation of audio localization and distance, the human perception of the placement of a sound source in the real world has to be understood. This type of perception is known as sound localization, and is dependent on cues for interpreting three dimensions, namely the horizontal angle, distance or velocity and the vertical angle. In determining the horizontal angle of sound listeners use three cues:

The first is based upon differences in time and intensity reaching the two ears. Any delay perceived between the time a sound reaches the opposite ear to the one first receiving the sound is called interaural time difference (ITD), binaural time differences or sometimes referred to as phase differences. The listener can use ITD cues to determine the angular direction, measured in degrees, of a sound source. If the sound is centered behind or in front of a listener, the ITD is zero. As the angle is changed by more than one degree (about 20 microseconds), a difference in direction can be perceived, until the sound moves toward a lateral position in front of the listener [Brown and Deffenbacher.1979:217]. The change in Binaural distance perception as the angle either increases or decreases is illustrated in Figure 1.

Figure 1



Binaural distance perception

[Brown and Deffenbacher.1979:217]

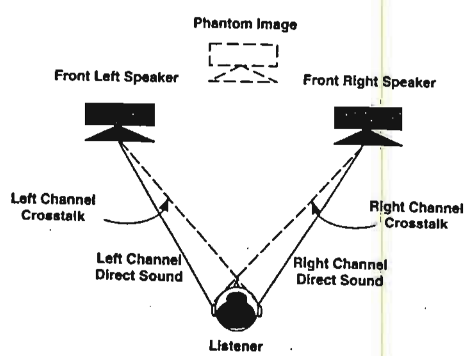
Interaural intensity differences (IID) provide another cue for determining direction. When the sound source is centered there is no intensity difference as the sound reaches both ears at the same time. When the sound source is not centered, the listener's head partially screens the ear opposite to the source, casting a shadow that diminishes the sound received by the ear, particularly at high frequencies.

Spectral cues provided by reflections off the outer ears, shoulders and upper torso also give directional cues. The ear flaps help orientate the source's angle of direction. Since the pinnae shield the external auditory canals from behind, above and below, higher frequencies are softer from other directions than from the front and subsequently front-back confusion decreases with increases in frequencies above 3000 kHz. Elevation experiments done by Martens [Begault 1994:75] using high and low sounds show that stimuli using higher frequency sounds tended to be heard to the front and those at lower frequencies towards the back. Experiments done by Stevens and Newman demonstrate that low frequency sine waves are ambiguous as to their front-back location when listened to through loudspeakers. [Begault.1994:75] Both ITD and IID cues are ineffective when the spectral energy of the sound resides below about 270Hz. Interaural time difference cues are most effective in the frequency range of 270 to 500Hz, but are useless as cues above 1400Hz where the shorter sound waves are reflected off the head and lost before reaching the opposite ear. A frequency below 50 Hz produces no major IID cues. Intensity differences increase with frequency so that above 1400Hz IID cues predominate. This is the reason why high pitched sounds with a sharp attack slope have good localization accuracy compared to tones with a softer curve.

2.2 Spatial imaging perception

The spatial relationships of reproduced sounds are conceived by the listener, through the concepts of the sound stage and imaging. The recording represents an illusion of a live performance which is conceived by the mind as existing in a real known physical space. Moylan says that "within the perceived performance environment is a two-dimensional area, both horizontal and distance" [Moylan.1997:211], but a third dimension could be included in the stereo sound stage, which is created by the illusion of the vertical movement and projection of sound within and beyond the arc of the normal stereo speaker positions. The area of the sound stage can be of any size, from a small area to an area filling a space beyond the stereo array. The placement of the front edge of the stage needs to be as close as possible to the listener with the furtherest sound marking the end and depth of the sound stage. The imaging of sound sources will be influenced by the characteristics of the unique performance environments of the individual sound source, as well as their placement within their own environment. In stereophonic reproduction the spatial distribution of the auditory stimuli is enabled through the creation of so called phantom images, which are not related to the hearing of natural sound. A phantom sound source is created exactly in the center between two stereo loudspeakers when both speakers emit identical signals simultaneously (Figure 2). This phenomenon is known as cross-talk which effectually limits the placement of auditory images between speakers.

Figure 2



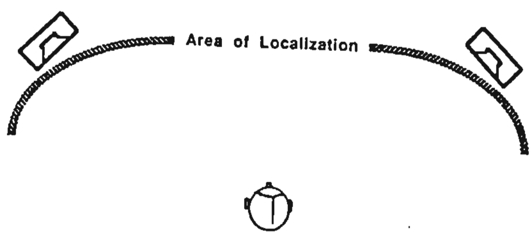
Placement of Phantom images

[Holman.1996:42]

2.3 Simulation of horizontal cues

Panning is used to describe the movement of sound image across a stereo field. As the source passes from one loudspeaker to another, the amplitude in the direction of the target loudspeaker increases, while the amplitude in the direction of the original source loudspeaker decreases. Initially a sound source from a pan-pot has zero width if panned towards the center. With the addition of stereo reverberation and stereo panning a greater width is obtained much like that obtained with real sources using a coincident pair of microphones, which works purely by amplitude differences at the two loudspeakers. As both ears hear the speakers the result is that the space between the speakers and the ears turns the intensity differences into time of arrival differences, giving the illusion of depth and distance as in natural hearing. Many different panning curves are possible, each giving a slightly different spatial impression of sound movement. For a symmetrical pan these curves assume that a listener sits in the exact center between the two loudspeakers. This is known as the normal stereo listening position as seen in Figure 3.

Figure 3



Normal Stereo listening position

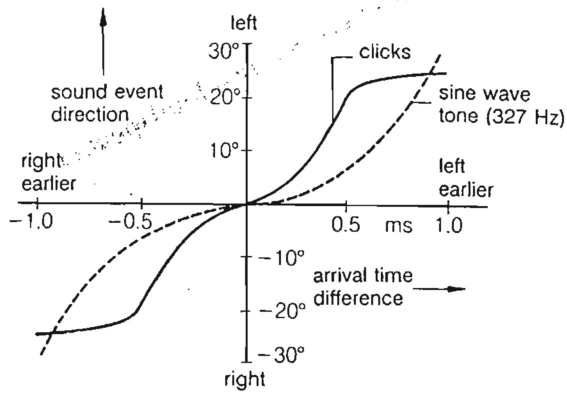
[Moylan.1992:27]

With linear panning a "hole in the middle" effect occurs since the ears tend to hear the signal as being stronger in the loudspeakers than in the middle where the intensity drops, due to the Law of Sound Intensity, which states that the perceived loudness of sound is proportional to its intensity [Dodge 1985 :460].

Constant power panning uses sinusoidal curves to control the amplitude emanating from the two loudspeakers. This creates the impression of a pan with a more stable loudness. The perceived pan is seen as rotating between the two loudspeakers at a constant distance from the listener, and heard at a constant sound level as it moves across from one side to the next. Another technique used to simulate movement is to use increasing time delay or level difference between loudspeakers. Here, the phantom source moves away from the center until it appears to come to rest in one of the loudspeakers. The louder or earlier arriving signal determines toward which loudspeaker the phantom sound source moves (precedence effect). The required time delay for a particular angular deviation (Figure 4a) depends somewhat on the type of signal. A time of arrival of 1 millisecond results in the greatest possible deviation of 30 degrees from the center in a usual stereo loudspeaker setup and even for small deviations sound sources appear to come from the nearest speakers.

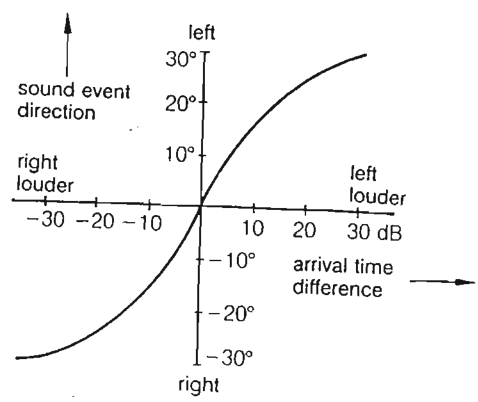
Figure 4

Delay difference between two loudspeakers for stereo reproduction and the resulting deflection of the phantom sound source



a. Delay differences between two loudspeakers

Level difference between two loudspeakers for stereo reproduction and the resulting deflection of the phantom sound source



b. Level differences between two loudspeakers

[Dickreiter.1989:80]

Level differences and IID cues lead to more stable phantom images. Level differences of 15 to 20 decibels permit the images to deviate to the extreme sides where they become stable sound sources in one of the loudspeakers, as shown in (Figure 4b).

The so-called stereo horizon within which true localization of phantom sources is possible is only 20- 40cm wide for normal stereo setups [Dickreiter.1989:86]. Otherwise the well known imaging problem occurs, where, when we move across the listening area the sound images that are meant to stay centered, move with us [Holman.1996:40].

2.4 Distance perception

Distance perception is dependent upon the time delay between the arrival of the direct sound and reverberant energy and upon the ratio of the amounts of direct sound versus reverberant sound [Moylan.1994:129]. Knowing the timbre of the sound will help the listener in perceiving the reiterations of the direct and the reverberant sound. Sounds are also judged according to their context with other surrounding sources. Distance sensation is improved markedly in enclosed spaces through evaluation of the ratio of direct to diffused sound, permitting recognition of even small distance difference. The listener must know the timbre of a sound in order to recognise that the sound is missing detail. The listener's brain will further calculate how much low energy information is missing, in order to determine the degree of distance. This perception will assist in calculating the amount of distance between the sound source and the listener. Previous experience and the listening skill level of the listener will play a major role in determining the accuracy in which distance judgements are made. Without prior knowledge of the timbre of a sound, perception of distance location is considerably less accurate. Through time the listener learns to know the sound qualities of sound sources within their immediate area. The space surrounding them serves as a reference from which to judge distance "near" and "far". Within this area of proximity, sounds may be altered by the environment although they will still have a great level of detail depending on how close the sound source is to the listener. Humans conceive "near" as being immediately outside of their area of proximity where sounds are moderately altered and are able to localize the sound's distance with detail and accuracy. Sounds cease to be considered "near" when the listener begins having difficulty with localizing because of their diminishing detail. The furthest "far" sounds may even be more difficult to recognise, but will have natural timbral alterations which could occur in nature as the reverberant energy which occurs over a deep valley. Other sounds may appear to be located at a distance outside of human experience, a distance beyond our world "otherworldly". These sounds have alterations in timbre which may have frequencies emphasized and de-emphasized which would not normally occur in nature, and become almost unrecognisable as a known sound [Moylan.1992:221].

