AN OVERVIEW OF 5.1 SURROUND SOUND WITHIN THE ELECTRONIC DANCE MUSIC CONTEXT

ANÉL SPIES

THESIS PRESENTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF PHILOSOPHY IN MUSIC TECHNOLOGY IN THE FACULTY OF ARTS, UNIVERSITY OF STELLENBOSCH.

Stellenbosch
September 2006

Supervisors:
Mr Theo Herbst
Dr Johan Vermeulen

STATEMENT

I, THE UNDERSIGNED, HEREBY DECLARE THAT THE WORK CONTAINED IN THIS THESIS IS MY OWN ORIGINAL WORK AND THAT I HAVE NOT PREVIOUSLY SUBMITTED IT AT ANY UNIVERSITY FOR A DEGREE, WHETHER IN PART OR IN ITS ENTIRETY.

30/11/2006
Date

ABSTRACT

This dissertation examines aspects around the 5.1 surround sound approach to mixing music. Although the use of surround sound systems has become thoroughly pervasive in numerous spheres in modern society, specifically in the context of home theatre systems, the present dissertation focuses mainly on 5.1 mixing within the context of electronic dance music (EDM). This focus was decided upon because EDM is the field in which the researcher is currently active.

After an examination of the physiological and cognitive aspects of the human auditory system, with specific emphasis on sound localisation, as context for discussion of 5.1 surround, an overview is given of currently available documentation providing specifications for the implementation of 5.1 surround. This is then related specifically to questions regarding mixing in the context of 5.1 surround and incorporates a discussion the views of producers currently active in the EDM industry.

The ultimate aim of the abovementioned is to establish the extent, or lack thereof, to which 5.1 surround is currently being implemented in the field of EDM. In response to this, the implementation of 5.1 in EDM is illustrated through practical application of 5.1 surround mixing in original music produced by the researcher and accompanying the present dissertation.

OPSOMMING

Hierdie dissertasie ondersoek aspekte rondom die 5.1 surround sound benadering tot die klankmengwerk van musiek. Ondanks die feitlik alomteenwoordige gebruik van ruimtelike klanksisteme in die moderne samelewing, veral in die vorm van huishoudelike teatersisteme, word hier veral gefokus op 5.1 mengwerk binne die konteks van elektroniese dansmusiek (EDM). Daar is op hierdie fokus besluit op grond daarvan dat dit die gebied is waarop die navorser tans aktief betrokke is.

Na die ondersoek van die fisiologiese en kognitiewe aspekte van die menslike gehoorsisteem as basis vir die bespreking van 5.1, met spesifieke verwysing na klanklokalisering, word 'n oorsig gebied oor die dokumentasie wat tegniese spesifikasies bevat ten opsigte van die implimentering van 5.1. Laasgenoemde word vervolgens in verband gebring met klankmenging binne die konteks van 5.1 en word uitgebrei deur verwysing na sienings van enkele vervaardigers wat tans in die elektroniese dansmusiekindustrie werksaam is.

Die uiteindelike doel van bogenoemde is om te bepaal tot watter mate, hetsy in groter of kleiner omvang, 5.1 tans in EDM aangewend word. Ter uitbouing hiervan word die aanwending van 5.1 in EDM in 'n praktiese afdeling geïllustreer in die vorm van oorspronklike musiek wat deur die navorser gekomponeer is en by die dissertasie ingesluit word.

DEDICATED

TO

THE LORD GOD MY SAVIOUR

AND TO

THE PEOPLE GOD USED TO SHOW ME

THE TRUTH, THE WAY AND THE LIGHT

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1. Motivation for this study

5.1 Surround Sound expands the sound field – and therefore the listening experience – beyond the arc of 60° and produces a more heightened envelopment and a more steady localization of sound sources. Albeit in its early phase (with regards to 5.1 mixing), the future and related potential of 5.1 Surround Sound is promising, although loaded with several technical and aesthetic questions and challenges. This concept of *5.1 Surround Sound* is defined as: The front centre channel is at equal distances between the left and right front loudspeakers. The extension of the front arc into a full horizontal 360° is provided by the left and right surround speakers.[2]

The motivation for this study revolves around three factors, namely (1) personal interest, (2) media consumerism, and (3) technical and psychoacoustic research.

Firstly, the author's personal interest in *5.1 surround sound* is based on the potential platform it creates for composing and arranging within the 5.1 domain and this within the music genre chosen for this dissertation, namely electronic dance music (EDM).

Secondly, in the commercial market, it is a well established fact that the term "surround" and all that it implies has established itself and gained a secure foothold in the home, studio and EDM genre (Swenson 2002: online). Even with the increasing confusion over the future of surround music mixes there is an increasing demand for efficiently-produced surround products for film

[2] Based on the ITU-R Recommendation BS.775-1 (1994). Please note that this recommendation was replaced by the ITU-R Recommendation BS.775-2 (2006) during the course of this study.

and television, as well as for music-only releases (AES Workshop [Abstract] 2006: online). The wide implementation of *5.1 surround sound* includes motor vehicle, conferencing, home entertainment, communication, and entertainment / dance / concert venues. Furthermore, we see the ability of television broadcast networks, satellite and cable operators, and terrestrial affiliates to deliver discrete digital 5.1 audio to significant numbers of households, straight into their home theatres. The latter prospect demands the final mixing of these productions to be in 5.1 (Bunish 2003: on-line).

Thirdly, the increase in the consumerism motivates the development of necessary algorithms and technical infrastructure by relevant organisations, e.g. Dolby Laboratories and the Fraunhofer Institute. Furthermore, this research project may lead to the publication of technical documentation regarding specifications on 5.1 implementation. This citation, however, has not been fully related to [homogeneous] standardisation practices [yet] and may therefore lead to confusion for the ill-informed producer/consumer. Furthermore, extensive research in psychoacoustic principles – to acquire a better understanding of surround sound – is being applied in the industry. For example, Dolby Digital Programming uses a new algorithm based on Auditory Scene Analysis in their *Dolby Model 585* time-scaling processor for multi-channel audio (Dolby News, Professional Audio Edition 2004: on-line).

1.2. Purpose of the study

The present research project consists of a close examination of the fundamental aspects of *5.1 Surround Sound* and its application within the EDM scene. Discussions of some of these fundamentals are illuminated by a demonstration in the concluding practical chapter of this thesis. Particular attention is paid, in this regard, to the audio mixing aspects of 5.1. The quantification of music – by applying certain rules – can inhibit the creative process of the composer. Therefore, the practical component of this dissertation shall be of a subjective nature, although essential *5.1 Surround Sound* specifications apply. The researcher interprets *5.1 Surround Sound* as a clean slate or canvas facilitating a mixing process where the opportunity exists to experiment with sound placement within a 360° sound environment. Because of this gain in space, the composer can also apply effects such as compression and EQ more creatively.

1.3. Sources

This dissertation examines surround sound from a technical as well as a creative point of view. In this regard, a range of technical documentation will be covered in the academic component of this dissertation. Technical specifications of surround sound will make use of standards that are provided by selected organizations in the industry, including the International Telecommunication Union (ITU) and the European Broadcasting Union (EBU). A number of constraining factors to this dissertation exist. These include limited academic publication regarding the practicalities of *5.1 Surround Sound*, especially mixing aspects and the fact that *5.1 Surround Sound* has not yet found a secure footing in EDM surroundings. The impact of this fact translated into the researcher relying on [severely] subjective interviews with producers and DJs.

1.4. Research methodology

This project was executed over a time period of two years. Preliminary research consisted of creative work in the recording studio in order to become familiarised with basic recording, programming and production aspects of EDM. Because of the importance of a thorough knowledge about Pro Tools HD and the use of a MIDI, a number of creative projects were completed in this time. These include an audio-visual DVD production, classical recordings, popular music recordings as well as EDM productions (see Appendix). Parallel to this, an extensive literature study was performed in order to keep up with technological progression. The literature study was followed by the implementation of a *5.1 Surround Sound* project in the specific music genre (i.e. EDM) selected for the project. This dissertation was restricted to 5.1 due to available infrastructure, including loudspeakers in the chosen facility.

1.5. Specific problems encountered

The implementation and interaction between recording (and playback) systems and spatial sound information, is not covered by a homogeneous body of publications emanating from a single, governing body (with related standardization). This may potentially lead to confusion in the application of these specifications.[3]

Furthermore, research covering the interaction between visual media and sound is vast and too extensive for the purposes of a dissertation of this magnitude. It was therefore decided to concentrate on the investigation of *5.1 Surround Sound* from an auditory viewpoint.

The budget and available infrastructure had vast limitations on the practical aspect – actual *5.1 Surround Sound* implementation – of the research project. In addition, it can be stated that the local EDM market is very small and the genre has not yet produced DJ's and Producers that are internationally ranked. The implications for this are also quite obvious.

1.6. Structure and scope of the study

Chapter 1 provides the purpose and motivation for this dissertation and outlines the sources, methods and problems relevant to this study. **Chapter 2** provides and overview of physiological and cognitive principles, to obtain a deeper understanding of sound. The purpose of this chapter is to demonstrate how human beings process sound by means of the auditory hearing mechanism – how sound waves are transformed into electrical impulses and whereby it is further interpreted by the human mind. **Chapter 3** discusses basic concepts within the field of acoustics. Following that, the processes involved in the human localisation of sound sources, are discussed. **Chapter 4** describes the technical specifications, provided by standardisation organisations regarding the implementation of 5.1 surround sound. Amongst others, these include room design, speaker placement, sound level calibration and bass management. **Chapter 5** focuses specifically on 5.1 mixing within the EDM genre. The application and

[3] According to Rumsey (2001: 128-9) the three important organisations are the International Telecommunication Union (ITU), European Broadcasting Union (EBU), and the Society of Motion Picture Television Engineers [USA] (SMPTE).

functioning of basic signal processors are discussed as well as methods in 5.1 mixing. Music cannot merely be quantified and therefore **Chapter 6** documents the creative component of the study and provides a discussion on 5.1 mixing in EDM, performed by the researcher. The conclusion of the study is discussed in **Chapter 7**, followed by the References section and Appendix.

CHAPTER 2

PHYSIOLOGICAL AND COGNITIVE

2.1. The anatomy and physiology of the ear

The following discussion of the anatomy and physiology of the human ear is based primarily on a discussion of the present topic in *An Introduction to the psychology of Hearing* by Moore (1998). Secondary sources include the following: Blauert (1999); Bruce (1997); Butler (1992); Correira (2002); Durant *et al* (1995); Ferl *et al* (1996); Grey (1918); Middlebrooks (1992); Munkstedt (2006); ProAV (2006); Pickles (1982); Raichel (2000); Sound Retrieval Systems [SRS] (1998); Von Békésy (1960) and Watkinson (1999). Cases where the latter sources are utilised, are indicated by means of references. It should be noted that differences occur in the Latin terminology used in the various sources. For the sake of consistency, a decision was taken to rely on Latin terminology provided in Moore (1998) and Blauert (1999).

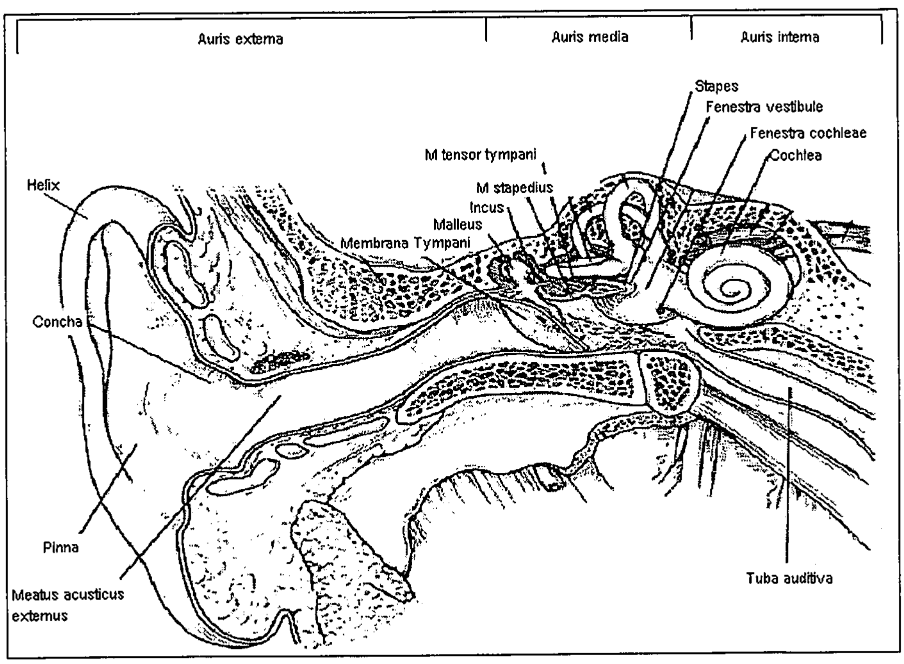


Figure 1: Anatomy of the ear (Pickles 1982:11)

Captions adapted from Moore (1998) and Blauert (1999).

In Figure 1 the three main sections of the human hearing mechanism can be identified. These are as follows: (1) the *auris externa* (outer ear), which is responsible for sound reception, -amplification, -localization and protection of the other sections of the ear; (2) the *auris media* (middle ear), which is responsible for impedance adjustment; and (3) the *auris interna* (inner ear), which is responsible for the conversion of sound.

The anatomical and physiological structure of the ear can be illustrated as follows:

2.1.1 The auris externa and the auris media

The auris externa consists of the *pinna* and the *meatus acusticus externus*, that is, the external ear canal. The *pinna*, attached to the outer extremity of the ear canal, stands at an angle of between 25° and 45° to the surface of the head (Blauert 1999: 53). The deep notch in the pinna is known as the *concha*, while the ridge at the outer edge of the pinna is called the *helix* (Grey 1918: 1033). Although not formerly realised, it has been proven[4] that the pinna positively influences the spatial perception of sound. From an acoustical perspective, the pinna functions as a linear filter of which the transfer function is dependant on the direction and distance of the sound source. This allows for the coding of spatial attributes of the sound field into temporal and spectral attributes.

The meatus acusticus externus extends from the concha to the *membrana tympani*. The full length of the meatus acusticus externus is approximately 2 cm, giving it a resonant frequency[5] of around 3400 Hz – incidentally also an important frequency for human speech perception. The sound pressure difference between the outer opening of the meatus acusticus externus and the membrana tympani is 10 dB (Raichel 2000: 207).

---

[4] Refer to the experiment by J.C. Middlebrooks (1992: 2607-2624) with the title: "*Narrow-band sound localization related to the external ear*".

[5] Resonance can be defined as "*...the tendency of a mechanical or electrical system to vibrate at a certain frequency when excited by an external force, and to keep vibrating after the excitation is removed*" (White 1991: 282)

The membrana tympani is an elliptical membrane acting as a partition between the meatus acusticus externus and the auris media. Positioned at an angle of 40° to 50° to the meatus acusticus externus, movement of the former takes place when undulations of air pressure occur within the meatus acusticus externus.

When acoustical energy reaches the ear, the pinna directs the sound towards the meatus acusticus externus, which in turn sets the membrana tympani in motion by means of the conversion of acoustical energy to mechanical energy. To a lesser extent, sound is also conveyed to the meatus acusticus externus via the temporal bone, but this is of secondary importance as far as spatial hearing is concerned.

In the auris media, a chain of three ossicles – the *malleus*, *incus* and *stapes* – conduct the vibrations caused by the membrana tympani through the auris media. This chain stretches from the membrana tympani to the *fenestra vestibule* (oval window), ensuring the transfer of sound waves to the fenestra vestibule within the *cochlea* situated in the auris interna (Ferl *et al* 1996: 732). Another important component of the auris media is the *tuba auditiva* (Eustachian tube)[6] (Raichel 2000: 207) which is normally closed, but opens during the action of yawning or swallowing in order to equalise the static air pressure on both sides of the membrana tympani. The *cavum tympani*, which is the cavity within the auris media, perform a further primary function by ensuring the effective transfer of sound from the air to the fluids in the cochlea. Incoming sound waves are largely reflected from the fenestra vestibule, because the amount of resistance it offers to movement differs from that of air. This is due to a difference in acoustical impedance.[7] The cavum tympani, in turn, acts as an impedance-matching device or transformer (Moore 1989: 17) that improves sound transmission and reduces the amount of reflected sound.[8]

It should be noted that the movements of the ossicles are influenced by two muscles that contract when exposed to intense sound, namely the *M tensor tympani* and the *M stapedius*. This contraction is known as the *middle ear reflex* and reduces the transmission of sound

[6] Named after Bartolomeo Eustachi (1514-1574). He extended the knowledge of the internal ear by rediscovering and correctly describing the tube which bears his name (The Columbia Electronic Encyclopedia 2002: on-line).

[7] Defined as "Acoustic impedance is a ratio of acoustic pressure to flow" (Wolfe 2006: 1).

[8] The transmission of sound within the auris media is most effective at frequencies between 500 and 4000 Hz (Moore 1989: 17).

through the middle ear in order to protect the cochlea at low frequencies. A further two functions have been proposed for this reflex. Firstly, it reduces the audibility of self-generated sounds, particularly speech.[9] Secondly, it effects the reduction of masking of middle- and high frequencies by lower ones.