2.5 Environmental Perception

"Environmental characteristics are determined by the listener, through comparing his or her memory of the sound source's timbre outside of the host environment to the sound source's timbre within the host environment" [Moylan.1992:227]. In the case of never having any recollections of the original sound source's timbre an approximation or comparison would have to be made.

" Differences in the spectrum and spectral envelope of the sound source, as remembered by the listener, and as heard in the host environment, form the basis for determining most environmental characteristics" [Moylan.1992:228].

2.6 Simulating distance cues.

To make a sound reside into the distance one can use techniques such as lowering the amplitude, apply a lowpass filter, adding echoes, or blending in reverberation. Amplitude increase and decrease can be simulated by using either a loudness control or by using compressors and expanders. Compression is used to describe conditions of gain reduction where the original dynamics are compressed or reduced. By the use of compression ratio controls, which simply specify the relationship between input and output levels, many different effects can be achieved. The attack and release time play a vital role in determining the effects of the compressed sound. The attack time will determine the characteristics and size of peaks allowed to pass through the system prior to attenuation; in effect it will dynamically modify the static sinewave response of the compression ratio [Belville.1977: 28]. As the attack time lengthens, a subtle change takes place in the spectral energy balance as increasingly high frequency content passes unattenuated, giving the sound a brighter timbre. Release time is important since it determines the moment-to-moment gain change, which in turn controls loudness. Under conditions of considerable compression and very fast release time the medium and lower level signal content is raised to peak level and subsequently sounds louder. In most recording applications, the purpose of an expander gate is not to expand the music, but to get in below the low-level signal and attenuate the channel gain in the presence of noise only [Belville.1977: 30]. Because of this attenuation in noise level the audio will sound brighter and clearer and particularly with short impulsive sounds or sounds with a sharp attack which can be created by using expanders are most easily located for distance cues as the initial attack determines the distance by its reverberation qualities. Long tones have the opposite effect and are difficult to estimate in distance perception. Intensity variation itself is not always a reliable cue to distance judgements. The amount of sound power at the eardrum can vary when distance does not. Loss of effective power with distance can operate as a good cue only when there is little or no reflection, such as in a anechoic chamber or over headphones [Begault.1994:51].

Loudness plays a major part in the perception of distance. In nature distant sounds are often much softer than near sounds, but may also appear to be louder in some cases. If, for instance, we were seated in an apartment in a city, the sounds of passing traffic would appear louder than the sound of a fridge motor coming from the nearby kitchen. Even though the fridge sounds quieter, it is perceived as being nearer because of its known position and timbre quality and not because of its loudness. The traffic appears louder due to the predominance of lower frequencies; higher frequencies which are shorter are lost into the atmosphere and reflected off buildings. In this instance timbre quality and reverberation time gives us the cue for its distance perception.

A fade out gives the illusion of a sound moving into the distance because of its decrease in loudness and change in timbre quality. Higher frequencies above the most sensitive hearing region (2700-3200Hz) need greater amplitude to get to the same loudness of lower frequencies. This theory is demonstrated in the constant-loudness contours or Fletcher-Munson curves [Roads.1996:1056]. Because of the drop in dynamic level during a fade out the higher frequencies are the first to lose their presence and subsequently lower frequencies predominate and give us this timbre change. Distance is not always accurately perceived; other elements such as timbre and reflection are often confused with the perception of distance as being degrees of loudness.

Timbral detail is an important cue for distance in relation to the characteristics of the environment in which a sound source is produced, and the perceived location of the sound source and the listener within that environment. In a microphone setup the closer the microphone is to the source, the greater the definition and timbre detail because of fewer high frequencies getting lost between the source and microphone field. By moving away from the microphone a greater distance is perceived because of this lack of definition created by the dominance of lower frequencies and indirect sound which might have reflected off a wall or obstacle in its path.

Timbre differences between the sound in its normal unaltered state and its state in the host environment are the primary determinants of distance location. Humans rely on timbral definition for most of their distance judgements and to a lesser degree, on the ratio of direct-to-reverberant sound. Distance perception is really a combination of both, as the amount of reflection affects the colour of the sound, by the presence of higher or lower frequencies. Equalizers are an important tool in changing tone colour within a sound; they either add presence and focus to a sound by boosts in certain higher frequencies or distance by certain lower frequency. Special filters such as high and low pass filters are also used to create distance and presence by attenuation of lower and higher frequencies respectively [Roads.1996:185].

In nature, distant sounds are often accompanied by a great deal of reverberant energy, which is important as an attributer of environmental characteristics, and for placing a sound source at a distance within the individual source environment. This reverberant energy can be controlled electronically by the use of digital reverberation modules where the depth, delay and length of reflection are manipulated according to desired specifications.

The sound source's degree of timbre definition, the perceived distance of the sound source within its host environment, and the perceived distance of the source's host environment from the perceived location of the listener, are combined into a single perception of distance location. This process will determine the actual perceived distance of the sound source. Research by Begault based on the work by Coleman, Sheeline and Von Be'skesy demonstrated that the relative mixture of direct to reverberant sound is a powerful cue for determining distance [Begault.1994:106].

2.7 Simulating Vertical Cues

Research shows us that high- frequency sounds reflecting off the outer ears and shoulders provide a critical cue to vertical localization. The surfaces of the pinnae and shoulders act as reflectors, creating short time delays that are manifested in the spectrum as a comb filter effect [Begault.1994:63].

The most significant influence on localization has to be the spectrum of sound as it reaches the outer ears, where, due to the shape and folds of the outer ear, the sound is filtered especially those high frequencies above 5 kHZ. This function by the outer ears is termed head-related-transfer-function (Hrtf). The folds of the ear cause minute delays with a range of 0-3 microseconds. These resonances and diffractions cause the spectral content at the eardrum to differ from the sound source. The Hrtf alters the spectrum and timing of signal, which is then recognised as a spatial cue. The shoulder, torso and ear canal also have an effect on filtering. Furthermore there are changes in interaural phase as a function of frequency that are caused by the Hrtf's. The Hrtf relates the transmitted source sound pressure developed at the ear drum. It varies with frequency, azimuth, elevation and range, and reveals the physical cues of sound localization [Begault.1991:57]. The directional aspects of the pinnae are considered to be particularly important for vertical localization. Several studies have shown that without the pinnae's effect on a broadband frequency source, vertical localization is less accurate.

Another kind of head shadow intensity difference is front to back. Here the ear flaps can help orientate the angle of the sound source. Since the pinnae shield the external auditory canals from behind, above and below short wavelengths, higher sounds are softer from other directions than from the front , and front back confusion decreases with an increase in frequency above 3000 kHz.

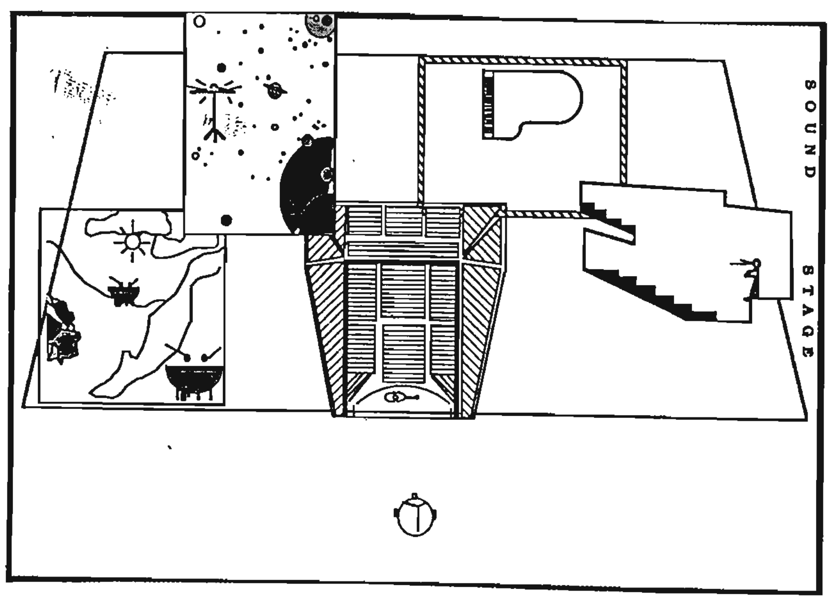
The understanding of filtering techniques for artificially produced sounds in space was anticipated by Bauer and Batteau [Begault.1991:70], who used analogue filters designed to produce the necessary elevation resonances of the pinnae from localization experiments [Begault.1991:71]. The simulation of vertical movement can also be created by using microphone techniques and/or by emphasis or de-emphasis of certain frequencies using filter techniques. Vertical cues can be simulated electronically, giving the impression that sound is arriving from above. This is done by filtering the input signal, imposing the change in spectrum caused by reflections off the head and shoulders. The use of Hrtf spectral shaping is a feature used in processing 3D sound. This is based on the theory that the most accurate means to produce a sample sound cue is to transform the spectrum of a sound at the eardrum as closely as possible to the way it would be transformed under normal spatial hearing. In this electronic synthesis of spatial sound, the electronic processing must create the effect of sound waves reacting with a real human head and ears, hence the sound must be transformed according to the properties of the Hrtf. However, this does cause problems as the listener's transform might be different to that of the original.

Directional perception in the vertical plane depends on the sound coloration, being a function of its vertical angle. If a signal moves upward in the vertical auditory plane, the level and time-of-arrival differences remain just about unchanged. The information about the rising sound source is mainly taken from the change in timbre caused by the diffracting effect at the head and ear lobes. This effect can be simulated by the use of filters on a broad band sound, primarily a sound with a good range of high frequencies. Sounds coming from different directions obtain direction specific boosts in certain frequency bands, the so called direction determining bands. Narrow band signals cannot have such complex spectral changes, therefore localization becomes difficult. Localization in the vertical plane depends clearly on the familiarity with the sound source and on the experience of the hearing mechanism. Localization in the vertical plane is not possible for narrow band signals, the signal direction is determined from the signal frequency.

2.8 Simulation of environmental cues

In current music an illusion of space existing within another space is created by each instrumental source being placed in its own host environment, within a larger overall environment or space. These environments change the timbre of the sounds as they appear in different depths and dimensions within the larger space.

Figure 5



Enviroments

[Moylan.1992:210]

Figure 5 presents different sound sources appearing to be performed in very unique environments:-

- Timpani located in an open air environment
- A string instrument placed in a large concert hall.
- A vocalist performing in a small hall.
- A piano sounding in a small room.
- Cymbal appearing to exist in a very unnatural environment.

The characteristics of the environment have specific component parts that contribute to its own unique sound quality. These characteristics are what must be determined, by identifying the differences between the sound quality of the sound itself and the quality of the sound source within the environment. These component parts are: the "reflection envelope", the "spectrum", and the "spectral envelope" [Moylan.1994:228].

The reflection envelope is created by the amplitudes of the initial reflections and reverberant energy of the environment. Without the opportunity to hear the environment's complete presentation, information related to the reverberant sound energy can never be audible. Computer aided spectrum waveform programs such as 'Cool Edit 96' help in this situation.

The "spectrum" of the reverberant sound and the initial reflections is a composite of the frequencies or bandwidths of pitch areas that are emphasized and de-emphasized by the characteristics of the environment itself. These are determined by carefully evaluating many appearances of the sound source in the environment and by listening to the way the sound source's timbre is changed by the environment over a wide range of pitch levels.

The "spectral envelope"of the environment is how frequencies that are emphasized and de-emphasized by the characteristics of the environment itself vary in loudness level over the duration of the sound environment.

3 Implementation of *Valley of dry bones*

By designing a stereo location graph according to Moylan (Appendix 1), one can plot the horizontal locations of all sound sources against a time line of the work. Such a graph was constructed for the composition *Valley of dry bones,* where I attempted to simulate dimensions and movement in sound through techniques discussed. The graph portrays the angle and direction of sources from the listener (positioned in the centre of the listening environment). The sources have been placed at approximate locations between the speakers. The boundaries of the horizontal axis are marked by the "left" and "right" loudspeaker locations, but may extend slightly beyond the speakers.

Once the direction and movement of sources has been decided upon, it's stereo location is then designed and created in the mixing process, as time and amplitude differences between the two loudspeakers and the listener's ears.

Three basic types of signal processing were used; each process is designed to function as a particular physical dimension of sound. An alteration in one of the physical dimensions of sound will cause a change in the other dimensions. The three processors used did not only cause audible changes in the physical dimensions but also altered timbre. They were:-

- Frequency processors: equalisers and filters.
- Time processors: delay and reverb units.
- Amplitude processors: expanders and compressors.

All mixing and processing was done on computer. The software program called 'Cool edit 96' was used to do most of the frequency and amplitude processing. The time and some frequency processing was done on the audio recording program "Cubase VST" and its automated mixing console was used to mix the audio for the desired effect.

Mixing is a process that combines sounds, creating an artistic blend of timbres, dynamic levels and assigned spatial location, distance and environment qualities and through signal processing, finalizes the shaping of the sound qualities. By combining simulation cues for angular location, distance and velocity (Doppler shift) one can create convincing illusion of moving sound sources between two speakers.

In my composition *Valley of dry bones*, the guitar feedback sound in bars 9 - 12 at 17 seconds gives the illusion of moving across the sound stage and back again. The sound first appears in the left speaker and as the sound decreases in amplitude the source moves towards the middle of the two speakers. This creates a phantom source in the centre of the listening area as an equal volume of sound is coming from both speakers.

A second wave or feedback sound follows the first wave and this moves across to the right speaker. This perceived change in direction is also enhanced by the fact that the second wave is much stronger than the first and that our ears are now directed towards the second wave front and this together with panning shifts the perception of the sound to the right speaker. As the initial sound of the second wavefront decreases the sound once again appears to come from the middle. It then moves into the distance as it is panned to the left speaker.

The results taken from a fast Fourier transform [Dodge.1885 :51] and spectrum analysis done on this effect, show that the initially lower frequencies are accentuated (500hz,1500hz,) as shown in Appendix 2 (1a.) The directional properties of the higher frequencies which are emphasised through the feedback sound, together with the increase in loudness, could possibly give the illusion of the sound source coming from the nearest speaker, in this case the right speaker.

Basic localization cues for static sounds can be extended to the simulation of moving sound sources. The Doppler effect describes the change in pitch that results when the source and the listener are moving relative to each other [Dodge.1985:46]. When a sound source and a listener are moving closer together the wavefronts of the sound will reach the listener more frequently, causing the perceived pitch to be raised. If the source and listener are moved apart, the pitch will appear lower.

This illusion of approaching and departing can be heard in the electronically produced effect of bars 42-46 at 1:24 seconds on the cd. The sound sources amplitude level and reverberation level were increased and decreased respectively as the sound was panned towards center stage. This together with a treble boost gives it the presence it needs to create the illusion of an approaching sound source, simulating the effect of sound waves becoming more and more compressed and rising in frequency as is the case in the Doppler effect [Vermooten.1980:12]. The pass-by effect made use of a lower frequency boost plus the addition of reverberation, which simulated the effect of a sound wave de-compression as if moving off into the distance as the sound was panned from left to right. A crossfade was done between the faders of the original and reverberant signal. Doing this gives the illusion of sounds moving to and from the listener. Fading between the equalizations in this way also helps to reinforce the illusion of pitch bend, if it was not originally used. By fading between original and reverberated sound and using appropriate equalization on sound sources, it is possible to get sounds moving backwards and forwards. By use of panning, sounds move between the two speakers and by combining these in the right proportions, it is possible to create swirling textures and 3-D movement.

A fast Fourier transform graph was done during the crossfade between the approaching and the close up sound. In Appendix 2 (2a) we can see the higher harmonics (in black) and the lower ones (in red). This boost in the higher frequencies gives the illusion of moving closer to the listener.

The opening voice recitation "spirit" was created using a combination of techniques. The intention was for the voice to project forward very quickly and then to disappear. To begin with the voice part was closely miked, giving it a close perspective. Here its timbre quality gives us the cue for its distance perception and not necessarily its loudness. The word "spirit" was spliced and placed on three different tracks in order to have control over the final processing and mixdown. The first part of the word had its gain reduced and together with added reverberation a simulation of an approaching effect is achieved. In the second part of the word the perspective is very close. This was achieved through using a high frequency equalization with no reverberation. The third part made use of the last syllable "it" where it's gain was reduction together with a lower frequency boost and reverberation. This gives the illusion of moving back into the distance.

The rumble sound in bar 38 at 1:15 seconds is divided into two parts. The first part begins right of middle and has a treble boost and bass cut (this gives the sound the illusion of rising) while the second part has a bass boost and treble cut, giving the opposite illusion. In the second part the presence of lower frequencies gives the illusion of the sound moving slightly towards central position, widening its image as it moves across. The added reverberation in the second part gives the illusion of moving further into the distance.

In Appendix 3 a graph indicates the distance location of sound sources in *Valley of dry bones*. Bar numbers and timings have been given for the sound sources and these distance cues are only approximations of how the sound was perceived.

In context with other sources the keyboard sound was designed to be located in the middle of the sound stage, on the border between near and far. From bars 26-75 at 55seconds it moves a little closer in perspective mainly because of an increase in gain. The bass part together with the bass drum gives us a close perspective: living in their own environment with little reverberation. The compression of the snare drum sound adds a sharp attack to its sound, thus causing the sound to stand out and appear closer to the listener. Because of the amount of reverberation applied to the snare drum it sounds livelier and if listened to very carefully the amount of reverberation places it at a distance in its own environment. The experience of the listener is necessary for judging the environment of the snare drum by knowing how it sounds without reverberation.

In Appendix 3 from the score *Valley of dry bones,* a Doppler effect moves through the far field into the otherworldy distance. Here the added long reverberation time together with amplitude reduction helps create the illusion of the guitar moving into the far distance. This effect is also achieved with the words "enter us, spirit." Here a pre-delay reverberant effect projects the voice into the distance which eventually becomes so muddled that it is difficult to perceive.

In Appendix 2 (1b) we see the effects of the new environment (reverberation) on the host environment. The guitar feedback's frequencies (1000,1500,2500,3000,4000 kHz) can be seen to be emphasized and appear to be much wider in dynamic range, which results in it's closer proximity and increase in amplitude.