2.1.2 The auris interna

As far as hearing is concerned, the cochlea is the principal component of the auris interna (Moore 1989: 17). The cochlea divides into three fluid-filled scalae along its length, the *scala vestibuli*, the *scala tympani* and the *scala media*. The two outer scalae, the scala vestibuli and the scala tympani, are joined by means of an opening found at the apex of the cochlea. This opening is called the *helicotrema*. The scala media, which forms a closed inner compartment, is on the one hand separated from the scala vestibuli by means of the *membrana vestibularis* (Reissner's membrane)[10] and on the other hand separated from the scala tympani by the *membrana basilaris* (Bruce 1997: 3).

Sound reaches the cochlea via the fenestra vestibuli resulting in an inward movement of the latter. This effects a change in pressure over the length of the membrana basilaris, which in turn results in displacement of the fluids in the cochlea, causing a wavelike movement of the membrana basilaris. This movement is directed towards a second window in the cochlea, the *fenestra cochleae* (round window), which opens into the base of the scala tympani and as a result undergoes an outward movement (Bruce 1997: 3). The movement of the membrana basilaris adopts a sinusoidal wave pattern stretching from its base to its apex. It should be noted that the membrana basilaris responds differently to various frequencies due to its mechanical properties. High-frequency sounds result in a maximum displacement of the membrana basilaris near the oval window, which means that in such instances there is little movement on the remainder of the membrana basilaris. Low frequency sounds, in contrast, result in a pattern of vibrations that extend along the entire length of the membrana basilaris, but which reaches a peak before the end of this membrane.

---

[9] It has been shown that this reflex is activated just before vocalization.
[10] This membrane is named after the German anatomist, Ernst Reissner (1824–1878). (Nsamba 1979).

Figure 2 demonstrates that the optimal displacement caused by different frequencies occurs at different points on the membrana basilaris.

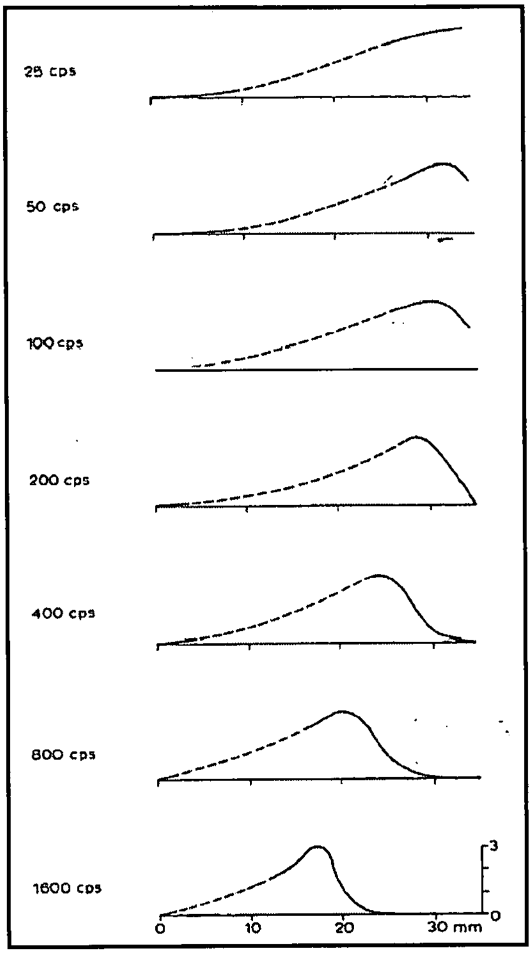


Figure 2: The shift in the place of maximum vibration amplitude along the basilar membrane for stimulation with different frequencies (Moore 1998: 19) [11]

[11] From *Experiments in Hearing*, by Von Békésy (1960), used with the permission of McGraw-Hill, by Moore.

From the above figure it can be deduced that the membrana basilaris in effect performs a Fourier analysis[12] on sounds it receives. This means that it acts as a mechanical frequency analyzer, a function which is essential in the perception and discrimination of phenomena such as pitch, timbre, consonance, dissonance as well as other auditory phenomena such as critical band, masking and the precedence effect (Watkinson 1999: 128).

The actual transduction of sound into auditory nerve signal, however, takes place in a further important component of the inner ear, namely the organ of Corti [13], which is situated inside the cochlea. The organ of Corti performs the following two important functions: Firstly, it is responsible for active filtering of the vibrations of the membrana basilaris; secondly it performs the transduction of sound energy into neural activity within the auditory nerve. Both these functions are performed by hair cells[14], which are situated within the organ of Corti (Bruce 1997: 4).

Because of its importance, the process of transduction that takes place in the organ of Corti will now be examined within the relevant scope of the present dissertation.

[12] This refers to a process in which a complex wave form is reduced to a series of sine waves with specific frequencies, amplitudes and phases (Moore 1989: 3).

[13] Named after the Italian anatomist Alfonso Giacomo Gaspare Corti (1822–1876) who discovered it (Science Clarified 2006).

[14] On account of discrepancies in the terminology used by various sources (amongst others: Henry Grey [1918]; Manning J Correira [2002]; Ferl *et al* [1996]) to describe the hair cells as well as the hairs that grow out of them, it has been decided not to make use of the Latin terminology in this instance.

2.1.3 The process of transduction and the hair cells

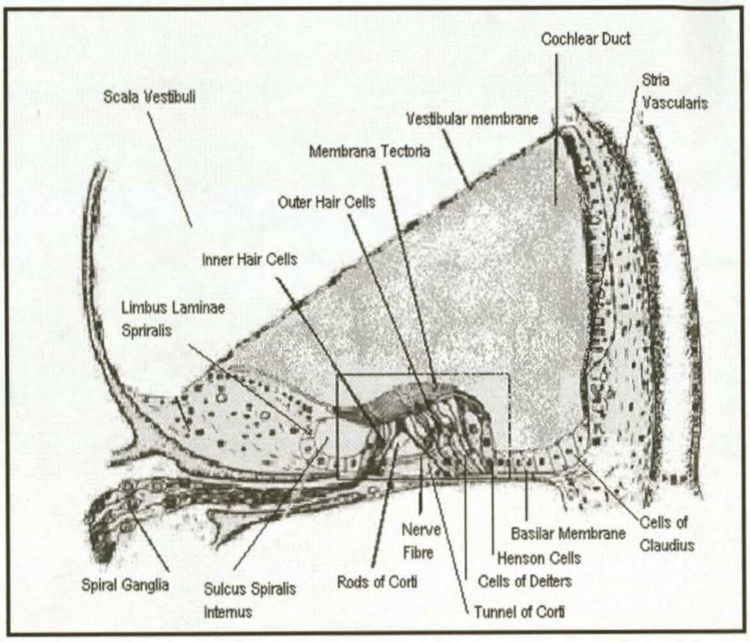


Figure 3: Cross-section of the cochlea showing the organ of Corti[15] (Moore 1989: 27)

The hair cells are divided into two groups (the inner and outer hair cells) by an arch forming what is called the tunnel of *Corti*. It should be noted that the outer hair cells are involved in the active filtering of vibrations of the membrana basilaris, while the inner hair cells are involved in the transduction of sound energy into neural activity (Bruce 1997: 4). Above the hairs cells lies a gelatinous membrane called the *membrana tectoria*. The hairs of the outer hair cells seemingly come into contact with this membrane, while it does not appear to be the case for the inner hair cells. When sound reaching the inner ear causes the membrana basilaris to move up and down, a shearing motion results between the latter and the membrana tectoria. This results in displacement of the hairs at the top of the outer hair cells and is thought to cause excitation of the inner hair cells. Excitation of the latter in turn results in the generation of action potentials

[15] The specific figure only contains the captions relevant to the present topic. The original illustration can be found in Moore (1989: 27).

in the neurones of the auditory nerve. The inner hair cells therefore are responsible for the transduction of mechanical movements into neural activity.

Watkinson (1999: 128) points to the fact that nerve firings are not perfectly analogue to the movement of the membrana basilaris. It appears that a nerve firing occurs at a constant phase relationship to the basilar vibration. This phenomenon is called *phase locking*. Firings do not necessarily occur on every cycle; it takes place irregularly in the case of higher frequencies, but they are all the same in phase relationship.[16]

2.2. The fundamental faculties of the ear

2.2.1 Sensorial characteristics of loudness

Moore (1989: 47) defines loudness as "… that attribute of auditory sensation in terms of which sounds can be ordered on a scale extending from quiet to loud."

In itself, however, the perception of loudness is a subjective reaction to sound level. Concerning loudness, Butler (1992: 78-82) notes that confusion around the terminology used can be addressed by making a distinction between the physical measurement of loudness and perceptual reactions to loudness. These concepts are interrelated to one another, though not interdependent.

Physical measurement of loudness

The most important factor that contributes to the independence of the physical measurement of loudness and perceptual reactions to loudness is the wide range of perceivable intensity values. This results in the mathematical measurements of and calculations with values involved in

[16] Although a more detailed discussion of the nature and functioning of the hair cells is not relevant to the present dissertion, more information on this subject can be found in *Spatial temporal coding of sound in the auditory nerve for cochlear implants* by Ian Christopher Bruce (1997: 7-14).

perceiving loudness, often being awkward. A number of units are used for measuring loudness: w/m^2, N/m^2, and the decibel (dB). The first two are rarely encountered even in research laboratories, hearing clinics, industry as well as in literature discussing acoustics and/or hearing. More customarily, it is the decibel that is encountered (Durant *et al* 1995: 54). A logarithmic scale is used to express values; the unit used for physical measurement of loudness being the decibel (one tenth of one Bel).

Illingworth (1998: 118) defines a decibel as "...a dimensionless unit used to express the ratio of two powers, voltages, currents, or sound intensities.[17] It is ten times the common logarithm of the power ratio." Thus if two values of power, P1 and P2 differ by n decibels then

n = 10 log10 (P2/P1)

i.e. P2/P1 = 10n/10

Thus, two powers, one of which is ten times the other, will differ by 1 bel; 10 Watts are 1 bel higher in level than 1 Watt.

If P1 and P2 are the input and output powers, respectively, of an electric network then if n is positive, that is P2>P1, there is a gain in power; if *n* is negative there is a power loss.

The dB-scale is preferred to the Bel because the latter is an inconveniently large unit.

Perceptual measurement of loudness

White (1991: 133) recounts that Fletcher and Munson undertook a series of experiments in the early 1930s relating to the sensitivity of the human hearing mechanism. One of these was to measure the sensitivity of the human hearing mechanism at different frequencies in order to establish the threshold of hearing, which is the softest audible sound. From this, they were able

[17] With regards to the representation of sound magnitude in decibels, Durant et al. (1995: 54) emphasise that this only has concrete physical meaning when an accompanying reference quantity is present, this being 10^{-5} N/m^2 for sound pressure. When measuring acoustic intensity (in dB), the result is the acoustic *intensity level* (IL); measuring sound pressure, however, the *sound pressure level* (SPL) is obtained.

to establish that the most sensitive range of human hearing is between 3 and 4 kHz. In addition, they found that sensitivity falls off rapidly at lower frequencies and somewhat more slowly at higher frequencies. They proceeded to plot these levels with respect to frequency and found that the resulting curve is not uniform, but varies drastically with frequency. Very soft sounds need to be more powerful at frequencies lower and higher than 3 to 4 kHz in order to be heard.

In a subsequent experiment, these same researchers chose a reference frequency of 1 kHz and increased the strength of the softest audible sound at that frequency ten times (10 dB). Subjects were asked to judge when other tones, generated at lower and higher frequencies and strengths, had the same loudness as the reference tone. Plotting the strengths of these tones against the threshold value, a "contour of equal loudness" was formed. This curve did however not run parallel to the "threshold curve", thus indicating that the human ear hears different frequency tones more uniformly in loudness when they are stronger than the threshold levels (White 1991: 133).

This is known as the Fletcher-Munson effect and has important implications for the reproduction of sound. Moore (1989: 52-53) points to the fact that the relative loudness of the different frequency components in a sound will change as a function of the overall level, so that unless the sounds are produced at the same level as the original, the tonal balance will be altered. The ear becomes relatively more sensitive to low frequencies at high intensities, while conversely becoming less sensitive to very low and very high frequencies at low levels. As a result, many amplifiers incorporate a loudness control[18] that boosts the bass, and to a certain extent the treble, at low listening levels.

[18] Moore emphasises, however, that such controls are of limited use since they do not take into account loudspeaker efficiency and the size of the listening room.

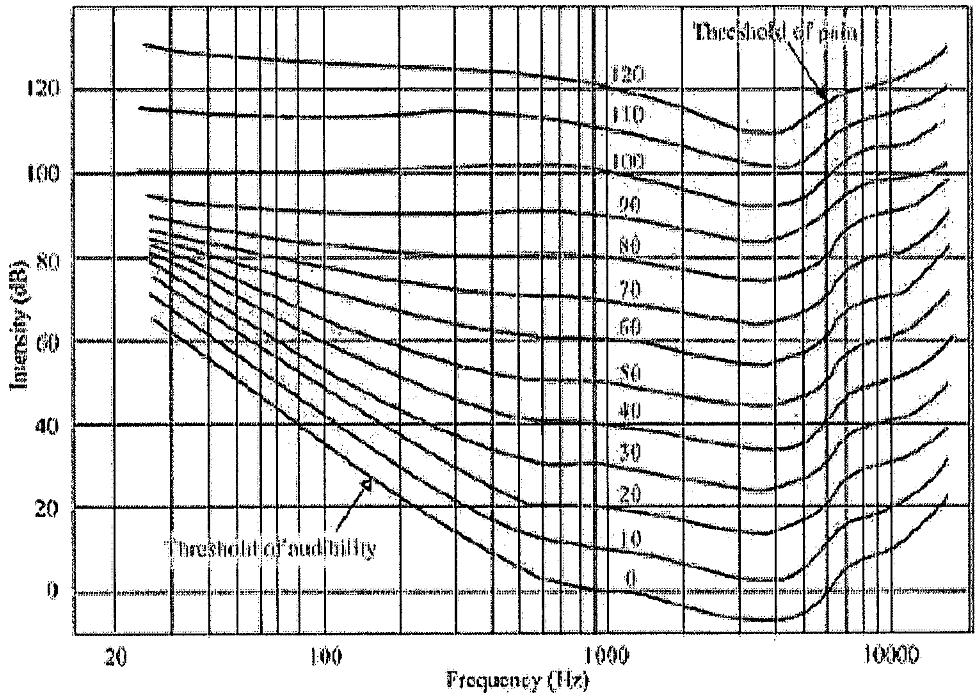


Figure 4: Fletcher-Munson equal-loudness contours (Munkstedt 2006: 5)

The levels measured by Fletcher and Munson were plotted on the dB-scale using a unit called the *phon*. This is a psychological unit of loudness and can be defined as the sound pressure level (SPL) of a 1 kHz pure tone that is judged to be the same loudness as the sound in question (White 1991: 245). Butler (1992: 81) adds that the term *phon* was created to describe loudness level, as distinct from intensity level. He states that phons and decibels are only equal at the frequency of 1 kHz, which is the frequency of the standard tone employed by Fletcher and Munson. The Fletcher-Munson Curve of Equal Loudness has been used in the design of sound level meters in an attempt to give an approximate measure of the loudness of complex sounds. The use of the phon has since been superseded by the use of weighting networks in such meters. This means that a given meter does not simply sum the intensity at all different frequencies, but rather weighs the intensity at each frequency according to the shape of the equal loudness contours. Sound levels measured using such meters are usually specified in terms of the weighting employed. A given level might be specified as 35 dBA, which means that the meter gave a reading of 35 dB when the 'A'[19] weighting was used.

In conclusion it should be emphasised that the assumption cannot be made that sound level meters necessarily give a true approximation of the loudness of a given sound. Such readings are closely related to the dB scale, which is a scale of physical magnitude rather than a scale of subjective sensation. Nonetheless, such meters do make it possible to roughly compare the loudness of different complex sounds.

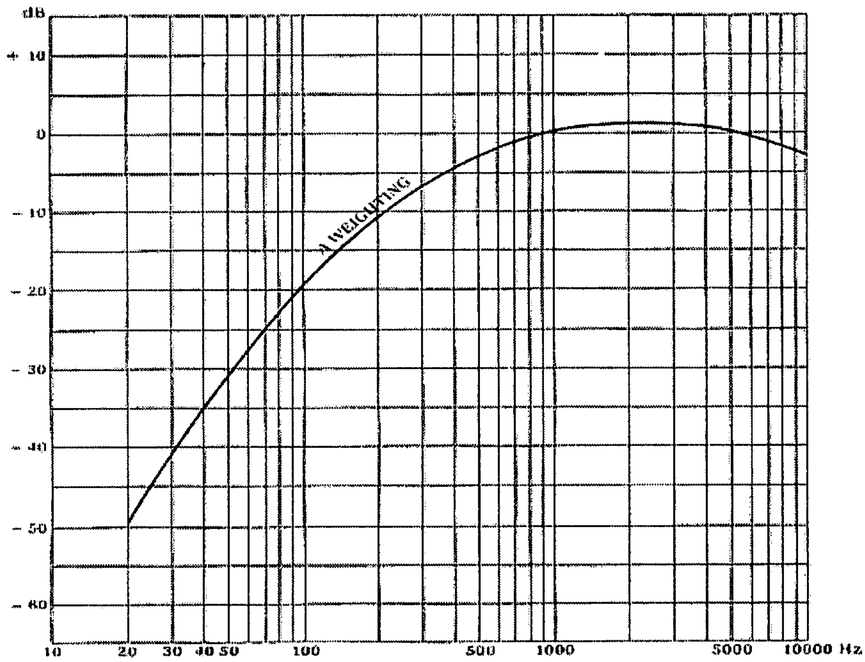


Figure 5: A-Weighting dB (A), Relationship between Frequency and Level (Adopted from ProAV, http://www.bnoack.com/index.html?http&&&www.bnoack.com/data/A-weighting.html)

[19] 'A' weighting is based on the 40 phon equal loudness contour.

2.2.2 Frequency selectivity, masking and critical band

The following section is primarily based on the explanation of frequency selectivity by the auditory system found in Moore (1989: 84-95), with the use of additional sources referenced where applicable. Frequency selectivity plays an important role with regards to auditory perception and also concerns the ability of the human ear to identify the sinusoidal components within a complex sound. This phenomenon can best be demonstrated by examining another phenomenon known as masking. Within this context the concept of critical band will also be addressed due to its strong bearing on masking.

To begin with, the American Standards Association[20] (1960, quoted in Moore, 1989:84) defines masking as follows:

- The process by which the threshold of audibility for one sound is raised by the presence of another (masking) sound. (2) The amount by which the threshold of audibility of a sound is raised by the presence of another (masking) sound. The unit customarily used is the decibel.

To this White (1991: 197) adds: "Masking is a subjective phenomenon wherein the presence of one sound will inhibit [the] ability to hear another sound."

More concretely, Butler (1992: 83) explains that a given sound, with a certain frequency content, forms patterns of excitation on the membrana basilaris once it has been processed by the mechanism of the auris externa and auris media. If a second signal, with similar frequency content, then reaches the auris interna and results in a second excitation pattern that coincides with that of the first signal, the result will be a loss of net energy. In other words, the sum of the loudness sensations will be lower than it would have been in the case of two tones with identical intensities but different frequencies.

---

[20] American Standards Association (1960) *Acoustical Terminology SI, 1-1960.* New York: American Standard Association.

It should further be noted that the respective frequencies of the two signals do not have to coincide exactly in order for masking to take place. The reason for this is that a given signal does not just stimulate one single point on the membrana basilaris, and in fact results in stimulation over a fairly broad region called the *critical band.* The latter consists of a central point of maximal stimulation with a section of diminishing response to the signal stretching in both directions around the central point on the membrana basilaris (Butler 1992: 83).

In terms of frequency, a critical band consists of a whole tone on each side of the point of maximal stimulation. By implication, an entire critical band comprises a frequency region that is roughly the equivalent of a major third, which is a third of an octave. The closer two points of maximal stimulation are to one another, the greater the competition becomes around the limited number of nerve receptors in the involved region. The situation is, however, compounded when the ear is dealing with complex tones because the upper partials of such tones stimulate different regions of the membrana basilaris. A further energy loss will therefore result should any of the critical bands of a second signal coincide with any critical bands of a first signal (Butler 1992: 83). Butler adds that a complex tone with a lower frequency has a certain advantage in competing with higher frequency tones for space on the membrana basilaris. Although the upper partials of the former may have to compete with the upper partials of the latter, the lower frequency partials of the former are not affected in any way by the fundamental and partials of the latter.

Characteristics of the transmission function of the auris externa

Motivation for a closer examination of the transmission function of the auris externa can be found in the following statement (Sound Retrievel Systems [SRS] Labs Inc. 1998):

- Due to the complex shapes of the pinna and concha, sound impinging on this area is subject to reflection, reinforcement, and cancellation at various frequencies. Effectively, the system functions as a multiple filter, emphasizing some frequencies, attenuating others, and letting some get through with no change. The response changes with both azimuth and elevation, and together with our binaural capabilities helps us determine whether a sound is coming from up, down, left, right, ahead or behind.

From the above it is clear that the auris externa plays an important role in identifying the position of sound sources. Blauert (1999: 63) further notes that sound reaching and travelling through the auris externa is altered due to the reflection , shadowing, dispersion, diffraction, interference and resonance that takes place therein. In view of the fact that sound travelling through the auris externa is moving through a linear system, these can be described as linear distortions. In this regard, Blauert states that these alterations can be described by the linear system's transfer function, the latter defined by him as follows (1999: 78): "The complex ratio of the Fourier spectrum of the output variable to that of the input variable."

As far as the auris externa is concerned, Blauert (1999: 78) provides the following three types of transfer functions:

- Free-field transfer function. This relates sound pressure at a point of measurement in the auditory canal of the experimental subject – preferably at the eardrum – to the sound pressure that would be measured, using the same sound source, at a point corresponding to the centre of the head (i.e., at the origin of the coordinate system) while the subject is not present.

- Monaural transfer function. This relates sound pressure at a point of measurement in the ear canal for any given direction and distance of the sound source to the sound pressure measured at the same point but with the sound source at a reference angle and distance.

- Interaural transfer function. This relates sound pressures at corresponding points of measurement in the two ear canals. The reference sound pressure is that at the ear facing the sound source.

The way in which the human brain processes signals it receives from the auris interna will now be examined more closely.

2.3. Cognitive aspects of human audition

2.3.1. Auditory Scene Analysis [ASA]

The present discussion of certain aspects of *auditory scene analysis* (ASA) is based principally on research done by Albert S. Bregman since 1960. His book, *Auditory Scene Analysis: the Perceptual Organization of Sound*, is regarded by scholars as a fundamental source in the field of *psychoacoustics*. The use of additional sources is referenced where applicable. To begin this discussion, it is sensible to start by examining Bregman's view with regards to the role that auditory scene analysis can play in the field of technology (1990: 3):

There are some practical reasons for trying to understand this constancy. There are engineers currently trying to design computers that can understand what a person is saying. However, in a noisy environment the speaker's voice comes mixed with other sounds. To the naïve computer the different sounds that a voice comes mixed with appears to be different words, or as if spoken by different people. The machine cannot compensate for the particular listening conditions the way human beings can. If the study of human audition were able to lay bare the principles that govern the human skill, there is some hope that a computer could be designed to mimic it.

Essentially, ASA aims to address perceptual questions such as the number, characteristics, and locations of the sound sources received by the auditory system. The latter system approaches such questions by splitting a received complex sound into smaller components and then grouping these components into streams (Chang 2004: 9). The grouping mechanism employed by the auditory system determines which segments belong to the same sound source, with the implication that each stream that is formed, constitutes complete perceptual representation of a given sound source. The latter process is known as *auditory scene segregation*, and represents the cardinal process involved in auditory scene analysis. Bregman (1990: 10) states that the stream plays the same role in auditory mental experiences as the object does in the visual sphere. He further distinguishes clearly between the term "stream", on the one hand, and "sound" or "acoustic event" on the other. It is important to note that Bregman (1990: 10) prefers the former term to the latter two terms and motivates this preference as follows:

The word "sound" refers indifferently to the physical sound in the world and to our mental experience of it. It is useful to reserve the word "stream" for a perceptual representation, and the word "acoustic event" or the word "sound" for the physical cause.

A next important aspect of ASA is the fact that many tenets of ASA are derived from studies in the field of Gestalt psychology (Chang 2004: 9). The latter is concerned with a theory formulated in the early 20th century in Germany by Max Wertheimer, Wolfgang Köhler and Kurt Koffka (Palmer et al, 1990: 84). Before the 20th century, most psychologists supported the structuralistic approach, which states that perception of the "whole" is made entirely of the sum of its parts. Gestalt psychologists, on the other hand, believed that perception is a much more complex process, and thus formulated several laws of perceptual organization to counter structuralism (Chang 2004: 9). Parncutt (2004: 14) provides a summary of the relevant Gestalt principles with regards to both vision and auditory perception in Table 1.

Table 1: Names and explanations of some well-known Gestalt principles in visual and auditory perception (Parncutt, 2004: 14).

Name	visual	auditory or musical
proximity	An object's contours tend to be physically close to each other.	The tones of a melody are close to each other in pitch and time; if not, the melody breaks perceptually into fragments (Noorden, 1975). The tones of a chord fuse when their onsets are synchronous (temporal proximity) and not too widely spaced (pitch proximity).
similarity	An object's contours tend to look similar to each other.	The tones of a melody are similar in timbre; if not, the melody breaks up perceptually into fragments (Wessel, 1979).
closure	Some of an object's contours may be imperceptible due to occlusion or masking by other objects.	Harmonic complex tones fuse perceptually even if one or more partials (including the fundamental) are physically missing or inaudible.
common fate	An object's contours tend to move in synchrony with and at the same speed as each other, when the object moves.	When the frequencies and/or amplitudes of the partials of a complex tone move in parallel (e.g. in a musical *vibrato*, in which frequency and/or amplitude ratios are held constant), the tone tends to fuse perceptually, even if the spectrum is not harmonic.
good continuation	An object's contours tend to be smooth (straight, or with a large radius of curvature) and not to change direction suddenly.[10]	This principle applies to continuations following melodic steps but not following large leaps, which are typically followed by a change in direction (Huron, 2001).

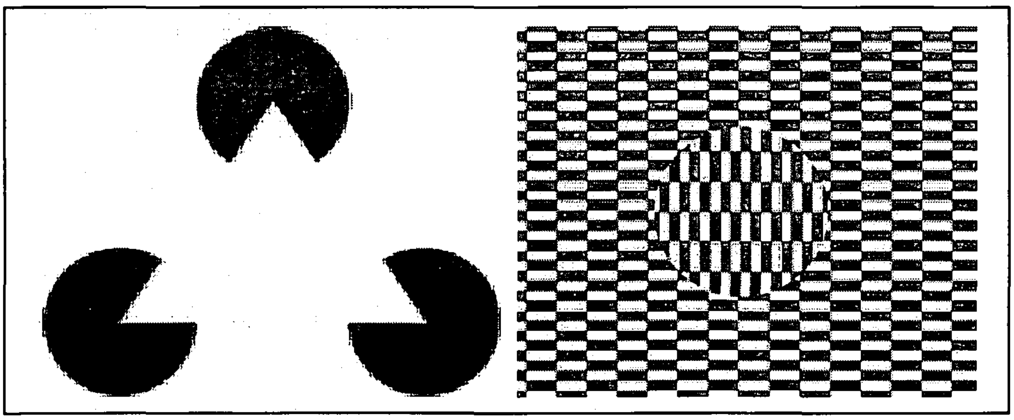


Figure 6: Gestalt at work (Smaragdis 2001: 50)

The illustration on the left-hand side of Figure 6 creates a percept of a white triangle covering three black circles. Although the triangle is not explicitly drawn, it is inferred by the placement of the black circles. Likewise, the ouchi-illustration on the right-hand side of Figure 6, creates a percept of a circle hovering over a plane, even though the drawing just reorients some of the rectangles (Smaragdis 2001: 50).

2.3.2. Primitive Auditory Scene Analysis

Bregman describes two mechanisms involved in ASA, respectively: (1) the use of primitive processes of auditory grouping (primitive); and (2) direction of the listening process in accordance with schemas (schema-based) that incorporate knowledge of familiar sounds. He believes that primitive segregation and grouping are inherent and relate more closely with Gestalt principles. These mechanisms have been described as being bottom-up[21] and involve the breaking down of sound signals into many elements for analyses. The grouping takes place in two dimensions across time, which involves so-called *sequential integration*, and frequency, which in turn involves *simultaneous integration*. These two forms of integration warrant closer attention.

Sequential integration

This takes place when a series of notes rapidly leaps up and down between different frequency regions. A simple example of this would be a swiftly repeated alternation between a high- and a low tone. If the speed of this alternation is fast enough, the listener will not perceive it as being a single stream of alternating tones, but will experience it as being two streams, each consisting of a repetition of one of the two tones. This will, however only take place if the frequency separation between the two tones is great enough. In such an instance where two streams are heard, two sounds will be perceived, respectively a high one and a low one, of which the tones happen to occur at the same time.

[21] Snyder (2000: 7) distinguishes between *bottom-up* (perceptual- or stimulus driven) and *top down* (cognitive- or concept driven) processing.

Simultaneous integration

Although spectrographically a complex sound may show overlapping elements in both time and frequency domains, simultaneous integration points to the ear being capable of recognizing similarities between sections of such complex spectral content not occurring at random. These include 1) Similarities between the auditory characteristics of sound events combined at different points in time, and 2) disintegration of the older sound spectrum from within the newer, creating a more audible remnant.

Grouping is dependent on different cues derived from the analysis of the elements. Chang (2004: 11) provides a summary of relevant cues in the auditory domain that shows similarity with Gestalt principles.

- Frequency/pitch proximity

Drawing on experiments conducted by Miller and Heise,[22] Chang concluded that two pitches in close proximity (in terms of time and frequency), tended to be grouped as part of the same "stream". Most importantly, this had to be understood in relation to the Gestalt principles "proximity" and "similarity". He pointed to the fact that important research in this regard was reported on by Van Noorden[23] in his dissertation, "Temporal coherence in the perception of tone sequences."

- Presentation rate

Tones separated by brief intervals are assigned the same "stream". Furthermore, Miller and Heise reported similar findings in a study, the *The trill threshold* of 1950.

[22] Heise, G.A. and Miller, G.A. (1950) *The trill threshold*, in: *Journal of the Acoustical Society of America 22*. pp. 637-638.
[23] Van Noorden, L.P.A.S. (1975) *Temporal coherence in the perception of tone sequences*. Ph.D. Dissertation in Eindhoven University of Technology.

- Similarity of timbre

In short, instruments emitting the same timbre are categorized or grouped together, a concept which is directly related to the Gestalt principle of "similarity".

- Spatial location

In accordance with the Gestalt "proximity" principle, Chang points out the fact that sounds were grouped according to the point from where they were emitted or where they originated.

- Spatial continuity

Sound sources are often not stationary. Furthermore, this movement is often not rapid and/or seamless. The latter, especially, plays an important role in Gestalt "good continuation".

- Sound continuity and smooth transition

The above can be expanded by adding that continuity with respect to intensity, spectrum and both time and frequency, assist in grouping.

- Onset / Offset

If two sound elements have the same onset or offset time, they are more likely to be grouped as one sound stream (A.S. Bregman and S. Pinker, "Auditory streaming and the building of timbre"). This cue is related to several of the Gestalt principles. Proximity and similarity should play a role since the elements share similarity in the time domain. This may also relate to common fate, since the elements [may] have the same temporal patterns.

- Loudness differences

The ease with which sound elements can be distinguished is to a degree reliant on the difference in loudness levels between them.

- Common amplitude and frequency modulation

Tones subjected to AM and/or FM modulation at the same time, are relegated to the same "stream" as are tones sounding together, and modulated in a similar manner, are grouped. In the case of tones rich in overtones, these are perceived as an entity – an occurrence related to "common fate".

- Cumulative effect

A cumulative effect will often determine whether the auditory system divides a sound sequence into separate streams or whether it remains a single stream. Chang (2004: 13) mentions that Bregman has found that the effects of the division of sound into streams can be influenced by sounds heard just a few seconds before the commencement of a given sound. To this, he has added that division into streams requires a few seconds after a period of silence to take place. He believes that the auditory system has to, as it was; lose the coherence in what it perceives. This means that the auditory systems sets out with the perceptual state of one single stream, and division into streams gradually takes place as auditory streaming sets in.

- Collaboration and competition

It should firstly be noted that this effect is not pertinently covered by Gestalt principles. This cue refers to the fact that, when looking at all the cues outlined above, some cues will be dominant, depending on the stimuli, while some cues will strengthen grouping or division in the presence of other cues. Consequently, this effect may be of some use when constructing a computational system that considers these cues for processing.

2.3.3. Schema-based grouping

Chang (2004: 11) notes that Bregman has described the other mechanism involved in ASA as schema-based. Bregman (1990: 734) provides the following definition for schema:

- In cognitive theory, an organization of information [inside the brain] affect to some regularity in his or her environment. Sometimes it is conceptualized as an active structure analogous to a computer program, and sometimes as similar to a complex organization of "memory" records in a computer. In all cases it is abstract enough to be able to fit a range of environmental situations. They are conceived of as being at different levels of generality. Examples are the schemas for "causality", "space" and "bread".

Schema-based integration firstly entails that a listener has to pay attention in order to "listen" for a sound. Secondly, it requires the use of previously acquired knowledge of, or familiarity with, the sounds to facilitate integration. While primitive processes split that which is received in accordance with evidence received by the auditory system, schema-based processes choose directly from the evidence. This then indicates that the latter processes can be regarded as being top-down.

In the final analysis, Auditory Scene Analysis fundamentals rest on primitive, data-driven cues for grouping. The latter was formulated around the 1930s by Gestalt psychologists. In addition, perception draws on scheme-based information where existing knowledge, concerning sound in general, assists in grouping sounds (Wrigley 2002: 21).

In addition to Bregman, David Griesinger[24] formulated theories regarding perception and grouping of spatial sound, emitted during artificial (recreated) listening conditions. According to Rumsey (2001: 44) these concern physical cues that control forms of spatial impression. Griesinger commented that the associated spaciousness (image) of a source was perceived as

[24] Griesinger, D. (2000) *The theory and practice of perceptual modeling – How to use electronic reverberation to add depth and envelopment without reducing clarity*. 21st Tonmeistertagung, Hannover, Germany. pp. 24-27.

part of the source. According to him it was the source and not the acoustic surroundings that conveyed spaciousness.

Finally, it should be noted that there existed an important link between perceptual streaming as discussed above, and Griesinger's CSI (continuous spatial impression), ESI (early spatial impression) and BSI (background spatial impression). Concerning CSI, Griesinger found that in the presence of a continuous sound, not segmented into events, the interaction with reflected energy and interaural variations in amplitude and time delay generated a feeling of "surroundedness" (Rumsey 2001: 44). ESI was related to "segmentable" sound events that generated a foreground stream during which energy of a reflected sound event is discharged within 50 ms. BSI concerned the energy reflections that occurred within larger acoustical environments within 50 ms of the demise of a sound.