In contrast, the Doppler effect shown in Appendix 2 (2a) has more of its lower frequencies emphasized and as a result distance perception is increased. With more lower frequencies than higher frequencies present, the effect has less localization accuracy. The effect of the environment on the Doppler effect of Appendix 2 (2b) can be heard on the CD at 1:24 seconds. Here the added reverberation changes the environment by giving it a much wider and deeper perspective as the sound passes by. The pre-delay effect used in bars (61-64) of Appendix 1 gives the effect of the voice receding into the distance. This can be heard on the compact disc at 2 minutes and 11seconds. The long reverberation time accentuates the lower frequencies and higher frequencies are almost lost completely in a swirl of reverberant energy. In the fast Fourier transform of the breath effect at both 12 seconds and at 1minute 44 seconds in Appendix 2 (4a) one can see amplitude boosts in the higher frequencies (2500-4000 kHz). The frequencies below this range have much lower amplitudes. The desired effect was a "breathy" sound to depict the spiritual world in a way that moves up into the unknown. This "breathy" type noise gives the illusion of rising out of the speakers. Here the delay of the original envelope was put through a high filter boost and this in combination with reverberation gave the illusion of rising beyond the speakers. This technique would simulate the effect of the timbre change as if the sound had been diffracted at the head and ear lobes. In Appendix 1 the metal pole sound in bar 16 at 32seconds, gives the effect of appearing directly above the head, somewhere beyond the speakers. Localization and elevation is very accurate mainly because of its sharp attack time (intensity) and bright timbre, making it a very distinct sound with a close perspective. The intention of the two short fly-past effects in bars 70-72 at 2:27 seconds was to create the effect of swift movement from side to side. A broad bandwidth sound (brown noise) was used to give it the desired clarity I wanted which made localization easier and together with panning this effect was achieved. The short and quick attack time of the effect also added to its accurate direction perception. To begin with the approaching sound source was given a bass frequency boost to create a little distance, its second half, which was spliced and put on a separate track for processing control was given a high frequency boost with no reverberation, this gave it a closer perspective . The second return effect was given the opposite treatment, that is, beginning with a close perspective followed by the perception of moving off into the distance. By cross fading between the original and reverberated sound, the effect of moving to and from the listener was achieved.

4 Conclusion

Firstly, in dealing with the psycho-acoustics of sound we need to consider that the electronic movement of sound is an illusion. And although these techniques are not as effective as a simulation of the real thing such as found in acoustic sound movement, they still present us with many new processing possibilities and control over an often neglected parameter [Strange.1983:214].

However, there does appear to be some problems in implementing spatial movement with two channel stereo. The first is the imaging problem of the sound which is meant to stay in the center as we move from one side to the other but instead moves with us. The only way out of this is to add channels or remain in a central listening position for a true stereo sound picture. The other method of alleviating many channels is to use phantom images, producing sounds from spaces between speakers as demonstrated in *Valley of dry bones*.

Depth is needed to create spatial dimension. The reason for little depth on normal stereo, besides only having two speakers, is that the in and out-of-phase information the ears need in order to determine the depth and direction, are not present. These are lost because playback environments do not match the original environment. Ambisonics might seem to be a solution, as this creates these phase differences, especially by the use of the sound-field microphone where convincing and accurate localized phantom images are created. Ambisonic recordings may be played back over two or four speakers in its simplest form. It uses an encoding system which transforms the two channel recording into a full 360 degrees horizontal front stage by utilizing both amplitude and phase techniques [Branwell.1983:140].

Other problematic regions are no doubt found in loudspeaker reproduction. The reason why this is a problem is firstly that the environmental context of the listening space (reflections) will be superimposed upon the incoming signal to the eardrums. Unlike the direct sound from headphones, the signals at each ear are a mix of the signals from the two speakers. Crosstalk plays a major part in our hearing of spatial sound [Begault.1994:217]. In natural hearing, analysis reveals the importance of the fact that cross-talk signal is effectively filtered by the HRTF before reaching the opposite ear by a 180 degree phase cancellation based on the theory that the addition of waveform 180 degrees out of phase results in a perfectly cancelled signal. It is difficult to know the reflections of speakers in certain settings and to predict the position of listeners or speakers. As one moves away from speakers the angle of incidence narrows, which changes the HRTF necessary for use in cross-talk cancellation. It must also be said that there is some individual variation in perception and not all people perceive the same spatial positioning on playback.

Progress is being made in the area of listening to 3D sound over loudspeakers. There have been developments in monitoring systems such as the BAP 1000, which is a virtual acoustic monitoring system, in that 3D sound techniques are imposed to allow simulation of a pair of stereo speakers in an ideal recording studio [Begault.1994:220]. This device filters the output from a mixing console using transfer functions that represent the HRTF impulse responses of loudspeakers in ideal conditions, at left and right 30 degrees azimuth.

There have been many new developments in computer hardware systems , one such development is the "Spatializer," which is a 3D audio processor developed (by Spatializer Audio Laboratories,) to improve stereo listening. Its processing creates phantom rear speakers based on psychoacoustic principles [http//www.catalog.com]. The technology behind "Spatializer" is similar to that found in the "MS" recording technique which uses the out-of-phase signal to enhance the spatial dimension of the audio. Developers are finding cost effective ways of modelling environments. By borrowing statistical techniques from the field of architectural acoustics they are developing methods for approximating the effects of listening environments [Astheimer 1993: no page]. If they can develop ways of matching the environment of the recording with the playback environment as is experienced over headphones, the dimensions in stereo listening could possibly be much improved.

Dimension and movement within stereo recording can only really cover the space between the speakers and possibly about 15 degrees either side of the speakers. This space is quite adequate as our main focus on hearing is directed towards the front. To get any projection to the back and sides of the head would require a third or fourth speaker. Although dimensional hearing with two speakers is greatly aided by modern technology, it still has a long way to go before it can be compared to natural hearing.

The Binaural recording process has achieved much in the direction of spatial hearing, its recording methods are based on principles found in natural hearing, but is limited by its playback environment which can only be fully realized through headphones.

Holophonics technology presents us with some interesting theories in that full spatial sound can be recorded without the use of traditional microphones. The technology is based on a new understanding of the physiology of human hearing, which is utilized in the recording process. Holophonics is based on the hypotheses that we do not perceive sound in a passive manner, with sound waves impacting our eardrums but rather that the human ear generates its own reference tone . The premise is that this tone interferes with incoming sounds to create the necessary spatial information for analysis by the brain and it is this interference between external sounds and the reference tone, that provides the brain with spatial information [www.holophonics.com/tech.html].

Perhaps with sound recording, processing and monitoring system developments of the future we could discover new spatial interpretations based on psychoacoustic understandings and placement of phantom images over loudspeakers. There are certain limits in stereo reproduction of sound and although the placement of phantom images does give us much movement, dimension and elevation within the normal stereo arch, more speakers are needed to reproduce sounds from other spatial angles.

BIBLIOGRAPHY

Astheimer, P. 1993. What you see is what you hear: acoustics applied in virtual worlds, IEEE symposium on Virtual Reality, San Jose, October, 1993.

Begault, D. 1994. Three-dimensional sound. London: Academic Press.

Belville, M.1997. "Compressors and limiters," Studio sound October: 28-32

Branwell, N.1983. "Ambisonic surround sound technology," Studio sound December: 140

Brown, E and Deffenbacher, K.1979. Perception and the senses. New York: Oxford University Press.

Damaske, P.1971. "Head related two-channel stereophony with loudspeaker Reproduction" Journal of the Acoustical Society of America vol.50, no4:1109-1105.

Dickreiter, M.1989. Tonmeister technology New York: Temmer Publications.

Dodge, C.D. and Jerse, T.A.1985. Computer music . New York: Schirmer Books.

Holman,T.1996. "Channel crossing," Studio sound February: 40-42.

Moore, F. R.1989. "A general model for spatial processing of sounds," The Music machine, ed.Curtis Roads. London: The MIT Press, 559-568.

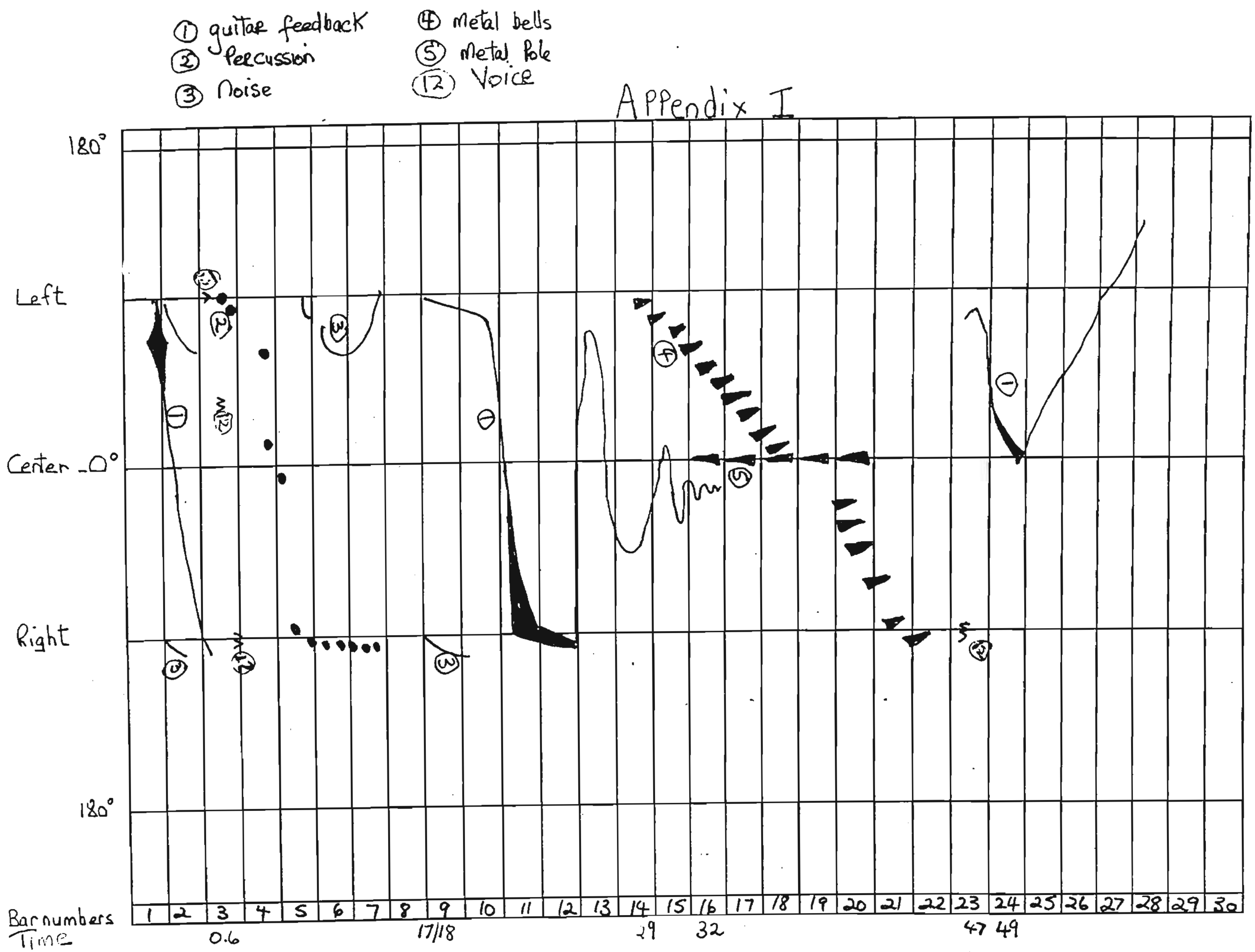
Moylan, W.1992. The art of recording. New York: Van Nostrand reinhold.