The evaluation of recreated sound in listening areas with short reverberation times relied on recorded BSI, rather than that supplied by the room (Rumsey 2001: 44). In addition, BSI could be subjectively evaluated with terminology that is dependant on surroundings. Descriptions of CSI and ESI benefited from hybrid terminology employed in describing sources' spaciousness.

CHAPTER 3

LOCALIZATION OF SOUND

A number of acoustical concepts and their associated terminology need to be addressed before an introductory discussion of a complex topic such as sound localization can be attempted. It should also be noted that the discussion of aspects of sound localization found hereunder is aimed at providing an overview of this topic. A more extensive discussion of this topic can be found in Blauert's *Spatial Hearing* (1999) and in *An Introduction to the Psychology of Hearing* (1989) by Moore.

3.1. Acoustics

Motivation for this section is found in the following statement by White (1991: 7):

> "Acoustics is the study of sound and its interaction with the human hearing mechanism."

The primary importance of this interaction relates to sound localization. Since the physiology of the human ear has already been discussed in Chapter 2, the following discussion will therefore be restricted to concepts dealing specifically with sound and its interaction with the human auditory system. Of particular significance for the topic of the present dissertation is the role of acoustics in the localization of sound.

3.1.1. What is sound?

> The German Standard DIN 1320 (1959) defines "sound" ... as "mechanical vibrations and waves of an elastic medium, particularly in the frequency range of human hearing (16 Hz to 20kHz)."

To this Moore (1989: 1-2) adds that sound is the result of the movement or vibration of an object. This motion of vibration then impinges itself on the circumjacent medium, usually air, affecting a series of changes in pressure. This means that atmospheric particles (i.e. molecules) are compressed more densely than usual by the sound wave – a process known as "condensation". "Rarefaction" describes the opposite or "thinning" effect, which particles undergo during the generation of a sound wave. Although the sound wave itself propagates or moves outward from its source, the molecules themselves do not move ahead with the sound wave, but only vibrate about an average point of rest. In general, a sound wave will lose strength as it moves further away from its source. Depending on the immediate surroundings of the sound source, a sound wave may also undergo reflections and refractions as it impacts on walls and / or objects in its path. A very important consequence of this is that the sound 'image', as Moore refers to it, that reaches the ear will be somewhat different from the sound that was originally generated.

Furthermore, sound waves can propagate in an omnidirectional manner, which is in all directions around its source, or a sound wave may take on directional characteristics resulting in propagation in a specific direction (Jenkin et al., 2003: 4).

3.1.2. "Near field" vs. "Far field"

Two concepts (Jenkin et al., 2003: 9) are significant when describing the distance to a sound source in the area of physical acoustics[25]: *Far field* (distance to sound source is large), where planar sound waves reach the listener and *near field* (distance to sound source is very close), where the sound waves are curved in relation to the listener's head so that spherical sound waves are prominent (Figure 7). To avoid ambiguity between "very large" and "very close"

[25] White (2002: 7) defines physical acoustics as a scientific discipline in which measurable objective parameters of sound, as well as its behaviour in any medium, is studied.

source distance, Jenkin et al (2003: 9) rely on the following definition for near field provided by Brungart and Rabinowitz[26]:

"...the region of space surrounding the listener within a fraction of a wavelength away from the sound source."

By reason of this inverse relationship between frequency and wavelength, the identification of near and far field is influenced by frequency. For a more pragmatic examination, far field is assumed if the distance to the sound source is greater than 1 meter and the propagating waves show a similarity with planar waves. Therefore, the source distance can be ignored in the presence of binaural[27] localization cues. In the near field, where spherical waves are prominent, ILD- and spectral cues are very dependent on the distance of the sound source and are influenced by the size of the head as well as the structure of the pinna.

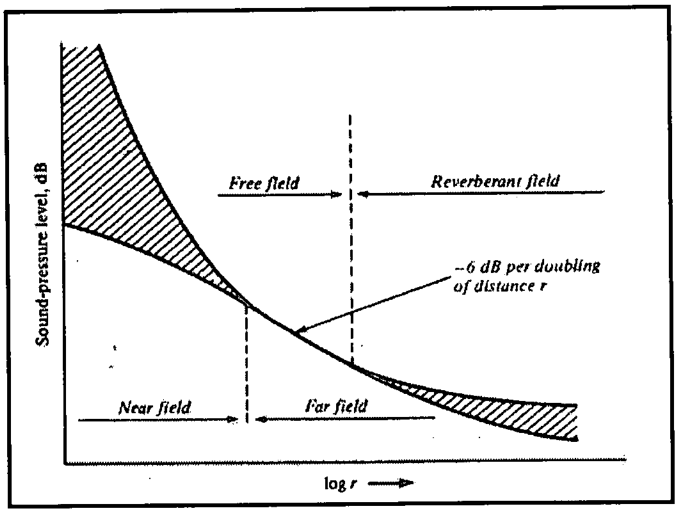


Figure 7: A Change in sound pressure level over source distance (Lamancusa 2000: 8.1)

[26] Brungart, D.S. and Rabiowitz, W.R. (1996) *Auditory localization in the Near-field.* The Proceedings of International Conference on Auditory Display. [on-line]. Available: http://www.icad.org/websiteV2.0/Conferences/ICAD96/proc96/brungart.htm#info [Jaunary 2, 2006].

[27] Bregman (1990: 730) states that this term concern two ears and a sound will resemble binaural characteristics if both ears are presented with the sound.

3.1.3. The coordinate systems

Sound events take place within a three-dimensional physical space where the position of a sound source is given relative to a chosen point of reference. In cases where only a single listener is involved, the listener will normally be chosen as the reference point and sound source positions will be defined relative to the listener (Jenkin 2003: 9). Two coordinate system types can be applied in order to represent any location point with reference to three coordinates, which will now be discussed.

The Cartesian[28] coordinate system

As can be seen in Figure 8, this system charts the relationship between the source and the listener's head by relying on coordinates utilising length, width and height (West, 1998: 4 [Chapter 2]). In this system the listener's head, as reference point, is divided by a lateral (positive x-axis connecting the ears), a longitudinal (positive y-axis horizontally pointing ahead) and a vertical (positive z-axis vertically pointing upwards) axis (Jenkin, 2003: 10). The angles, at which the axes intersect in this system of reference, result in the formation of three planes, namely the median-, frontal- and the horizontal plane.

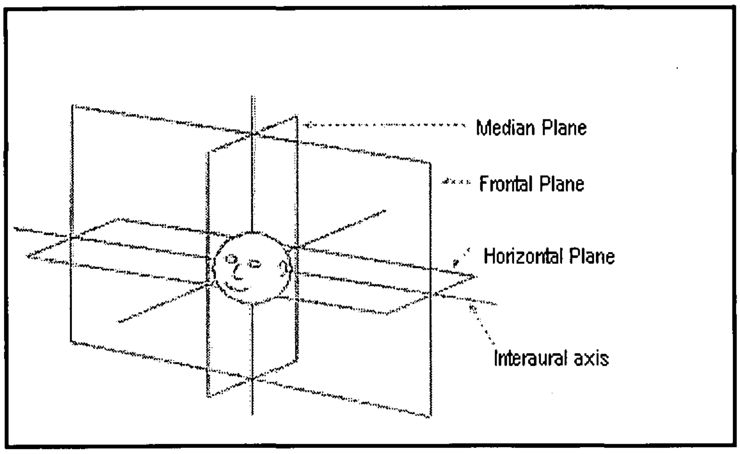


Figure 8: The Coordinate system (Jenkins et al.,2003: 11)

[28] This system, named after the French philosopher and mathematician R. Descartes (1596-1650), refers to a system for locating a point by reference to its distance from two or three axes intersecting at right angles (Dana 1995).

The spherical coordinates system

This system is a more listener-oriented system that uses the centre of the listener's head as the origin, with coordinates being identified by means of azimuth, elevation and range (Mackensen 2004: 11). It should be noted that this coordinates system can be approached in one of the following two ways (Jenkin et al., 2003: 10):

(1) **Single polar**, to which Jenkin et al. (2003: 10) ascribes the following characteristics:

> "... the centre of the head defines the origin while azimuth (θ) and elevation (ϕ) are specified by lines of latitude and longitude respectively. An azimuth angle of 0° is directly in front (e.g. median plane) while an angle of -90° is directly to the right (e.g. moving clockwise from 0° results in negative azimuth angles). The horizontal plane is at an elevation of 0° and moving upwards from this point, elevation increases positively, with +90° directly on top of the head. Range specifies the distance between the origin (centre of the head) and the point of interest."

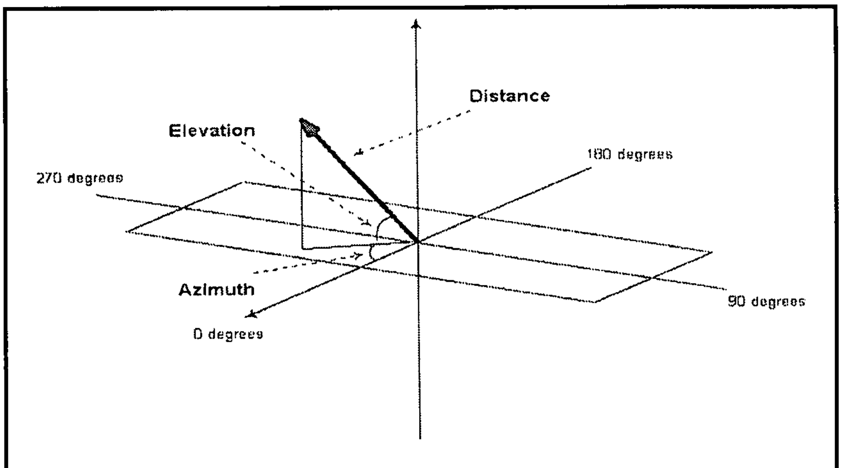


Fig. 2.4. Position of the sound event's position relative to the center of the head, with reference azimuths labeled [16].

Figure 9: Position of the sound event's position relative to the centre of the head, with reference azimuths labelled (West, 1998: 6 [Chapter 2])

While the single-pole approach is often applied, aspects have been identified within this system that may cause problems, particularly with regards to the measurement of the *head related transfer functions* (HRTFs). A notable instance of such a problem is the fact that the length of a semi-circle between two angles of azimuth is dependent upon elevation. The length of the semi-circle between 0° and 90° azimuth at an elevation of 0°, for example, is greater than the length of the same semi-circle at an elevation of 75° (Jenkin et al., 2003: 10).

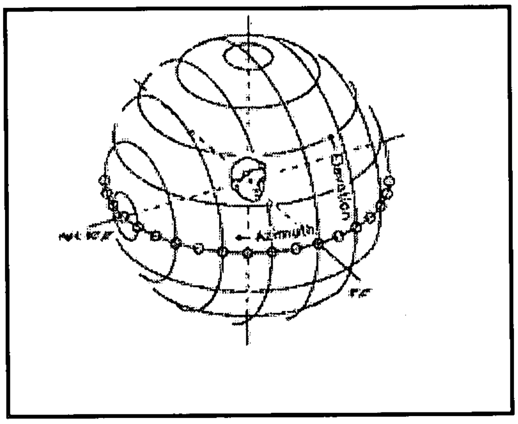


Figure 10: The single polar coordinate system (Jenkin et al., 2003: 11)

(2) **Double polar**

Although elevation is identified in the same way as in the single pole system, azimuth is given as a series of rings, which are parallel to the midline (the z-axis) and centred at the poles at each interaural axis (positive x-axis). In this approach, the length of the semicircle between two angles of azimuth is independent of elevation. This system is, however, not as intuitive as the single-pole system and is therefore not widely used (Jenkin et al., 2003: 10).

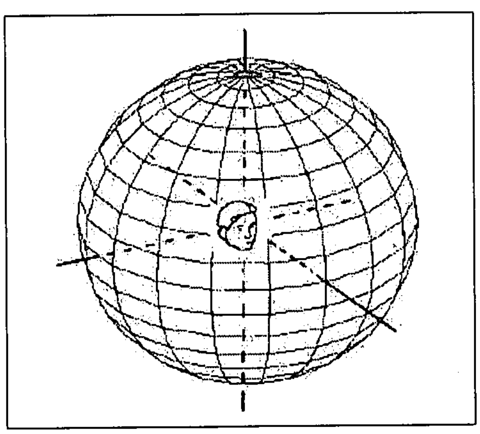


Figure 11: Double Polar System (Jenkin et al., 2003: 11)

3.1.4. Sound absorption

In the first instance a degree of sound absorption takes place as a sound wave propagates through a given medium. It should be stressed that the amount of sound absorption that will take place in such an instance, is greatly determined by the characteristics of the given medium. Although a detailed discussion of the influence of these characteristics falls outside the scope of the present dissertation, it is important to note that the absorption a sound wave undergoes while proceeding through a given medium, will alter the sound spectrum of that wave to some extent (Jenkin 2003: 5).

In addition to this, a sound wave will be subject to further absorption by obstacles in its plane of motion, within its medium of propagation. While travelling sound waves pass small, non-porous obstacles virtually unhindered by bending around them, larger surfaces that are both more flexible and porous (e.g. room boundaries) will tend to exert a greater influence on [impacting] sound waves. When sound waves hit a larger surface, this surface will respond to the impacting sound waves in any or all of three possible ways: (1) a certain amount of energy may be reflected; (2) a certain amount of the energy may be absorbed; and/or (3) a certain amount of the energy may propagate through the boundary and be transmitted as sound waves on the opposite side thereof. Although the degree to which sound waves may undergo any or all of these effects is not significantly influenced by sound intensity in a typical context; sound frequency and the angle of incidence may have the most significant effect in this regard (JBL Professional 1999: 5-1).

It must be stressed that sound absorption is of distinct consequence in relation to the topic of the present dissertation. This is due to the fact that the majority of sound reinforcement equipment is designed to function within an enclosed area. The acoustical characteristics of such an area will necessarily play a substantial role in determining the requirements and performance of any given sound reinforcement system (JBL Professional 1999: 5-1). After all, most enclosed listening environments will be echoic rather than anechoic, with the former referring to a degree of reflection and reverberation that will take place and the latter pointing to a distinct lack of this. It must be emphasised that, although anechoic environments can be artificially created, they are very rarely encountered in nature (Jenkin et al., 2003: 5).

Consequently, one needs to be able to determine how much energy will be lost every time a sound wave hits a boundary surface or object in the room when confronted with the acoustic characteristics of an enclosed space. For this use is made of absorption coefficients that, unless otherwise noted, indicate average absorption over all possible angles of impact. Furthermore, such absorption coefficients are usually provided for a number of different frequency bands. Generally such frequency bands are one octave in range and possess standard centre frequencies of 125 Hz, 250 Hz, 500 Hz, 1 kHz, and so forth. For the purposes of sound design it usually suffices to take into account the absorption characteristics of a given material, in three of four frequency ranges (JBL Professional 1999: 5-2).

The scale on which absorption coefficients are measured, assigns a value from nought to one to the absorption characteristics of a given material. This scale is based on a concept established by Wallace Clement Sabine (1869-1919), the pioneer of modern architectural acoustics, in 1889 (Acoustic Engineering 2006). Sabine proposed that an open window should be taken as an example of a perfect sound absorber since it reflects no sound at all. He postulated that such a window would then possess an absorption coefficient of 100%; therefore 1 (one) on his scale. At the opposite extreme of his scale would be a material that reflects absolutely all sound, giving it an absorption coefficient of nought (JBL Professional 1999: 5-2).

Although the absorption coefficient, as such, is mainly concerned with how much energy a given sound wave will loose when coming into contact with a given medium, it also gives an indication of the intensity of the reflection of sound that will take place. When a given material possesses an absorption coefficient of 0.2 at a determined frequency and angle of impact, for example, it means that 20 % of the sound energy will be absorbed while the remaining 80 % will be reflected. This implies that the material has what can be called a reflection coefficient of 0.80. In order to express the energy loss of the reflected sound in terms of decibels, the following 10 log function is necessary: $10 \log_{10} 0.8 = -0.97$ dB. In this case the result means that the ratio between reflected and direct sound energy is approximately -1 dB, indicating that the reflected wave is 1 dB weaker than it would have been if the surface it hit had been 100 %

reflective. Table 2 gives an indication of the relationship between absorption coefficient, reflection coefficient as well as the latter expressed in terms of dB:

Table 2: Absorption coefficients with their corresponding reflection coefficient with the latter expressed as a function of the former (JBL Professional 1999: 5-3)

Absorp. Coeff. (α)	**Refl. Coeff. 1-α (γ)**	**Refl. Coeff. dB**
.01	.99	-.044
.02	.98	-0.88
.03	.97	-.13
.04	.96	-.18
.05	.95	-.22
.06	.94	-.27
.07	.93	-.32
.08	.92	-.36
.09	.91	-.41
.10	.90	-.46
.20	.80	-.97
.30	.70	-1.5
.40	.60	-2.2
.50	.50	-3.0
.60	.40	-4.0
.70	.30	-5.2
.80	.20	-7.0
.90	.10	-10.0
.96	.05	-13.0

With regards to absorption coefficient, it is important to note that more recent publications have the tendency to prefer to express absorption in an enclosed space in terms of *average absorption coefficient*. This has the distinct advantage of not being attached to a particular measurement system. An average absorption coefficient of 0.15, for example, remains the same irrespective of the unit of distance in which the surface area of the enclosed space is measured. Another advantage of the use of an average absorption coefficient is that it facilitates establishing *reverberation time*,[29] direct-to-reverberant sound ratio, and steady-state sound pressure (JBL Professional 1999: 5-3).