Roads, C.1996. The computer music tutorial . London: The MIT Press.

Strange, A.1983. Electronic music. Dubuque: Brown.

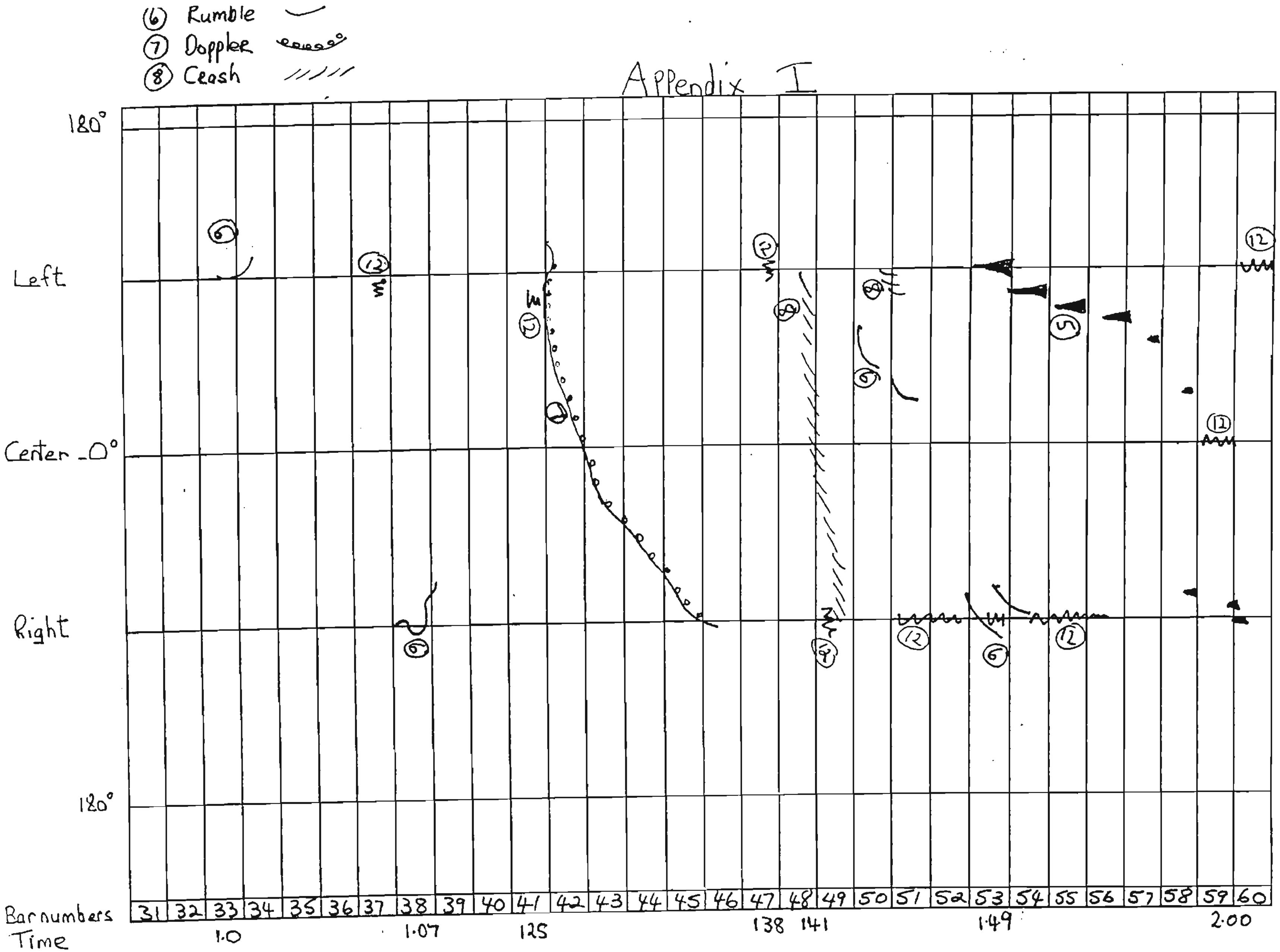
Vermooten, D.1980. Sound and acoustics. Johannesburg: n.p.

Appendix I



① guitar feedback
② Percussion
③ Noise
④ metal bells
⑤ Metal Pole
⑫ Voice
180°
Left
Center _0°
Right
180°
Bar numbers
Time
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30
0.6
17/18
29
32
47
49