3.1.5. Acoustics of enclosed spaces

Regarding the acoustics of enclosed spaces, it is necessary to briefly mention two concepts that can be used to test the suitability of a given room. These are: (1) reverberation or liveliness; and (2) background noise levels (Lamancusa 2000: 8.1).

Reverberation or liveliness

This is primarily a function of the sound absorption in a room and is quantified in terms of *Reverberation Time*. This concept shall be addressed in greater detail in the next section.

Background noise levels

The main sources of background noise in enclosed spaces are normally collectively referred to as *Heating, Ventilation and Air Conditioning* (HVAC) noise (Lamancusa 2000: 8.1). Due to the fact that the human ear is relatively insensitive to low levels of low-frequency noise as can be inferred from the Fletcher-Munson effect, a comparatively greater degree of low-frequency noise can be present in auditoriums and recording studios (White 1991: 217). The desire to translate the apparent noisiness of a given room into an objective measure has lead to the development of the *Noise Criterion* (NC)[30] curves (Lamancusa 2000: 8.1 and White 1991: 217). These curves are basically contours of equal loudness that are spaced 5 dB apart and numbered. NC-15, for example, describes an environment that is so quiet that the average person would not even be aware of the presence of any background noise. Furthermore, the NC value of a given room can be determined by using a sensitive sound level meter equipped with a band filter with an octave range. The measured sound levels can then be plotted on a graph showing the NC curves (White 1991: 217-218).

---

[29] According to White (2002: 283), "The time of reverberation is defined as the time it takes for the Sound Pressure Level to decay to one-millionth of its former value. This is a 60-decibel reduction in level".

[30] *Noise criterion* referres to the background noise in a room (White 1987:217).

3.1.6. Reverberation with enclosed spaces

White (1991: 282) describes reverberation as the remainder of sound that exists in a room after the actual source of the sound is stopped, sometimes miscalled "echo". All rooms have reverberation, and an important subjective quality of a room is its reverberation time, although other factors, such as ratio of direct-to-reverberant sound, are probably more important. In a real room, the sound heard by a listener is a mixture of direct sound from the source and reverberant sound from the room. Reverberant sound is diffuse, coming from random directions, the direct sound allowing localization of the sound source. Moving further away from the source, the direct sound becomes weaker whilst the reverberant sound becomes relatively stronger. At a certain point the two would be equal in strength, and this is sometimes called the critical distance. The objectively measured reverberation time of a room is not necessarily heard by a listener, for the reverberant sound may be high or low in level compared to the direct sound. Most music recordings have some reverberation recorded on them along with the direct sound, and this leads to a sensation of room ambience (special reality). Usually, recording studios are lacking in reverberation, and so synthetic reverberation is mixed with the music signal when compiling the master tape.

Direct sound and early reflections

As sound waves travel at a speed of approximately 345 m/s, the sound produced by a sound source in a large room will reach listeners in that room after a time of between 0.01 and 0.2 seconds. Scavone (2006)[31] points out that the strength of direct sound will decrease by 6 dB each time the propagation distance is doubled. He adds that the auditory system's identification of the direction of a sound source is based on the direct sounds reaching the ear. Reflections reaching the listener within approximately 50 to 80 ms after the direct sound will not even be perceived as being distinct from the direct sound, and will actually tend to reinforce the direct sound. With regard to the localization of sound source, Scavone (2006) makes the following important statement:

[31] No page numbers indicated in the publication of Scavone.

The source is perceived to be in the direction from which the first sound arrives provided that: (1) successive sounds 'arrive' within about 35 milliseconds, (2) the successive sounds have spectra and time envelopes reasonably similar to the first sound, and (3) the successive sounds are not too much louder than the first. This is referred to as the *precedence effect.*

The *precedence effect*, also known as the *Haas effect*,[32] refers to the way in which the human auditory system will respond to a sound that arrives at the listener's ear from two locations. The sound source will be localized based on the earlier of the two sounds, even if the second sound is stronger than the first. This is of great significance in the case of sound reinforcement systems where a number of loudspeakers are often employed. In such cases the sound will be localized at the loudspeaker which provides the earliest sound, with the others not being heard at all (White 1991: 153-154).

Soon after the direct sound reaches the listener, however, a series of less distinct reflections will be conveyed to the listener from reflective surfaces in the room, such as the walls and ceiling. These are referred to as *early reflections* and will normally reach the listener between 50 and 80 ms after the direct sound (Scavone 2006). It is important to note that, if the enclosed space in which sound waves are travelling is a large room, an echo may be perceived by the listener. For a reflection of sound to be considered an echo, an interval of more than 65 ms between the direct sound and the early reflections is required. In terms of distance, this translates into a sound path of approximately 18 m (Nisbett 1995: 38). If sound is travelling in a smaller room, on the other hand, sound colouration may occur to certain degree. This can be defined as the selective emphasis of certain frequencies or frequency bands within reverberation (Borwick 1997: 113).

Late reflections or reverberation

After the early reflections, another wave of reflections will reach the listener, but these will generally be much lower in amplitude and will typically be spaced very densely apart in time. These reflections are collectively referred to as the *reverberant sound* or *late reflections*.

[32] Refer to H. Haas (1972) "*The influence of a single echo on the audibility of speech*," Journal of the Audio Engineering Society, 20, pp. 146-159.

Important in this respect is also the *reverberation time* of a given room, defined as the time it takes for the sound level in a room to fall by 60 dB (or t_{60}). Importantly, however, the reverberation time will vary with frequency. In an empty room, where all surfaces absorb the same amount of energy from the impacting sound waves, the theoretical reverberation time will be proportional to the ratio of volume to surface area. When reverberation time is expressed in terms of cubic and square metres, it can be determined by the formula RT = 0.161 V/A, with V being the volume of the room and A the effective total absorption area[33] (Scavone 2006).

It is important to remember that air absorption will also exert an influence in determining reverberation time, in particular due to its substantial absorption of high-frequency sounds. Calculating reverberation time with this factor taken into account can be done with formula RT = 0.161 $V/(A + mV)$, where m is a constant taking into account air temperature, humidity and frequency (Scavone 2006).

3.2. Basic sound localization terminology

Within the context of spatial hearing, Blauert (1999: 37) defines two concepts to assist the understanding of spatial hearing, more specifically, *Localization* and *Localization Blur* (Carlsson, 2004: 6). Used in conjunction, information concerning placement in terms of distance and direction of the sound source, forms the basis of psychoacoustics. It is also of importance to distinct between the sound- and auditory event.

[33] Scavone (2006) states that the total absorption area "is calculated as the sum of all surface areas in the room, each multiplied by its respective absorption coefficient for a particular frequency".

3.2.1. Defining localization

In this instance, it will suffice to quote two definitions[34] of localization. Blauert (1999: 37) defines it as follows:

> "...is the law or rule by which the location of an auditory event (e.g., its direction or distance) is related to a specific attribute or attributes of a sound event, or of another event that is in some way correlated with the auditory event."

Theile (1980: 7) provides the following definition:

> "Localization is the mechanism or process that maps the location of an externalised auditory event to certain characteristics of one or more sound events."

3.2.2. Localization blur

Blauert (1999: 37) defines this term as follows:

> "It is the smallest change in specific attributes of a sound event or other event correlated to an auditory event that is sufficient to produce a change in the location of the auditory event (e.g., direction or distance again)."

Mackensen (2004: 12) expands on this by stating that localization blur takes place when the listener does not perceive any specific, clearly defined position for an auditory event. In such an instance, the listener rather perceives a blurred impression or auditory event region as opposed to a single point of sound origin.

---

[34] It should be noted that there are distinct differences in the usage of the terms *sound event* and *auditory event* between Blauert (1999: 37) and Theile (1980: 7). While Theile makes a very clear distinction between these two terms in his usage, Blauert uses them more interchangeably. The present dissertation, however, will follow the usage of these terms as found in Theile.

3.2.3. Sound event

This term refers to a physical event which results in the emission of sound waves. Such events have in common that a measurable physical quantity (e.g. air density) varies in time or other respects. Because a sound event possesses a physical character and can be measured, it occurs in the physical world and is generated by a sound source (Mackensen 2004: 10).

3.2.4. Auditory event

Where a sound event takes place in reality and can be measured, an auditory event only takes place in the mind of the listener and its occurrence cannot be proven by using measuring techniques. An auditory event is perceived and felt, be it consciously or unconsciously. An auditory event can even take place in the absence of a corresponding mechanical vibration or sound event (Mackensen, 2004: 10). In his definition of sound event, Theile (1980: 7) draws the following connection between sound event and auditory event:

> "A sound event is that part of a sound, which stems from a single sound source and which determines or influences the associated auditory event with respect to its location and gestalt."

Adding to this, Blauert (1999: 3) refers to Lungwitz[35] in emphasizing that these are distinct in terms of time, space and other attributes. This means that they take place at particular times and places and with particular attributes. It is within this context that the concept of spatial hearing is given its meaning, since it infers that auditory events are inherently spatially distinct.

---

[35] Lungwitz, H. (1933) *Lokalisation der akustischen Gegenstände [The localization of acoustical objects]*, in: *Lehrbuch der Psychobiologie, vol. 2.* Berlin: Walter de Gruyter.

3.3. Spatial theory

There are many factors involved in understanding a three-dimensional sound event. According to Carlsson (2004) there are many factors involved in understanding a three dimensional sound field. The most important parameters that contribute to a listener's perception, as stated by Carlsson, will be discussed hereunder.

3.3.1. The listener's acoustic environment

The listener's acoustical environment, where actual sound events occurs, may involve various conditions, for instance *free field-* or *enclosed space* (Carlsson 2004: 7). In the former, an omnidirectional source radiates a sound without it being reflected. Consequently, for a listener or measuring device, sound levels drop rapidly (6 dB for every doubling of the distance). A Comparison between far and near fields show that the level will drop more rapidly in the case of the latter. An anechoic room (highly absorbent, non-reflective enclosure) eliminates the unwanted influence of reflected signals over captured sound signals (Carlsson 2004: 7).

Rumsey (2001: 33) states that the qualities of "discrete" point sources are influenced by reflections, with early reflections strongly determining perceived room dimensions. Carlsson (2004: 7), drawing on Shinn-Cunningham,[36] pointed out that although reverberation aided distance perception, it clouded directional hearing.

3.3.2. The influence of the listener's torso and head (on the wave front)

Body, shoulder and especially reflections from the pinna change spectra arriving at the membrana tympani, assisting in localization (Rumsey 2001: 23). According to Blauert (1999:

[36] Carlsson, K. 2004) *Objective Localisation Measures in Ambisonic Surround-sound.* Master Thesis in Music Technology, Department of Speech, Music and Hearing, Royal Institute of Technology, Stockholm. [on-line]. Available: www.speech.kth.se/publications/masterprojects/2004/KarinCarlsson.pdf [July 9, 2006].

63), the pinna is responsible for distinguishing time and frequency components of perceived sounds. Depending on their location, its shape is responsible for their idiosyncratic change whilst influencing and linearly distorting them. Essentially, the influence of the pinna amounts to phenomena discussed in Chapter 2, including reflection, shadowing, dispersion, diffraction, interference and resonance. Furthermore, Moore (1998: 205) reports that if a perceived sound possesses high energy levels across a broad frequency range, the information provided via the pinna will be optimally effective (Carlsson 2004: 8).

In this regard, frequencies above 5 kHz seem to be of particular importance, as only frequencies with such short wavelengths will interact significantly with the pinna. Low frequencies, however, also contribute since their levels are affected by the torso and head (Carlsson 2004: 8). Because the head further presents a notable obstacle to the free propagation of sound, its perturbation of the sound field exerts a great influence on the sound signals in the pinna and meatus acusticus externus (Blauert 1999: 304).

3.3.3. The influence of the listener's outer ears on the total auditory wave front

Research done on directional sound perception, studies by Blauert (1999) and Moore (1989) being important instances, has pointed to the existence of two primary mechanisms, commonly referred to as *cues*, involved in decoding and interpreting a three-dimensional environment. These are *interaural time difference* (ITD) and *interaural level difference* (ILD), respectively, and are often referred to collectively as the *Duplex* theory. This theory was formulated by Lord Rayleigh (1842-1919)[37] and can be defined as follows (Jenkin et al., 2003: 12):

> "... a theory of human sound localization based on the two binaural cues, interaural time delay (ITD) and interaural level difference (ILD) and on the assumption that the head is spherical with no external ears (pinnae)."

[37] Before receiving his peerage he was known as John William Strutt (O'Conner *et al.*, 2003).

Interaural Time Delay (ITD)

ITD results when a sound wave originating off the centre front reaches each right and left membrana tympani at different times. The additional time it takes sound waves to travel to the ear furthest from the sound source, is influenced by the angle of incidence of the incoming sound (Rumsey 2001: 22). The path difference between the two ears can be used to calculate ITD with the following equation (Carlsson 2004: 9):

$$ITD = r\,(\theta + \sin\theta)/c$$

r = half the distance between the two ears

θ constitutes the angle of incidence in radians

c = 344 m/s (20° C)

ITD further only applies to frequencies below a certain point – generally lower than about 700 Hz – and the occurrence of ITD is most noticeable at the onsets and offsets of sounds (Rumsey 2001: 22). Carlsson refers to Blauert (1999) which states that this cue enables humans to establish the direction of sound sources to within 1°. In the case of sinusoidals, a time difference will be equal to a phase difference between the two ears. This, however, depends on the frequency of the sound and location of the sound source that is being perceived, since the distance between the two ears remains constant (Carlsson 2004: 9).

Phase shift [38] assists the listener in determining the location of sounds in the lower frequencies. As frequency increases and the signal period reaches approximately twice the ITD, uncertainty creeps in. For a sinusoid at 725 Hz at a 90° angle of incidence, mirroring waveforms are produced at the two ears. Unless the head is moved, localization is compromised as a particular waveform may be 90° out of sync with the other. Moore (1989: 197) states that above 1.5 kHz, phase difference is clouded. Consequently, by relying on their substantial interaural phase

[38] White (2002: 244) defines this term as follows: "...is a characteristic of a device and is the change in PHASE impressed on a SIGNAL that passes through the device. An AUDIO device will always add a time delay to an applied signal. If this time delay is constant at all frequencies, the phase shift between the input and output of the device will be a LINEAR function of frequency".

information, harmonically rich or complex waveforms are more easily localized than pure, high tones. Naturally-occurring sounds, fortunately, resort under the former, harmonically rich, category.

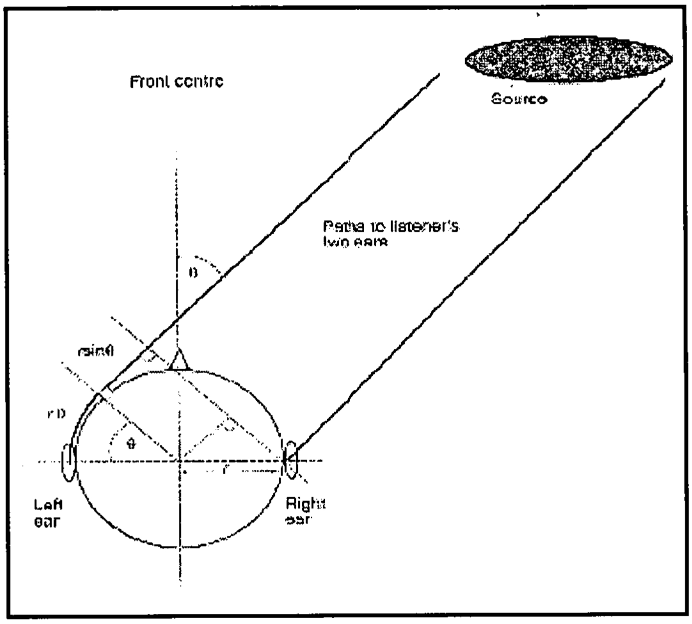


Figure 12: The Principle of ITD (Carlsson 2004: 9)

ILD

A second, important mechanism is *interaural level difference*, or ILD. This, in short, derives from the fact that higher frequencies are obstructed by the head. For high frequencies, the more distant membrana tympani is reached after diffraction occurs. Essentially, the head imposes a "shadow" to one side which leaves lower frequencies unaffected and localization biased towards the harmonically richer and "louder" side (Malham 1998: 1). See Figure 13.

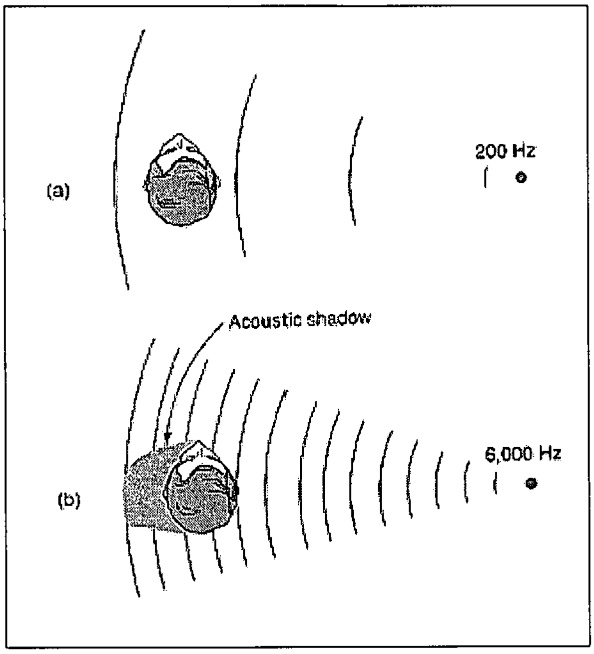


Figure 13: The principle of ILD

Additional observations on interaural cues

Research, following Rayleigh, concluded that ILDs and ITDs offered insufficient motivation for the occurrence of spatial hearing (West 1998: 8). Furthermore, in experiments where signals with ILD or ITD where presented, the listeners could make left-right judgements, but failed to distinguished between front-back placements. Woodworth refer to this uncertainty of location as the "cone of confusion which is presented in Figure 22. (Woodworth, in West 1998: 8).

HRTF

The *Duplex theory* concerns the [dominant] theory regarding human sound localization. Batteau's (1967)[39] theories involving the pinna's filtering effect (HRTF), however, brought a paradigm shift in thinking concerning sound localization.

[39] Batteau, D.W. (1967) *The role of the pinna in human sound localization.* Proceedings of the Royal Society, 168:158-180.

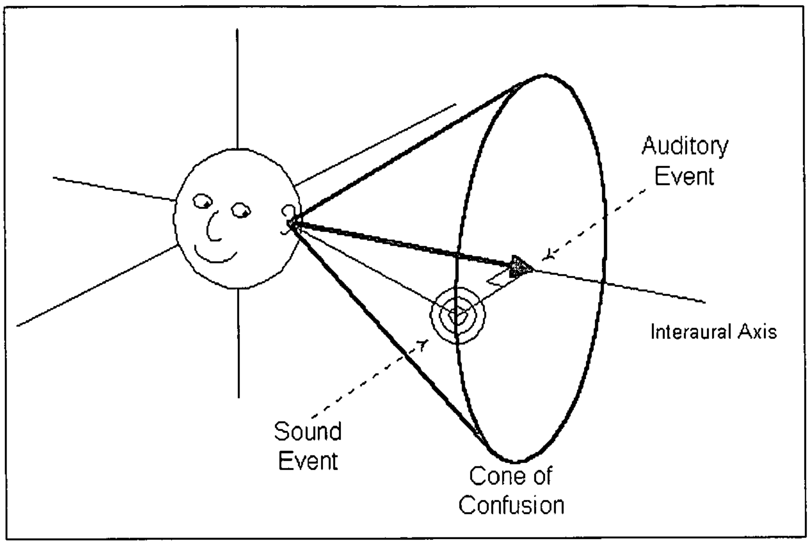


Figure 14: Graphical representation of the Cone of Fusion (West, 1998: 8 [Chapter 2])

Jenkin et al (2003:14) provides a definition of this phenomenon:

> "The filtering of the sound source spectrum caused by the complex interactions of the sound waves with the head, shoulders, torso and particularly the auris externa (pinna or auricle) prior to reaching the ear drum (in addition to the interaural time delays and level differences), are collectively known as the head related transfer functions (HRTF)."

By comparing the sound spectrum reaching respective eardrums to that of the source, the *Head-Related Transfer Function* (HRTF) as dB-ratio is deducted. This function varies systematically with direction of the sound source relative to the head and is unique for every direction, since for every angle of incidence of a sound wave, the spectrum that reaches the eardrum will be different (Carlsson 2004: 10). Individual pinna-shapes differ, implying the same for HRTF's, which inhibits generalization concerning spectral characteristics. HRTF's play a more prominent role in the horizontal plane compared to the vertical plane (Jenkin et al. 2003: 15).

3.3.4. The nature of the source.

Localizing an omnidirectional source depends on the *nature of the source* (i.e. amplitude, frequency, complexity). Most sound sources have a [directivity] pattern reflecting the deviation from omnidirectional radiation at different frequencies (Rumsey 2001: 4). The difference in deviation is sometimes expressed as a dB-gain, compared with the omnidirectional radiation at different frequencies. According to Rumsey (2001: 4-5), sources tend to radiate more directionally at high frequencies, and increasingly omnidirectional at decreasing frequencies.

3.4. Correlations between spatial attributes and physical factors

3.4.1. Sound direction

To achieve localization of a number of different sound sources, the auditory system has to establish the direction of each such source, despite coincident bombardment of reflections and superposition-effects, to which it is subjected. The two localization cues that come into play in achieving this are ITD and ILD[40]. These, however, are only relevant in instances where [only] the direct sound of a single source possesses a substantial amount of significant energy within the critical band, meaning that ITD and ILD are indicating the direction of the sound source (Faller et al. 2004: 3075).

[40] As discussed in Section 3.3.3.

3.4.2. Source Distance

In this respect, distance cues of monaural and binaural sound sources have to be considered. These involve the following: (1) frequency content; (2) binaural cues under free-field conditions; and (3) loudness and direct-to-reverberant sound. The so-called *Doppler Effect* is also relevant in view of its role in the perception of movement of a sound source.

Jenkin et al. (2003: 25) divide sound source distance cues into two categories: (1) *exocentric*, that is to say relative, providing information with respect to the relative distance between two sounds originating from two points at different distances from the listener; and (2) *egocentric*, that is to say absolute, providing information about the actual (or absolute) distance between the listener and the sound source.

Frequency content

Neher (2004: 19) found that the perception of alteration in the frequency content of a direct sound is effected by means of monaural cues. In terms of the latter, he established that distance strongly influences the available monaural cue in the case of a continuous, harmonically rich median plane source. High frequencies are absent for sound sources positioned in the far field (> 15 m), a fact that can be explained by air absorption (Blauert 1999). A simplified description of this phenomenon is a low-pass filter, where the cut-off frequency decreases and the slope increases in conjunction with a decline in source proximity (Savioja et al. 1999: 684).

With nearby sources, low frequency energy will be more significant than high frequency energy. This can be explained by the difference in nature of the wave fronts as they reach the listener, if propagated from distant (planar) vs. nearby (curvature) sound sources (Neher 2004: 20). Regardless of possible spectral distortions of the signal for sound sources closer than 3 m, Blauert points to the small impact they have for distances greater than 25 cm (Neher 2004: 20).

Loudness

As discussed earlier, the perception of loudness for any source is connected to SPL, where the inverse distance (1/r) law determines the SPL (Neher 2004: 21). The result is a drop of 6 dB per doubling of a unit of distance. Furthermore, it is worthwhile to mention the correlation that exists between loudness and tone colour, although this is of secondary importance in distance perception. An increase in the signal level will result in a heightened perception of loudness whilst the tone colour will become darker. This can be understood by revisiting the equal loudness contours discussed in Section 2.2.1

Chomyszyn (1992: 257-260) studied the accuracy with which listeners determine the distance of sound sources within a reflective environment, after their loudness levels had been matched and therefore eliminated as a cue. The conclusion was made that in the absence of loudness cues, distance could still be determined, if parameters specific to the reverberant environment was present.

Although the 1/r law announces a 6 dB increase in DS level when the physical distance from the sound source is split in half, a different level increment may be required to draw forth the corresponding psychological change. Von Békésy published a localization curve in 1949 (Neher 2004: 23) which is based on distance judgements (under anechoic conditions) from five [blindfolded] listeners. By constantly increasing the distance of the sound source, listeners could not localize the position of the sound. Consequently, Von Békésy could speculate that auditory space is limited to a certain area in space (Neher 2004: 23). In conclusion, Blauert (1999: 118) provides a classification scheme of the characteristics of the ear input signals which are dependent on the distance from the sound source:

> 1) At medial distances (3-15 m from the sound source) for point sources, only the SPL of the input signals at the ear depends on the distance from the sound source, assuming the radiated signal is constant.
>
> 2) For distances greater than 15 m from the sound source, the path between the sound source and listener is subjected to distortion. Nevertheless, the interdependency between SPL and the distance of the sound source is still current. It is independent of frequency and follows the 1/r law. But there is an additional attenuation that depends on the length of the air path and varies with frequency, where higher frequencies are attenuated more than lower ones. Not only does the SPL depend on the distance from the sound source, but the shape of

the spectrum also depends on it. More precisely, the relative level and phase curves as a function of frequency depend on it.

3) Close to the sound source (less than 3 m for point sources), the effects of curvature of the wave fronts arriving at the head can no longer be neglected. The linear distortions of the signals due to the head and the external pinna vary with distance from the sound source. Close to the sound source, the sound pressure level changes, though the way the spectrum changes is not the same as at great distances.

Direct-to-reverberant sound ratio (D/R)

Nielsen (1993: 755) undertook a study regarding the influence of several physical factors on the perception of distance. Within the context of direct-to-reverberant sound ratio (D/R), he identified two cues, namely *absolute* and *relative*. Absolute cues contain information about the distance of a given auditory event. Relative cues can only identify changes in distance. Within context, it was stated that the D/R depends on the acoustical properties of an enclosure. The listener will subconsciously gather information about the reflected sound, and will consequently be able to tell whether a source is near or far away, even when not having being exposed to the sound event before. It can, therefore, be said that D/R always provides absolute cues for distance perception. In conclusion, it can be said that reflections play a crucial role for distance hearing, which is confirmed by Nielsen (1993: 764):

> "One of the factors varying as a function of the source distance is the ratio between the direct and the reflected sound energy. This ratio is highly signal dependant as the absorption of the room is frequency dependant."

The Doppler shift

Named after C.J. Doppler[41] (See Figure 15), this effect occurs when there is an active change in the source range (Neher 2004: 35) which implies the following:

> "...that as a sound-emitting source approaches a receiver the peaks and troughs of the radiated wave fronts are compressed (i.e. closer together in space), whereas if the source retreats they are rarefacted (i.e. farther apart in space). This results in changes of the effective wavelength (and therefore frequency) at a listener's ears, that in turn manifest themselves in the form of pitch variations. "

[41] An Austrian physicist (1803-1853).

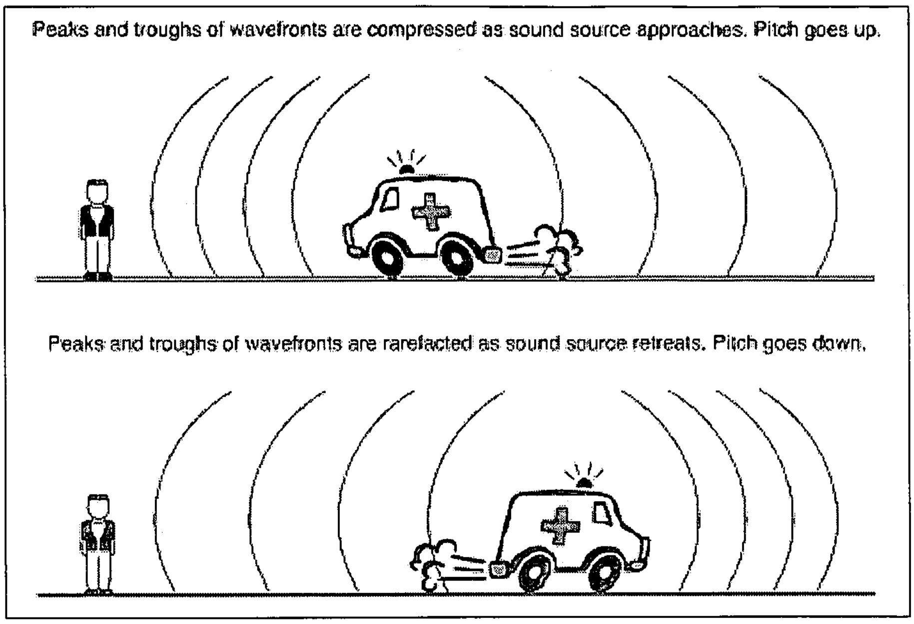


Figure 15: The Doppler Effect (Neher 2004: 35)

Neher (2004: 35) refers to Strauss,[42] who examined the influence of Doppler shifts in virtual auditory displays, by simulating a situation of a fast car passing the listener. The parameters used in the study were of more significance than the perceptual change caused by the Doppler Effect. As a result, the majority of subjects found stimuli with Doppler shifts to be more convincing than those without them. Therefore, we can conclude by saying that the Doppler Effect can be considered a significant cue for perceiving a front-back movement of a source.

3.4.3. Spatial Impression

Bradley et al. (2000:651) point out that spatial impression is a significant acoustic factor when evaluating a good quality concert hall. Furthermore, Bradley et al. (1994: 2263) points out the two- dimensional aspect of spatial impression, namely *apparent source width* (ASW) and *listener envelopment* (LEV) that will be discussed in the following section. ASW and LEV correlate to distinct acoustic factors, which include *early lateral reflection energy* (leads to a

[42] Strauss, H. (1998) *Implementing Doppler shifts for virtual auditory environments.* In: *Audio Engineering Society Preprint*, 104th Convention, preprint no. 4687

sense of ASW) and *later-arriving lateral sound energy* (leads to a sense of LEV) (Bradley et al. 2000: 1).

Apparent source width

ASW has been defined as (Iida et al. 1994: 1):

> "...the width of a sound image fused temporally and spatially with the direct sound image."

Blauert (1999: 347) defines the perception of ASW by using the term *auditory spaciousness* and explains that it is used:

> "...to describe the characteristic spatial spreading of auditory events, so that they fill a larger amount of space than is defined by the visual contours of an ensemble of sound sources."

An understanding of ASW originated in the field of concert hall acoustics, where its basic psychoacoustic principles were established (Neher 2004: 42). More specifically, the perception of the perceived source width seems to be dependant on lateral reflected sound energy. This is only true if the reflected sound reaches the listener within 80 ms of the direct sound. In turn, this influences the degree of interaural cross-correlation that takes place. The latter concept is an interaural process that measures the similarity between signals at both ears (Rumsey 2001: 37). The fact that the frequency spectrum of reflections effects ASW hearing has also been established in concert hall acoustics.

Therefore, according to ASW, a sound source can fill a larger space than that suggested by its physical boundaries (Neher 2004: 42). However, within the context of reproduced sound, research[43] has shown that listeners find it difficult to evaluate the left and right boundaries of such a source. This can be explained by the contribution of early reflections to the "fuzziness"

[43] See Beranek's publication in the *Acoustical Society of America*, namely "Concert and Opera Halls: How They Sound" (1996).

and localization difficulties and inaccuracies of sounds (Rumsey 2001: 36). For these reasons, listeners often find terms like 'source focus' or 'diffuseness' more applicable (Neher 2004: 42).

Listener envelopment

Bradley et al. (2000: 651) defines LEV as:

> "...the sense of being surrounded by a diffuse array of sound images that are not associated with particular source locations."

Similar to ASW, our understanding of LEV also derives from research done in the field of concert hall acoustics. Experiments done by Bradley et al. (1994: 2263), show that the sense of LEV is perceived because of reflected energy reaching the listener at least 80 ms after the DR[44]. Furthermore, they point out that reverberation is only relevant when lateral reflections reach the listener. Bradley et al. (2000: 655) also points out that an increase in early sound energy tends to decrease the perceived LEV.

Current research, examining LEV within the context of multi-channel sound reproduction (using the ITU specification), has resulted in a better understanding. The following section presents some of the findings of experiments done by Soulodre et al.,[45] Griesinger[46] and Rumsey,[47] as summarized by Neher (2004: 45-48).

- Soulodre et al. (2003): The perception of LEV can be influenced by changing the relative level or the angular distribution of late energy and, but to a lesser extent, the reverberation

[44] "Direct-to-reverberant sound ratio" (Neher 2004).

[45] Soulodre, G. A., Lavoie, M. C. and Norcoss, S.G. (2003) Objective measures of listener envelopment in multichannel surround systems, in: Journal of the Audio Engineering Society, vol. 51, no. 9. pp. 826-840.

[46] Griesinger, D. (2000) *The theory and practice of perceptual modeling – How to use electronic reverberation to add depth and envelopment without reducing clarity.* 21st Tonmeistertagung, Hannover, Germany. pp. 24-27.

[47] *Rumsey, F. (2002): 'Spatial quality evaluation for reproduced sound: Terminology, meaning, and a scene-based paradigm', Journal of the Audio Engineering Society, vol. 50, no. 9, pp. 651-666*

time[48] (RT) of the sound field. Furthermore, their results show that a higher playback level in the total mix, give rise to a heightened experience of LEV and vice versa.

- Griesinger (2000): Reverberant energy with a delay of 150 ms associated with the direct sound will result in the best sensation of LEV.

- Rumsey (2002): Because of the lack in internationally agreed terminology regarding LEV, he suggested to distinguish between *source-*, *ensemble-* and *environmental envelopment*[49] where LEV would most accurately be described as *environmental envelopment.*

To conclude, Neher (2004: 49) provides us with a summary of the relationships between spatial dimensions and their physical correlates. It should be noted that not all of the characteristics apply to the ITU- BS 775-1 loudspeaker set-up that is used in this dissertation.

[48] White (2002: 282) defines this term as follows: "The time of reverberation is defined as the time it takes for the SOUND PRESSURE LEVEL to decay to one-millionth of its former value. This is a 60-decibel reduction in level".
[49] Relevant in the context of source width perception (Neher 2004: 48).