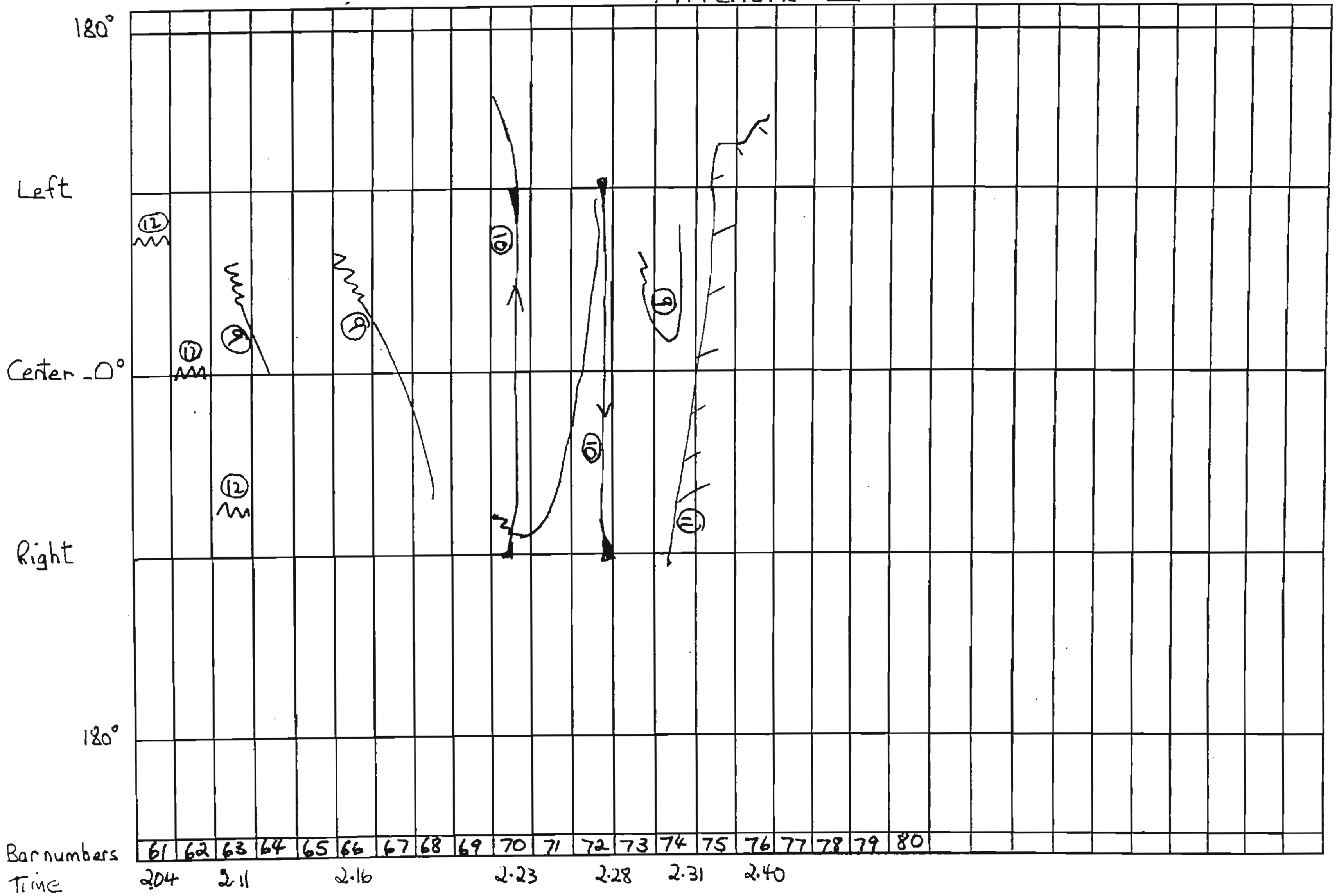
Appendix I



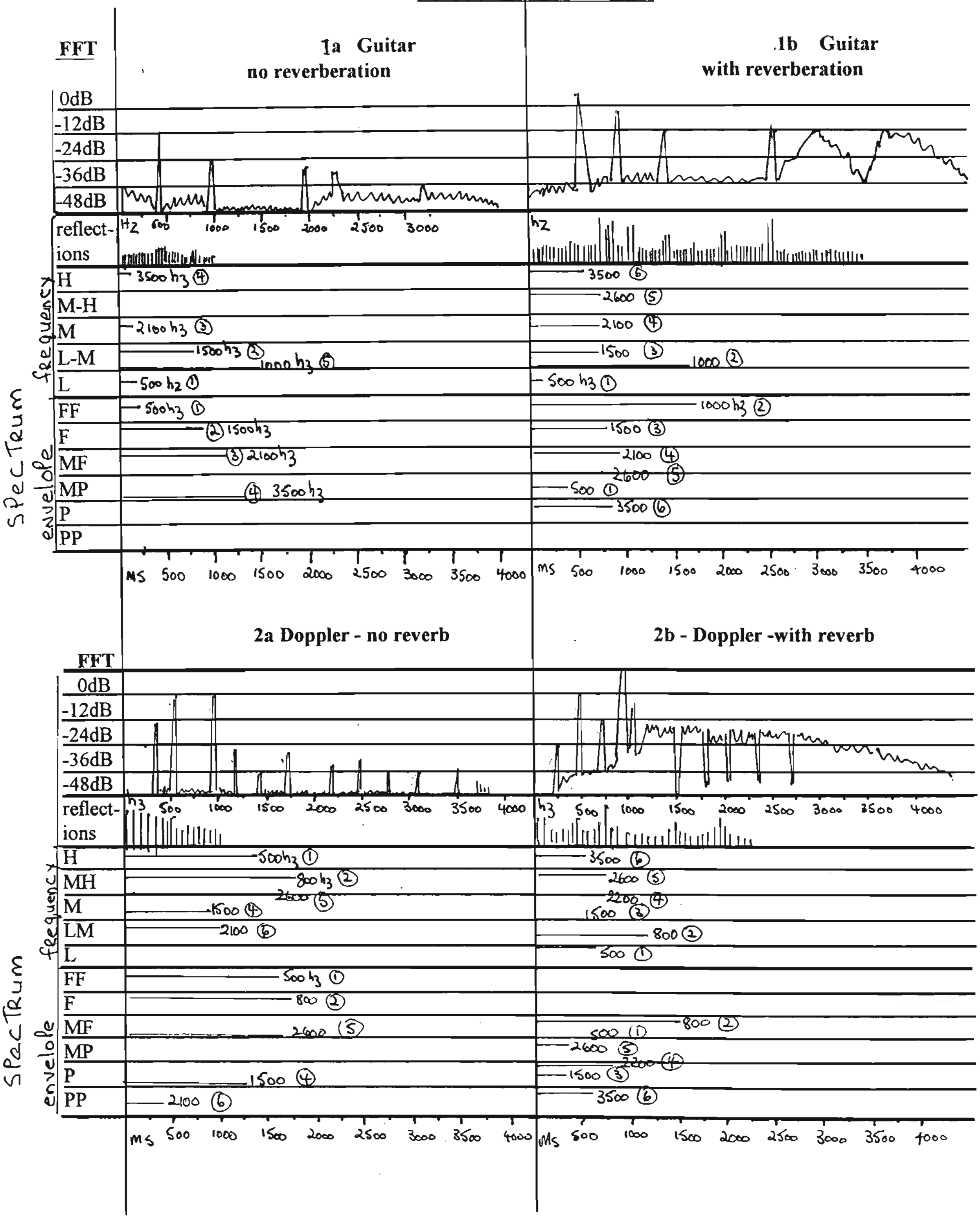
⑥ Rumble
⑦ Doppler
⑧ Crash
180°
Left
Center _0°
Right
180°
Barnumbers
Time
31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
1.0
1.07
125
138
141
1.49
2.00
6
12
7
8
5

(9) Metallic Rasp
(10) Heavy whistle
(11) Rising Crash.
(12) Voice

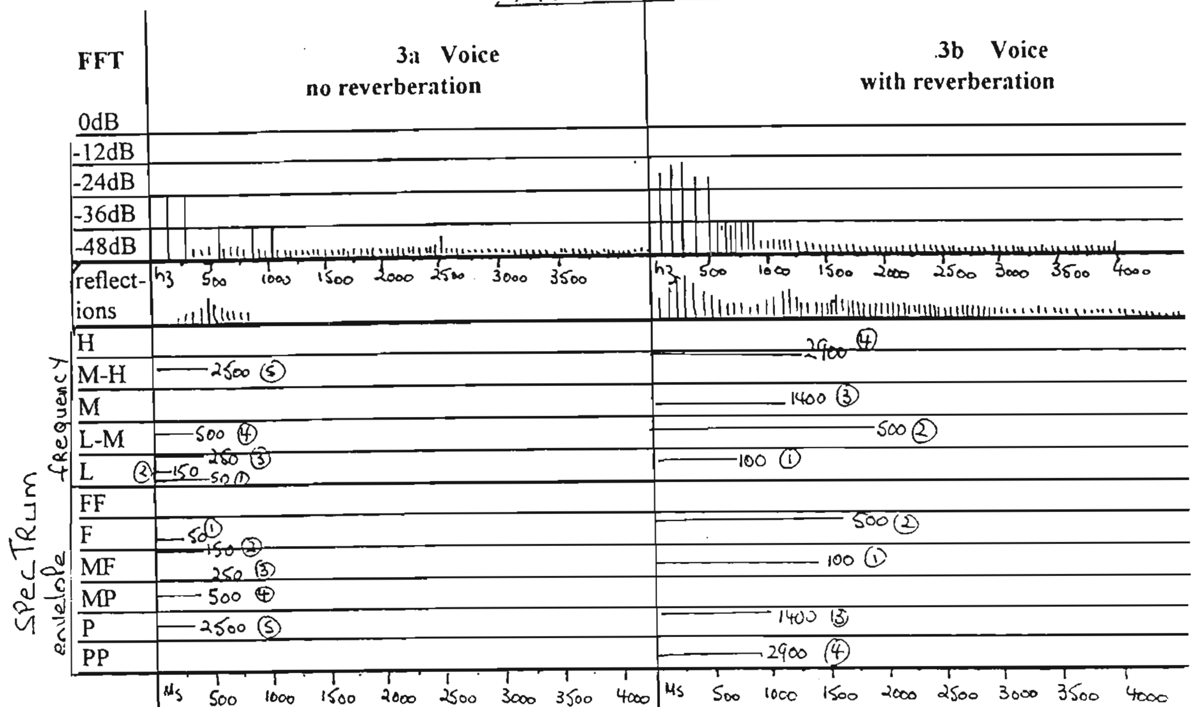
Appendix I

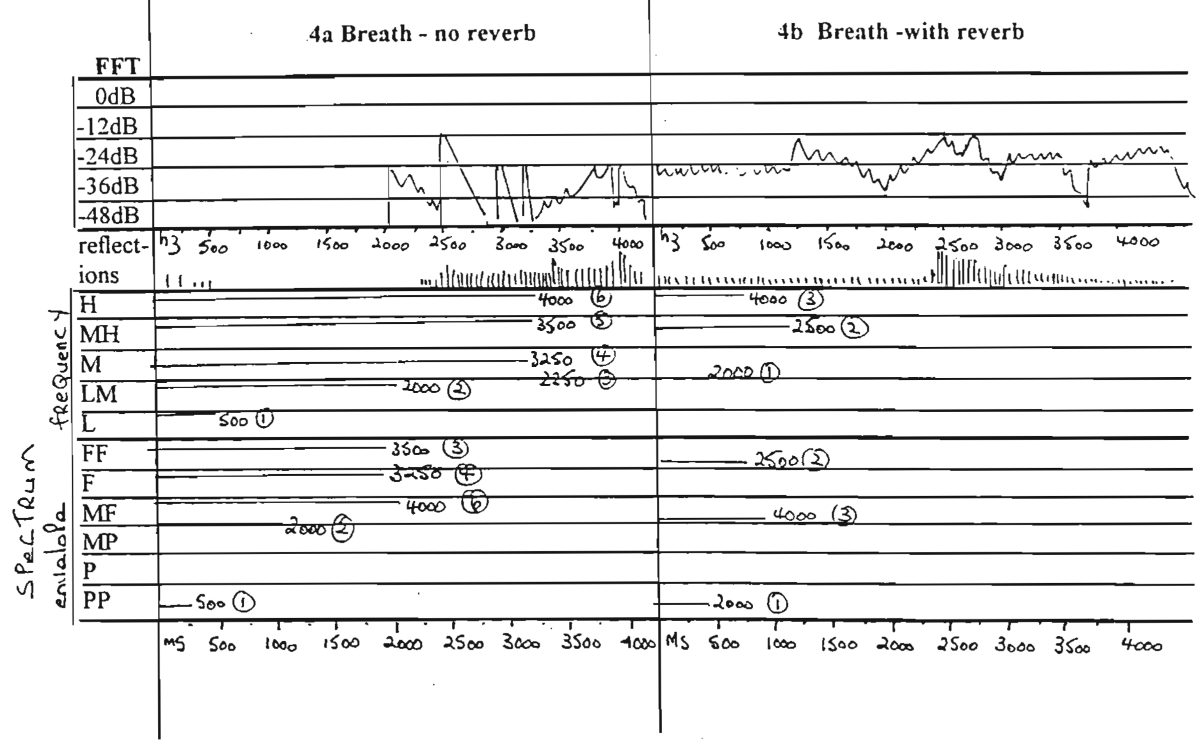


Appendix 2

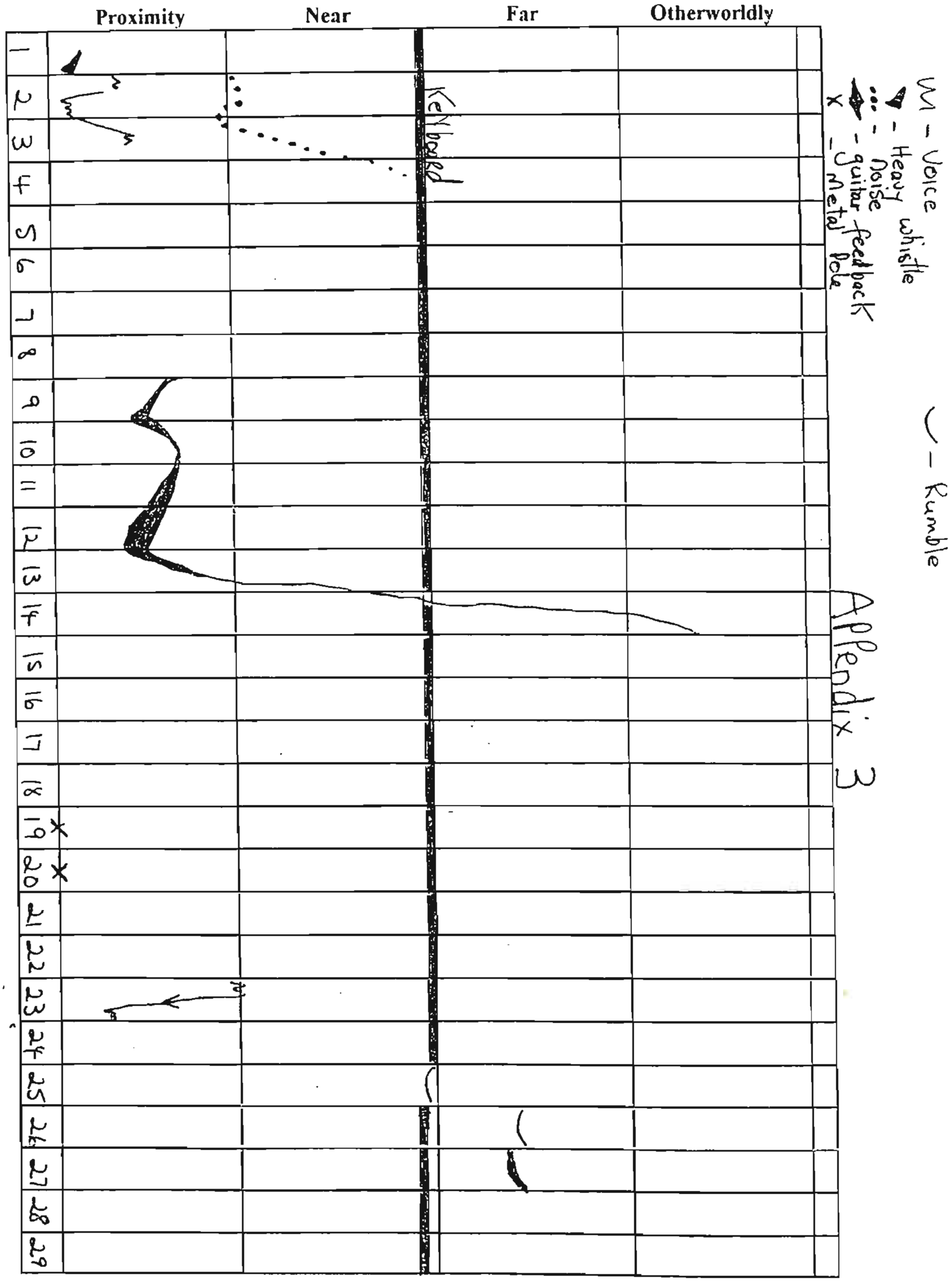

FFT
1a Guitar
no reverberation
1b Guitar
with reverberation
0dB
-12dB
-24dB
-36dB
-48dB
reflect-
ions
H
M-H
M
L-M
L
FF
F
MF
MP
P
PP
frequency
envelope
SPecTrum
2a Doppler - no reverb
2b - Doppler -with reverb
MH
LM

Appendix 2


FFT
3a Voice
no reverberation
3b Voice
with reverberation
0dB
-12dB
-24dB
-36dB
-48dB
reflect-
ions
H
M-H
M
L-M
L
FF
F
MF
MP
P
PP
Spectrum
frequency
envelope
2500 (5)
500 (4)
250 (3)
150
50 (1)
50 (1)
150 (2)
250 (3)
500 (4)
2500 (5)
2900 (4)
1400 (3)
500 (2)
100 (1)
500 (2)
100 (1)
1400 (3)
2900 (4)
MS 500 1000 1500 2000 2500 3000 3500 4000

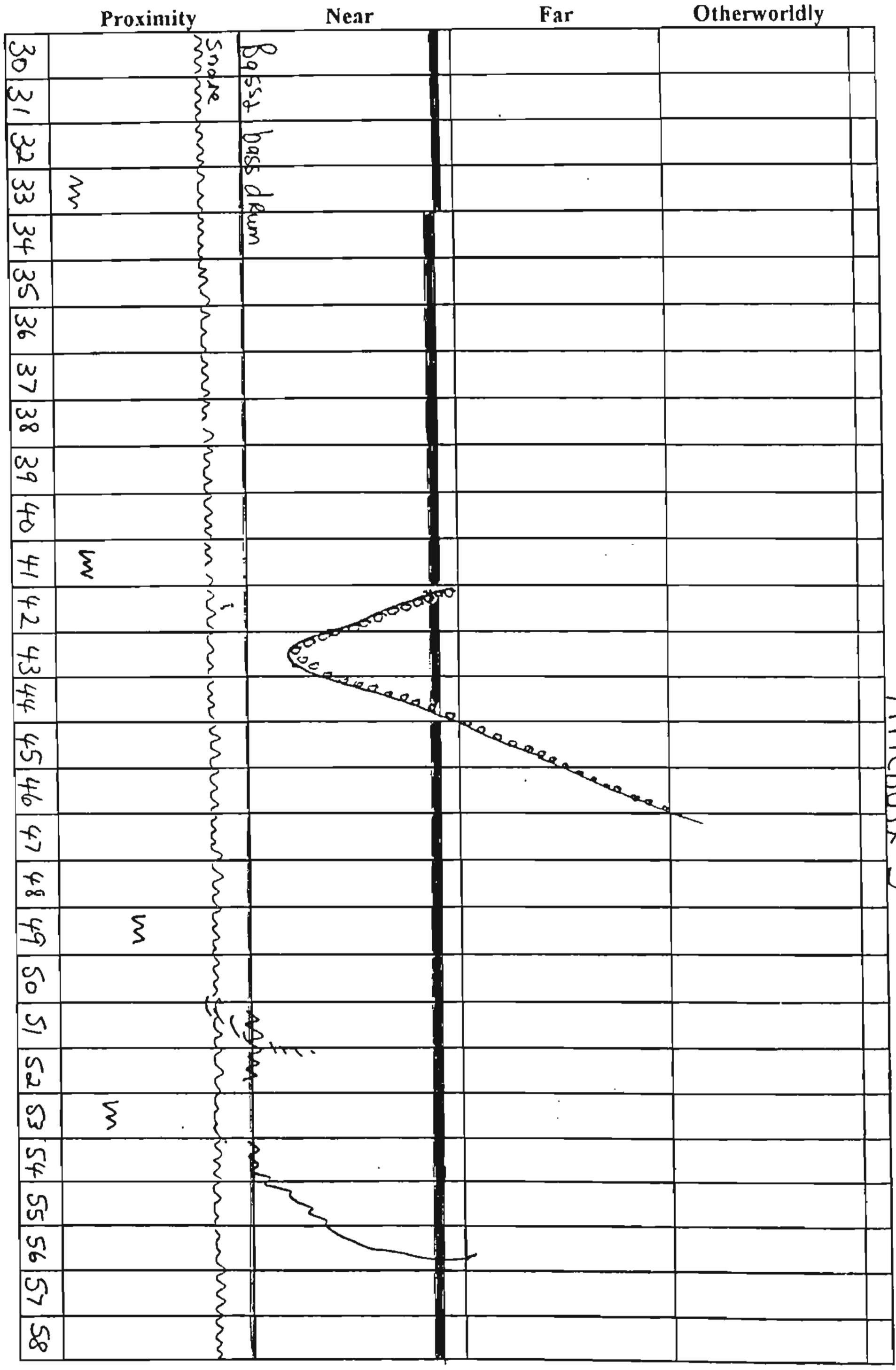

4a Breath - no reverb
4b Breath -with reverb
FFT
0dB
-12dB
-24dB
-36dB
-48dB
reflect-
ions
H
MH
M
LM
L
FF
F
MF
MP
P
PP
Spectrum
frequency
envelope
4000 (6)
3500 (5)
3250 (4)
2250 (3)
2000 (2)
500 (1)
3500 (3)
3250 (4)
4000 (6)
2000 (2)
500 (1)
4000 (3)
2500 (2)
2000 (1)
2500 (2)
4000 (3)
2000 (1)
MS 500 1000 1500 2000 2500 3000 3500 4000

Appendix 3


Proximity
Near
Far
Otherworldly
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
Keyboard
M - Voice
- Heavy whistle
- Noise
- guitar feedback
X - Metal Pole
- Rumble