Table 3: Reviewed spatial attributes, their physical correlates (relevant to this work) and respective references (Neher 2004: 49)

Spatial attribute	Physical parameter	Reference
Source distance	High-frequency content	[Blauert, 1997]
	Loudness	[Blauert, 1997] [Begault, 1987]
	D/R	[Nielsen, 1993]
	Relative distribution of early and late sound energy	[Michelsen & Rubak, 1997]
	Early (t < 50ms) lateral reflections	[Gerzon, 1992a] [Griesinger, 2000]
	Spatial and temporal distribution of early reflections	[Pellegrini, 2002]
	Cross-correlation coefficient (f < 250Hz)	[Martens, 1999]
	Doppler effect	[Strauss, 1998]
Source depth	Cross-correlation coefficient (f < 250Hz)	[Martens, 1999]
Ensemble depth	Early (t < 50ms) lateral reflections[15]	[Wöhr *et al.*, 1990] [Theile, 2001]
Externalisation	Decorrelated reverberant energy	[Kendall, 1995] [Begault *et al.*, 2000]
Source width	Absolute value of cross-correlation coefficient	[Martens, 1999]
	ITD fluctuations during sound segments with audible DS (as caused by asymmetrical, lateral reflections)	[Mason, 2002a]
Ensemble width	Source-specific interchannel time/level differences	[Rumsey, 2002]
Environment width	ITD fluctuations during reverberant sound segments (as caused by asymmetrical, lateral reflections)	[Mason, 2002a]
Source envelopment	Direct sound of single (wide) source; Interaction of continuous sound with early reflections	[Rumsey, 2002]
Ensemble envelopment	Direct sound of sources distributed around listener	[Rumsey, 2002]
Environmental envelopment	Lateral reflections with minimum delays of 45ms (125Hz octave band) to 160ms (8kHz octave band)	[Soulodre *et al.*, 2003]

3.5. The multidimensional nature of spatial quality

Within the framework of the present dissertation, it should be noted that an increased interest in multi-channel audio in the recent past has resulted in intensified research, aimed at breaking down spatial quality into its discrete components (Neher 2004: 12). In order to better place in context the multidimensional nature of spatial quality, it will be useful to examine briefly the main findings of two important research projects as discussed by Neher (2004: 12-16).

3.5.1. Berg and Rumsey

Berg and Rumsey (1999a; 1999b; 2000a; 2000b; 2002)[50] undertook to identify the individually perceivable components of spatially reproduced sound, as well as the relative weights. Table 4 provides a summary of their findings.

Table 4: Spatial attributes obtain from Berg and Rumsey [2002] (Neher 2004: 13)

Spatial attribute	**Definition**
Naturalness	How similar to a natural (i.e. not reproduced) listening experience is the sound as a whole?
Presence	The experience of being in the same acoustical environment as the sound source
Ensemble width	The perceived width/broadness of the ensemble from its left to its right flank
Source width	The perceived width of an individual source/the angle occupied by this source. Excludes sounds coming from the environment.

50 Berg, J. and Rumsey, F. (1999a) *Spatial attribute identification and scaling by Repertory Grid Technique and other methods*. Proceedings of the AES 16th International Conference on Spatial Sound Reproduction, Rovaniemi, Finland, April 10-12. pp. 51-66.
Berg, J. and Rumsey, F. (1999b) *Identification of perceived spatial attributes of recordings by Repertory Grid Technique and other methods*, in: Audio Engineering Society Preprint, 106th Convention, preprint no. 4924.
Berg, J. and Rumsey, F. (2000a) *In search of the spatial dimensions of reproduced sound: Verbal Protocol Analysis and Cluster Analysis of scaled verbal descriptors*, in: Audio Engineering Society Preprint. 108th Convention, preprint no. 5139.
Berg, J. and Rumsey, F. (2000b) *Correlation between emotive, descriptive and naturalness attributes in subjective data relating to spatial sound reproduction*, in: Audio Engineering Society Preprint. 109th Convention, preprint no. 5206.
Berg, J. and Rumsey, F. (2002) *Validity of selected spatial attributes in the evaluation of 5-channel microphone techniques*, in: Audio Engineering Society Preprint , no. 5593, 112th Convention.

Localization	How easy is it to perceive a distinct location of a source? How easy is it to pinpoint the direction of a source?
Source distance	The perceived distance from the listener to the source.
Source envelopment	The extent to which the source envelops/surrounds the listener. Excludes sounds coming from the environment.
Room width	The width of/angle occupied by a sound source's reflections in the room.
Room size	In cases where a room/hall is perceived, this denotes the relative size of that room.
Room envelopment	The extent to which a sound source's reflections envelop/surround the listener. The feeling of being surrounded by reflected sound.

In light of its rank in their findings, "naturalness" also appeared in the above table. Other qualities inevitably contribute to "naturalness" as a sound appears to emanate from a matched source if it is perceived to be located away from the listener, and is spread out. Collected data were submitted to "correlation analysis" and "factor analysis" processes in order to investigate their relationship. From this, 'source width-ensemble width', 'source width-localization' (negative correlation) and 'source width-source envelopment' dualities were identified.

It can also be observed that Table 4 is devoted primarily to portraying auditory phenomena such as "source width", "source envelopment" and "room size". Finally, although "localization" did not stand in close association with "naturalness" and "positive sensations", it could be argued that the findings in Table 4 were actually to be expected.

3.5.2. Zacharov and Koivuniemi

Zacharov and Koivuniemi[51] also investigated the impact and description of auditory spaciousness. Table 5 summarizes the outcome of their extensive experimentation in the form of eight *spatial attributes*.

[51] Zacharov, N. and Koivuniemi, K. (2001a) Unravelling the perception of spatial sound reproduction:Techniques and experimental design. Proceedings of the AES 19th International Conference on Surround Sound, Schloss Elmau, Germany, June 21-24. pp. 272-286.
Zacharov, N. and Koivuniemi, K. (2001b) Audio descriptive analysis and mapping of spatial sound displays. Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-Aug. 1.
Zacharov, N. and Koivuniemi, K. (2001c) *Unravelling the perception of spatial sound reproduction: Analysis and external preference mapping*, in: *Audio Engineering Society Preprint, 111th Convention,* preprint no. 5423.

Table 5: Spatial attributes as derived by Zacharov en Koivuniemi (Neher, 2004: 15)

Spatial attribute	**Definition**
Sense of direction	How easy is it to discriminate sound source locations? Can several sources be distinguished in terms of their directions?
Sense of depth	Can several sound sources be discriminated in terms of their distances?
Sense of space	Is there a strong sensation of being in the recording space?
Sense of movement	Does the sound source appear to move in space?
Penetration	Does the sound seem to originate very close to or even inside the head?
Distance to events	Does the sound event appear to be at a certain distance from the listener?
Broadness	How wide is a sound event perceived to be?
Naturalness	How well does the perceived sound event conform to what is considered to be realistic?

In the final analysis it can be observed that, although the descriptive terms used in Table 4 and Table 5 do differ slightly, their findings support and reinforce one another.

3.5.3. Elevation localization

A basic understanding of elevation localization can provide a platform for the extension of *5.1 Surround Sound* towards the median plane. It should be noted that, strictly speaking, height perception falls outside of the scope of this dissertation. Firstly, Blauert points out that directional hearing in the median plane[52] differs from that in the horizontal plane. He motivates this by the following statement (1999: 44):

> "When the sound source is in the median plane, the signals at both ears are identical to a first approximation; interaural signal differences are therefore rarely available to aid in interpretation of the signals."

Variations in the direction of sound incidence in the median plane will affect monaural ear-signal attributes (Blauert 2003: 105). Since identical signals will reach the ear in the case of a symmetrical head, the effect of interaural cues[53] will thus be eliminated.

[52] A surface which vertically divide the head through its middle (Blauert, 2003)

[53] The interaural differences will equal zero (Blauert, 2003).

The theories of localization in the median plane

To explain the localization in the median plane, several theories have been discussed in the literature. The most important ones, as discussed by Blauert (1970b: 206) are: (1) the theory of timbre differences; (2) the theory of bone conduction, and (3) the theories of visual, tactile and vestibular cues.

1) This theory focuses on the linear distortions to which a sound wave is exposed when hitting the head of the listener. As stated earlier, these distortions can be explained by diffraction-, resonance- and shadow effects caused by the head, pinnae and ear-canal. These distortions are observed by the listener as timbre differences because of the correlation that exists between the spectrum of the sound wave and the angle of incidence. Blauert points out that the experimental results of Roffler and Butler[54] (1968a) reported the same spectral cues to be responsible for sound localization. Although it should be presumed that a reference-timbre is necessary for localization, experiments do not confirm this. The only necessary condition is for the sound signal to be of sufficient bandwidth, where prior knowledge of the sound signal is therefore not actually necessary.

2) The theory of bone conduction points out the role of the ear canal, ear-drum, the skull etc. as possible means by which sound can reach the inner ear. Furthermore, Blauert mentions the work of Sone (1968), which states that bone conduction plays a vital role in establishing the direction of sound.[55] Nevertheless, when signals fell under the perception level for bone conduction (e.g. masked by noise), experiments demonstrated that bone conduction did not constitute a necessary cue for localization in the median plane.

3) According to visual, tactile[56] and vestibular theories,[57] two bodily faculties are necessary to determine the direction of the sound source. Visual stimuli seem to play an important role in auditory localization.

[54] Butler, R.A. and Roffler, S.K. (1968a) Localization of tonal stimuli in the vertical plane, in: JASA, no.43.
[55] Bone conduction varies with the angle of incidence (Blauert 1970: 206).
[56] This stimulus is perceived by the pinna or by the skin of the neck, but does not appear to exert a significant influence on auditory localization (Blauert 1970: 206).

Localization and localization blur in the median plane

The following section provides a summary of Blauert's findings with regards to localization and localization blur in the median plane, as well as his references to other studies done on this topic (1999: 44-50). As previously mentioned, interaural signal differences will not occur when sound signals reach the auditory system from the median plane. As a result, the auditory system will not be able to rely on such differences as an aid in the interpretation of sound signals. This has been confirmed by a number of studies, most notably those done by Blauert,[58] Damaske et al.[59] and Wettschurek.[60] Blauert (1999: 44) has summarized the findings of these studies as follows:

> "The localization blur $\Delta(\delta = 0)_{min}$ for changes in the elevation angle δ of the sound source in the forward direction is approximately 17° for continuous speech by an unfamiliar person ... about 9° for continuous speech by a familiar person ... and 4° for white noise ..."

Blauert (1999: 44) expands upon this by recounting the findings of a study done in 1971 by Plenge & Brunchen.[61] He notes that they identified a tendency for auditory events to shift to the rear sector of the median plane, in cases where signals are very short and possessed impulse content. This tendency, however, appeared to be absent in cases where research subjects were exposed to the same signals as were used shortly afterwards in the actual experiment, thereby suggesting that previous acquaintance with a sound signal exerts an influence in directional hearing in the median plane.

In an experiment performed in 1968, Blauert (1999: 45) found that neither localization nor localization blur could be established in the median plane in the case of sound signals with a bandwidth of less than two thirds of an octave. This indicated that the direction of the auditory

---

[57] If head movements is taken into consideration, these cues will influence auditory localization (Blauert 1970: 206).

[58] Blauert, J. (1970b) *Ein Versuch zum Richtungshoren bei gleichzeitiger optischer Stimulation [An experiment in directional hearing with simultaneous optical stimulation]*, in: Acustica 23.

[59] Damaske, P. and Wagener, B. (1969) *Richtungshörversuche über einen nachgebildeten Kopf [Investigations of directional hearing using a dummy head]*, in: Acustica 21.

[60] Wettschurek, R. (1971) *Über Unterschiedsschwellen beim Richtungshören in der Medianebene. [On difference thresholds in connection with directional hearing in the median plane].* Gemeinschaftstagung für Akustik und Schwinggungstechnik, Berlin 1970, VDI-Verslag, Düsseldorf.

[61] Plenge, G. and Brunschen, G. (1971) *Signalkenntnis und Richtungsbestimmung in der Medianebene bei Sprache* [A priori knowledge of the signal when determining the direction of speech in the median plane]. Proceedings, 7th Int. Congr. on Acoustics, Budapest, 19 H 10.

event is determined only by the frequency of a given signal, and not by the direction of the sound source. In response to this, Blauert concluded that there is ultimately no clear correlation between the direction of the sound source and that of the corresponding auditory event.

Spatial hearing with a single sound source in the median plane

It is known that sounds emanating from the quotidian context, such as speech and music, will generally be localized by the auditory system in the actual direction of such respective sound sources. Blauert (1999: 104) affirms that this also applies to the median plane and notes that differences in the angles of incidence in the median plan will principally exert an influence on monaural ear-signal attributes. He makes the following statement in order to expand this point (Blauert 1999: 105):

> "When the sound signals to the ears pass the external ear they are always raised or attenuated in specific spectral regions depending on the direction of sound incidence."

This means that the auditory system draw a parallel between the various angles of incidence and their perceived directions, by strengthening the relevant class of directional bands in the perceived signal (Blauert 1999: 105).

3.5.4. Spatial hearing with multiple sound sources

In practice, cases of a single sound source in a completely non-reflective environment are almost never encountered. This particularly holds true for enclosed spaces. Blauert (2003: 105) provides some general observations for instances where multiple sound sources are involved. He does this by discussing an environment with two sound sources as basis for generalised speculation about cases with multiple sound sources. In the context of the present dissertation it will suffice to mention three phenomena that can take place in an instance where two sound sources are involved (Blauert 2003: 105):

1) Summing **localization**: This occurs when there is only one auditory event of which the position is dependent on both sound sources involved.

2) The **precedence effect:**[62] This refers to cases where only one auditory event – of which the position is determined by one sound source only – takes place.

3) **The echo effect**: This is produced when a second auditory event is heard in addition to a first auditory event. In such cases the position of the primary auditory event will be established according to the sound source that radiates first, with the position of the echo being determined by the second sound source.

To study these effects more closely, Blauert (2003: 107) conducted an experiment using two loudspeakers (set up in a standard, stereo arrangement) as sound sources. For this experiment one [and the same] signal was sent to both loudspeakers, however, with the addition of inter-loudspeaker delay. One of the important findings of this experiment was that, with a number of sound sources radiating correlated signals, the individual sound sources were not as a rule perceived as separate auditory events. In many such cases, only one sound source – specifically the one of which the position was first identified through the first wave reaching the listener – was perceived.

[62] This is sometimes alternately referred to as the "law of the first wave front" (Blauert 2003: 106).

Blauert (2003: 107) further found that an increase in the spatial expansion of an auditory event is experienced, when the correlation between the radiated signals decreases. When a number of uncorrelated signals were radiated, these were generally perceived as being a number of distinct auditory events. Blauert (2003: 107) provides the following example in order to illuminate this instance:

> "The latter situation is given, for example, when listening to an orchestra, where the individual instruments are to be seen as sources weakly correlated to each other. It is important for the listener to be able to distinguish the single instruments within the global auditory picture. The orchestra should sound transparent. It is also said that the individual instruments should not mask each other."

In conclusion, Blauert (2003: 107) outlines the conditions in which signals from various sources are least likely to mask each other. According to this, such a state will be achieved when the signals reaching both the listener's ears from separate sources, are sufficiently distinct with regards to their interaural features. This then implies that the occurrence of masking would be less, if there is a difference in azimuth between the two sound sources.

CHAPTER 4

TECHNICAL CONSIDERATIONS

The development and introduction of surround sound technology has had a pronounced influence on a range of aspects and qualities of the recording chain. The impact has resonated through the professional and so-called "home" domains. As is often the case, the introduction of new technological solutions initiates and stimulates debate, regarding formats and protocols, in short, standardization. Typically, bodies such as the EBU (European Broadcasting Union), ITU (International Telecommunication Union) and SMPTE (Society of Motion Picture and Television Engineers) set and add to those technical standards deemed necessary in order for a particular technical innovation to thrive.

Chapter 4 is derived from the Dolby 5.1-Channel Music Production Guidelines documentation[63]. (Dolby Laboratories Inc. 2003). The decision to proceed from this particular document is motivated by the fact that it has been compiled in order to act as a guideline applicable to home and professional installations. Furthermore, it incorporates the following three important specifications: ITU-R-Recommendation BS 775-1, BS 1116-1,[64] EBU 3276[65] and EBU 3276-E (2004: Revised).[66] This is important for the following reasons. The first provides the international standard for a "set of uniform discrete multi-channel systems" (Surround Sound Forum, 01.1-E-2002: 2) while the second and third describe and discuss the suitability of playback environments.

It should be noted that any discussion of especially the technical specifications regarding surround sound is highly problematic because of the fact that no single specification covering

[63] Included in the Appendixes section.
[64] ITU-R Recommendation BS.1116-1. (1997) *Methods for the subjective assessment of small impairments in audio systems including multichannel sound.* [on-line]. Availible: http://www.itu.int/rec/recommendation.asp?type=folders&lang=e&parent=R-REC-BS.1116. [Augustus 30, 2004].
[65] EBU Tech 3276 (1998) *Listening conditions for the assessment of sound programme material: monophonic and two-channel stereophonic.* [on-line]. Available: http://www.ebu.ch/en/technical/trev/trev_home.html [January 16, 2005].
[66] EBU Tech 3276-E, Supplement 1. (2004 Revised) *Multichannel sound.* [on-line]. Available: http://www.ebu.ch/CMSimages/fr/tec_doc_t3276_s1-2004_tcm7-12765.pdf [January 16, 2005]

all possible implementations has been drafted and agreed upon. For example, 5.1 surround sound designed for the home environment has a set of preconditions that differ from those applied to surround sound in cinema.

4.1. Control room design

4.1.1. Dimensions

The model control room would be symmetrical along the line between the centre speaker and the reference listening position. Furthermore, to prevent a build-up of low frequency standing waves[67], any parallel surfaces, such as walls, ceiling and floor, should be avoided. The suggested parameters are outlined in Table 6:

Table 6: Room Dimensions (Dolby Lab.Inc. 2003: [3-2]).