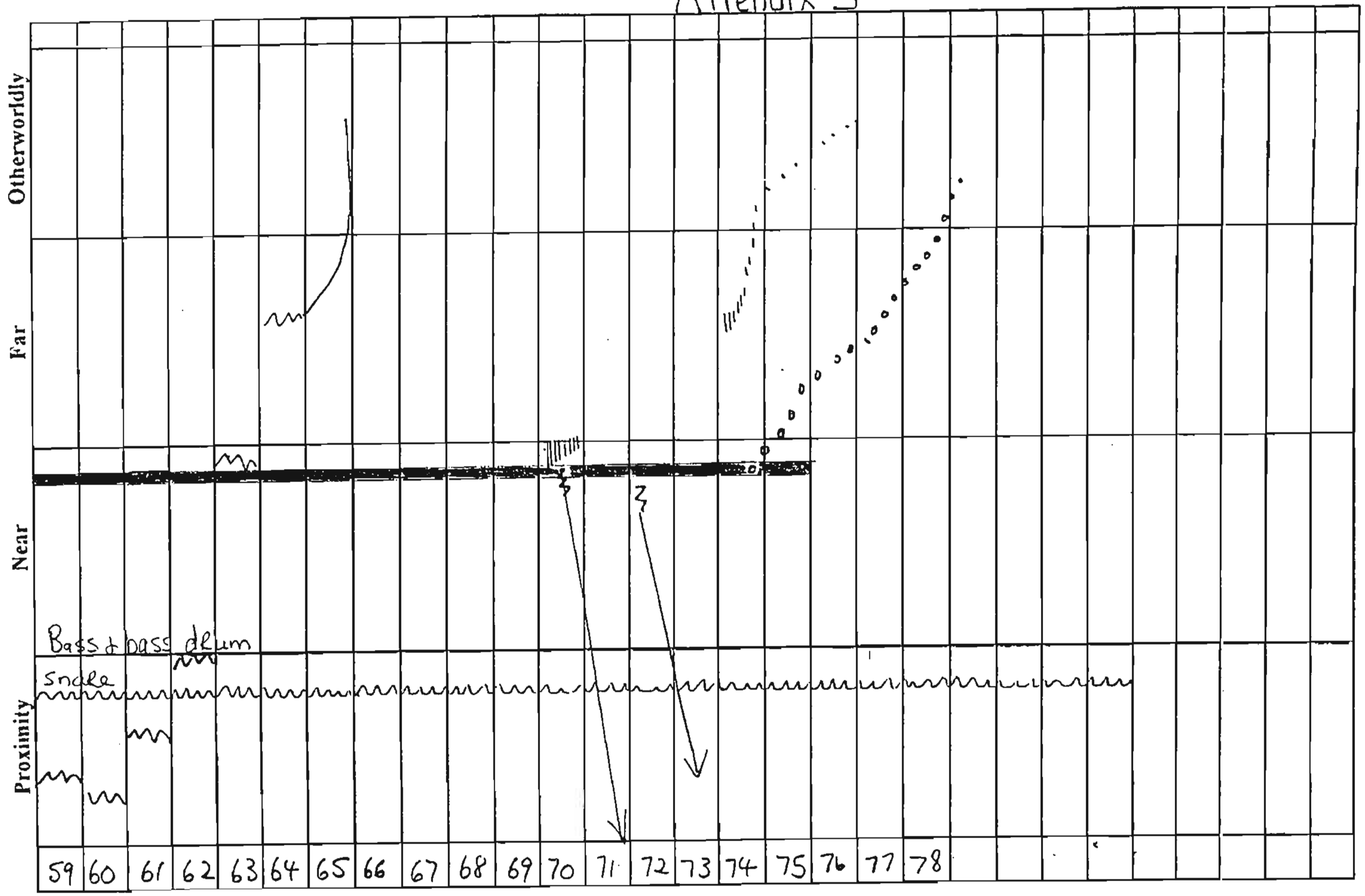
Appendix 3

- Doppler effect



|||| - metal rasp ·
m→ - Heavy whistle
ooo - Rising clash

Appendix 3



PROJECT

Valley of wet bones

Aim: To produce spatial localization using natural environments through the binaural recording process, relayed through two speakers.

Methodology:

Binaural recordings were made using a dummy-head manikin with omni-directional microphones placed inside the ear canals where the eardrums would be in a person. The microphones were connected to a mixer and panned to the extreme left and right. This was then recorded onto a dat tape machine.

Three different environments were used:

Outside environment
Studio environment
A room with dry acoustics.

All sound effects were recorded using the binaural process except for some moving sword effects, imitation fly buzzes, squeaks and marching effects used from commercial sound effects recordings. These can be heard on cuts 8,15 and 19 from the CD recording of this project.

The imitation buzzing fly sounds at 0.13 (also heard separately on cut 6), the approaching flying saucer sound at 1.14, the swishing and squeaking sounds at 2.00 (heard separately on cut 19) and the marching effect at 3.15 (heard on cut 8) were relayed over six speakers in a studio environment surrounding the dummy-head and recorded onto tape.

All other sounds were recorded naturally via the binaural process using neither panning nor equalization techniques, and their unique environments were kept without using artificial reverberation in any way. These were then recorded onto a computer based 16-track recording system, manipulated and placed at different positions in order to complete the sound picture.

The context of the theme is a valley of human bones which are the remains of a defeated army. This is followed by a flying saucer effect representing God's spirit coming to resurrect this huge army of bones. The various popping, clicking and squeaking effects represent the bones coming together. This is then followed by the army rising up and marching off into the distance.

Valley of wet bones:

Track 1(4.10)

Major localization cues: These were determined through listening on Spendor SA 200 monitor speakers, which have a reverberant and spacious sound quality. The standard stereo listening setup should be used in order to get the full benefit of this work.

0.08 - Faint rainbird harmonic sound behind head.
0.13 - Single mosquito sound close to left ear.
0.17 - Bird chirps above head.
0.27 - High pitched mosquito sound on the right, close to the head.
0.35 - Bee buzzes in circular motion- to the left front. Higher pitched bee sound gives sensation of moving around head.
0.45 - Sunbird chirps above head.
1.00 - Strong chirp above from left and then outwards towards the back.
1.14 - Loerie call on the right and slight echo towards the back.
1.52 - Birds overhead.
2.00 - Finger clicks in front, move towards forehead.
2.27 - Above on the left , low squeak.
2.45 - Creak directly above head, and illusion of coming from behind.
2.53 - Drill from above on the right, comes down to ear level on the left, front.
2.55 - Plastic bottle in front to the right at between 70-80 degrees elevation.
3.00 - Sawing in front at elevation angle between 70-80 degrees.
3.13 - Loerie front left and behind.
3.53 - Birds chirping above and behind to the right.
4.00 - Footsteps fade out as in real life situation when all has passed by.

Isolated Extracts (Tracks 2-19)

Track 2 - Outdoors effects with small bird sounds overhead and Loerie call with illusion of a 45 degree elevation to the front with call also giving the effect of coming from an angle of about 190 degrees to the left.

Track 3 - Outdoor environment with small high frequency bird chirps giving the illusion of sounding directly above the head at an elevation angle of about 70-90 degrees.

Track 4 - Pneumatic drill from the right and fading out to the left.

Track 5 - Sawing in front, with elevation of about 45 degrees.

Track 6 - Effect of flies' buzzing in a circle to the front with a single mosquito to the right-front and above. The effect of flies' buzzing in front and moving slightly to the left and up by about 10 degrees was obtained by relaying the sounds with delayed times between front, middle and back speakers, in a six speaker set-up surrounding the dummy-head. The intention was to get the effect to move around the head. This was difficult to achieve without at least four speakers. The single mosquito sound was intended to appear as close to head as possible. This was recorded by playing the effect back through headphones via the microphones in the ear canals of the dummy-head, giving it a inside-the-head effect.

Track 7 - Plastic coke bottle being crushed and blown at close proximity to the dummy-head, at elevation of 45 degrees. With a fairly dead acoustic background the effect of being very close to the head was achieved when listening back through speakers.

Track 8 - Marching soldiers: also relayed over six speakers, beginning in front, middle then to the back. The intention was to get the effect of soldiers approaching, surrounding the head, and then moving off into the distance.

Track 9 -Bee buzzing in circular motion. This effect was copied and repeated a number of times to emphasise the effect of moving around in a circular pattern. Following the bee effect is an imitation effect of a single mosquito buzzing close to the head and ear.

Track 10 -Forest environment with sunbird flying overhead and outwards towards the back.

Track 11 -Sunbird flying above and past the head, with other bee and bird effects.

Track 12 -Close-up plastic bottle squeaks, in a fairly dead environment.

Track 13 -Forest effects with sunbird flying overhead and to the left.

Track 14 - Effect of elastic band being shot at a wall in a enclosed environment to the left in front of the head.

Track 15 - Sword swishes and cork being screwed out of a bottle- neck.

Track 16 -Bee buzzing with plane and birds flying overhead.

Track 17 - Dog barking in the far distance to the left front of dummy-head.

Track 18 -Digital theme music played on keyboard.

Track 19 -Comical effect taken from a commercial sound effect recording, of sword swishes, creaking sounds and buzzes. This was also relayed over the six speakers surrounding the dummy-head giving the illusion of sounds moving back and forth.

Conclusion

In recording 3D sound, the sound material must contain dimensional information derived from the real or synthesised interaction of that sound with a human head and ears.
The perception of 3-dimensional hearing and the recording of binaural sound on two stereo channels are intimately related, it is a combination of psychoacoustics and physics. This is

particularly relevant when using the outdoors environment where the brain is accustomed to hearing sounds from all angles. The moving and static sounds, in this case mainly bird sounds, provide movement, whilst the acoustic environment provides the depth and spacial cues. The varying frequencies of the sounds whether moving above, towards or away, are recorded by the microphones which are separated from each other as the ears are in natural hearing, by the shape of the head. The shape and folds of the ears also assist in determining these cues by acting as filters to different frequencies. These cues are then interpreted by the brain due to past associations and experiences as being in that particular known environment.

In the case of the effects of cuts 6,8 and 19, the attempt to simulate movement over 6 speakers was not as effective as found in the outdoors environment. Although it is effective to some degree the illusion associated with outdoors environment is not captured in studio environments unless artificial reverberation is used to simulate this environment. The only other way to simulate this movement would be to multitrack and relay the audio through more than two speakers.

The dryer acoustics of an indoors environment gives sound a closer perspective as is the case with some of the effects in this work, although other characteristics such as timbre and frequency also aid in producing the desired effect. The crunching bottle effect on track 12 sounds closer than the sawing effect of track 5.

In general, the use of the binaural recording process does enhance the illusion of 3D. Binaural recording only seems to effectively work in live situations, where natural acoustics and spatial cues are captured, especially through live 3D sound effects.

It is difficult to capture the same dimension from recorded music without the help of artificial processing and multi-speakers setups. One of the reasons for this being that our brain does not associate music as coming from all around us but mainly from the front.

Interest in 3D sound processing is on the rise. There is renewed interest in binaural recordings with some claiming to produce true 3D sound, but until listening environments and speakers can match the acoustics and spacial cues of a particular environment, multi speakers system will still be necessary.