Parameter	**Units/Conditions**	**Value**
Room Floor Area		> 30 m^2 (320 ft^2)
Room Volume		< 300 m^3 (10, 500 ft^3)
Room Proportions	L = Length (larger dimension, irrespective of orientation) W = Width (shorter dimension, irrespective of orientation) H = Height	1.1 W/H ≤ L/H ≤ 4.5 W/H -4 with L/H < 3 and W/H < 3 No ratios of L, W, and H within ± 5 % of an integer value

[67] This effect is explained by White (2002: 317) as: "A phenomenon in room acoustics whereby a sound is reflected back and forth between two parallel surfaces, such as two sidewalls. The sound waves interfere with one another to produce a series of places where the sound pressure level (SPL) is high and another series of places between them where the SPL is very low."

4.1.2. Acoustics

Early reflections

Early reflections reaching the listener within 15 ms of the direct sound should ideally possess a level at least 10 dB below the level of the latter, for frequencies between 1 and 8 kHz.

Reverberation field

Reverberation time[68] is dependant on frequency where the nominal value (T_m) describes the average reverberation times measured in one third of an octave band from 200 Hz to 4 kHz. These times should not exceed the following range in either extremity: 0.2 s < T_m < 0.4 s. It should also be noted that T_m must increase in accordance with the size of the room. The formula in Table 7 can be applied to establish T_M:

Table 7: Reverberation values (Dolby Lab.Inc. 2003: [3-2]).

Parameter		Units/Conditions	Value
Reflected Sound	Early Reflections	0–15 ms (in region 1–8 kHz)	< –10 dB relative to direct sound
	Reverberation Time	$T_m[s]$ = nominal value in region of 200 Hz to 4 kHz V = listening room volume V_0 = reference room volume (100 m^3 [1075 ft^2])	$\approx 0.25(V/V_0)^{1/3}$

[68] Dolby Laboratories (2003: [3-2]) point out that it is important for the reverberation time to be consistent with the tolerance mask. This, however, only applies to frequencies between 63 Hz and 8 kHz.

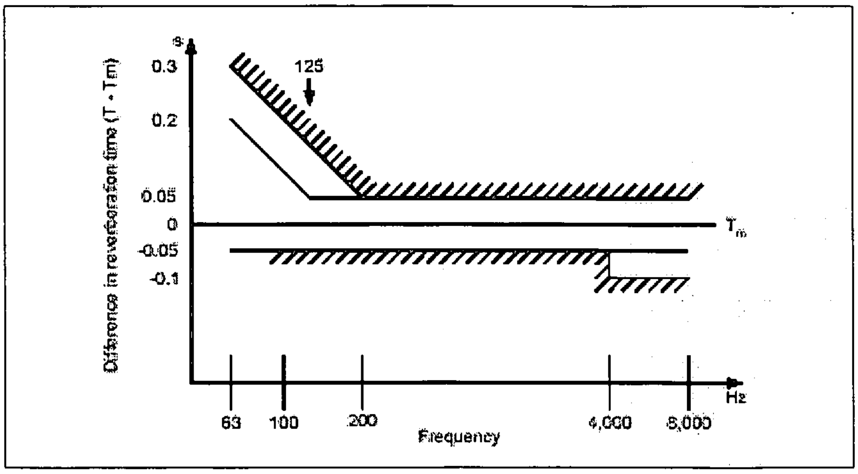


Figure 16: Demonstrates the tolerance levels for room reverberation time (Dolby Laboratories 2003: [3-3]).

Reflective- and absorbent surfaces

The presence of large flat surfaces possessing reflective characteristics must be minimized within the mixing environment. A smooth *RT delay time* within the range, indicated in Figure 16, can be obtained by combining diffuse reflectors and absorbing material in the said room.

Although not guaranteed, an analyzer[69] performing real-time spectrum analysis can accurately and rapidly measure room qualities in order for them to be tailored, if necessary. Equalization, utilizing graphic equalizers is applied only after exhausting options to arrive at a smooth *room response*.

Low frequency interaction between loudspeakers and echoic rooms

Room modes express the response of a room concerning low frequencies. Because of the presence of reflections, static patterns of pressure, which is directly related to the wavelength of sound, is generated which is spread around the room. Loudspeakers in a room are perceived due to there coupling with these room modes. Furthermore, the position of the speakers in

[69] This refers to a spectrum analyzer, consisting of a group of band pass filters. Each one of these filters has a constant bandwidth percentage, for example one octave, or one-third octave (White 2002: 275).

connection with the pressure patterns of these modes, control the amount of coupling that takes place room (Rumsey 2003: 123).

> If a loudspeaker is placed at a pressure minimum (a node) then it will couple weakly or not at all to the mode, whereas it will couple strongly when placed at a maximum (antinode). This has a substantial effect on the perceived frequency response of the loudspeaker.

Background noise

An NC rating of ≤ 10 is acceptable in the listening room. This value must be obtained at the reference position when the equipment is turned off. Control rooms containing auxiliary equipment such as video projectors, should demonstrate a value of ≤ NC 15 while the power is switched on.

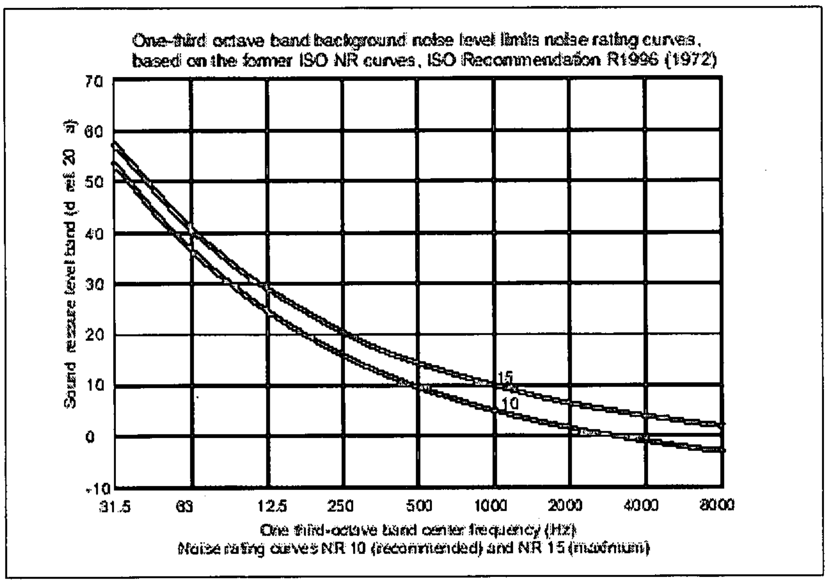


Figure 17: Noise Rating Curves (Dolby Lab.Inc. 2003: [3-4])

4.2. Monitoring

4.2.1. Reference monitors

Within the context of 5.1, a distinction can be made between types of loudspeakers[70] suitable specifically for the reproduction (in surround sound) of music in the cinema, as well as in the home environment. The former utilizes speaker arrays; the latter spreads an ambient soundscape through *dipole speaker* systems.

In both cases, surround loudspeakers often serve to distinctly place sound elements through *direct-firing/monopole speakers*.

These allow sufficient control over parameters, specifically "level", "timbre" and "image location". Ideally, they are dependant on the reflective qualities of walls, particular placement and a scattered surround environment, therefore not particularly suitable to surround sound monitoring.

And, in terms of the parameters listed in Table 7, matching speakers (L, C, R, Ls, Rs) ought to be relied upon.

---

[70] Attention must be devoted to the suitability of different types of surround loudspeakers (Dolby Lab. Inc. 2003: 3-5).

Table 7: Spesifications for the Reference Loudpseakers (Dolby Lab. Inc. 2003: [3-5]).

Parameter	Units/Conditions	Value
Amplitude/frequency response	20 Hz to 20 kHz* on axis (0°)	4 dB
	±10°	Deviation to 0°, 3 dB
	Horizontal ±30°	Deviation to 0°, 4 dB
Difference between speakers	In the range >250 Hz to 2 kHz	.5 dB
4 Directivity Index	250 Hz to 16 kHz	8 dB ±2 dB
Nonlinear distortion attenuation (SPL = 96 dB)	<100 Hz	–30 dB (=3%)
	>100 Hz	–40 dB (=1%)
5 Transient fidelity Decay time t_s, for reduction to a level of 1/*e*	t_s[s]	<5/*f* [Hz] (preferably 2.5/*f*)
6 Time delay	δt	≤10 μs
7 System dynamic range Maximum operating level (per IEC 60268, §17.2, referred to 1 m distance)	$L_{eff\ max}$	>112 dB (at IEC 60268 program simulation noise or special condition)
Noise level	L_{noise}	≤10 dBA

* 20 kHz is a minimum value. Some delivery formats contain content up to 96 kHz. Choice of speakers may depend on the production format in use.

Although a particular speaker system may meet the requirements of Table 7, assessment, implying pre-selection, may still be necessary. The following parameters must be adhered to (Dolby Lab Inc. 2003: [3-6]):

> ...the frequency response curve over the range 20 Hz to 20 kHz, measured in one-third octave bands using pink noise on the main axis (directional angle = 0°), should preferably fall within a tolerance band of 4 dB. Frequency response curves measured at directional angles of ±10° should not differ from the main axis frequency response by more than 3 dB and at directional angles ±30 ° (in the horizontal plane only) by more than 4 dB. The frequency response of different loudspeakers should be matched. The differences should preferably not exceed the value of 1.0 dB in the frequency range of at least 250 Hz to 2 kHz. The amplifier/speaker system should be capable of reproducing 120 dB without significant distortion.

4.2.2. Subwoofers

Ideally, subwoofer frequency response[71] must be flat: ±3 dB between 20 Hz and 120 Hz. If, in addition, the *bass management*[72] *crossover frequency* conforms to the parameters provided in Table 8, the subwoofer can be placed anywhere in the room.

Table 8: Reference Subwoofer Specification (Dolby Lab. Inc. 2003: [3-6]).

Parameter	Value
Crossover frequency	80 Hz
Out-of-band harmonic distortion levels	≤–50 dB (0.3%)
Filter order	Fourth

4.3. Reference positions

The position of the head of the listener is used as reference point regarding the distance and aiming angles for speakers in a 5.1 arrangement. The following parameters define this position:

At the centre of the console; Equidistant from the sidewalls; Directly above the rear edge (arm rest) of the mixing console; 1.2 meters (≈ 48 inches) off the floor.

4.4. Speaker placement

Speaker placement should be approached in the Figure below. These specifications are taken from the ITU-Recommendation BS 775-1 (ITU-R 1994), which outlines the international standard for a hierarchy of uniform discrete multi-channel systems. The ultimate aim of speaker placement is a multi-channel monitoring system that accurately images and expedites

[71] Loudspeakers will demonstrate a flat frequency response when "...its output is at the same level for all frequencies of interest, provided its input also has uniform amplitude over the same frequency range" (White 2002: 133).

[72] According to The Recording Academy's P&E Wing (2004: [3-9]) this term refers to "...the redirection of low frequencies from the main channels to the subwoofer, so that it reproduces all the low frequencies in a surround mix, including the dedicated ".1" LFE channel".

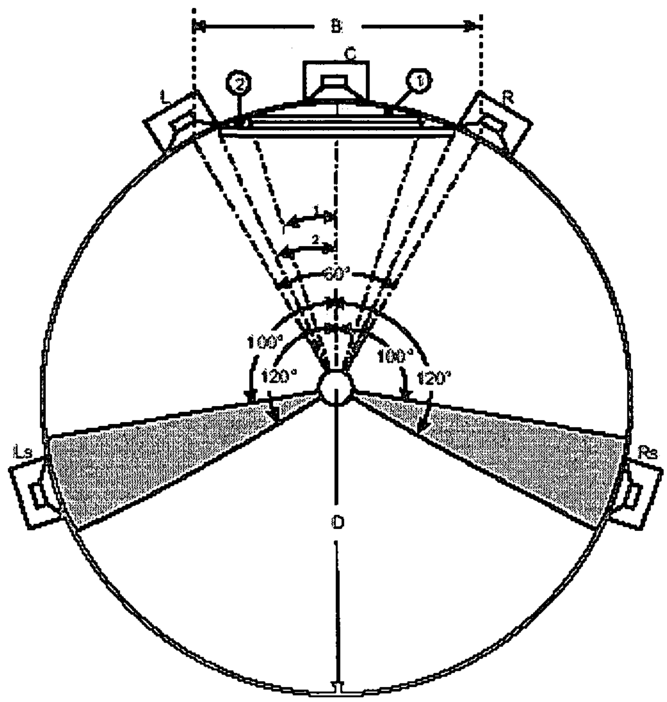
interchangeable critical listening judgements. However, any placement will still be subject to uncontrollable circumstances such as the physical dimensions or constraints of the facility and/or equipment.

4.4.1. Front-speaker placement

It is to be preferred that the three front-speakers (L, C, R) be placed at equal distances from the reference position. The centre channel has to be to the front of the reference position. The L and R should then be at an angle of approximately 30° in relation to the line formed between the centre speaker and the reference position. Additionally, the individual speakers should be at the same height as the reference position, that is 1.2 m, and front speakers should be vertically positioned at an angle of approximately 0° to each other. Lastly, the speakers all need to be the same distance from the reference point and should be directed towards this point.

4.4.2. Surround speaker placement

Ideally, the surround speakers should be placed the same distance from the reference position as the front speakers. The former should in addition be located at an angle of 110° with a permissible play of approximately 10° in either direction. As long as the surround speakers are kept equidistant from and directed towards the reference position, they may be elevated to a position not exceeding 15° above the reference position.



H: height of screen
B: loudspeaker base width
D: listening distance
Screen 1 HDTV: reference distance = 3H (2 1 = 33°)
Screen 2 = 3H (2 2 = 48°)

Loudspeaker	Horizontal Angle from Center (degrees)	Height (meters/feet)	Inclination (degrees)
C	0	1.2/4	0
L, R	30	1.2/4	0
Ls, Rs	100–120	≥1.2/4	0–15 down

Figure 18: Reference loudspeaker placement (ITU-R 1994: 3)

Table 9: Reference position (Dolby Lab. Inc. 2003: [3-6]).

Parameter	Units/Conditions	Value
Base Width	B [m]	2–3 m (6.5–10 ft)
Basis Angle	[°] referred to L/R	60°
Listening Distance	D [m]	= B

4.4.3. Subwoofer placement

Subwoofer placement can be a more cumbersome undertaking and the relative placement positions will differ between rooms. In general, a degree of experimentation will be required, especially in retrofitting an existing production room. An effective approach towards establishing the most favourable position is placing the subwoofer(s) close to the listening position and referencing according to material with considerable low-frequency content. Consequently, different [possible] locations should be experimented with in the room, in order to identify where the smoothest bass response can be obtained.

It must be kept in mind that the signal to the subwoofer will be band-limited anywhere in the range of 80 to 120 Hz and, in particular, that the ability to localize loudspeaker position will increase along with a rise in crossover frequency. Because the LFE-channel may possess content of up to 120 Hz, it is advisable to set the sub-crossover at 120 Hz or to bypass it. In the latter instance, the bass-management filter[73] will act as a roll-off[74] filter in reducing the output as the frequency is increased (White 1991: 286). On the other hand, greater flexibility with regards to the positioning of the subwoofer will be achieved if the crossover frequency for the bass-managed channels is kept lower, for example 80 Hz.

When placing a subwoofer in a symmetrically designed control room, it is advisable to be careful to try and position it in a [similarly] symmetrical position, for example along the centre line under the front speaker. In a symmetrical environment, this can result in symmetrical standing waves, which in turn causes an uneven room response. It may be then that a somewhat more asymmetrical placement will provide a more pleasing result, and a second subwoofer could be added in order to aid in smoothing out an uneven room response.

[73] The process of bass-management is discussed in greater detail in Section 4.5.3

[74] Paul White (1991: 277) defines this as "The rate at which a filter attenuates a signal once it has passed the filter cut-off point".

4.5. Sound level calibration

Calibration, implying referencing, plays an important role in surround sound. Although standards set by the AES, SMPTE and ITU have been determining factors, important procedures concerning, for example, the levels in audio mix-downs had earlier not been standardized. Clearly, standardization in terms of levels between media is preferable. Also, current [high-quality] equipment add to demands placed on technicians and others responsible for the sonic end-products from the studio, as challenges regarding monitoring, mix-down and mastering increase in complexity. It can be added that increased resolution affects dynamic range, frequency spectrum and loudness level and, acting in combination, these harbour destructive potential.

4.5.1. Alignment Signal Level

1 kHz Sine Wave Alignment Level

In the document – The Dolby 5.1 music production guidelines (2003: [4-2]) – on which this discussion rests, it is suggested to:

> "Use –20 dBFS as the 1 kHz sine wave alignment signal level, per current multi-channel DVD production standards. The RMS level of all test signals (see Section 5.6) will be at **Alignment Signal Level**, that is, –20 dB with respect to dB full-scale (FS) digital level in digital devices [3]. A 1 kHz sine wave at this level typically produces +4 dBu (= 1.23 Vrms relative to 0 VU [8]) from professional consoles."

For digital devices it is suggested that

> "...the alignment level must be a 1 kHz sine wave 20 dB below the maximum possible coding level of the particular digital system, irrespective of the total number of bits available."

Values listed in Table 10 can be used to align 16-bit audio systems.

Table 10: Digital codes for 1 kHz Sine Wave Alignment Levels (Dolby Inc. 2003: [4-2]).

Number of Bits	Audio Alignment Level	
	Negative Peaks	Positive Peaks
16	F333	0CCD

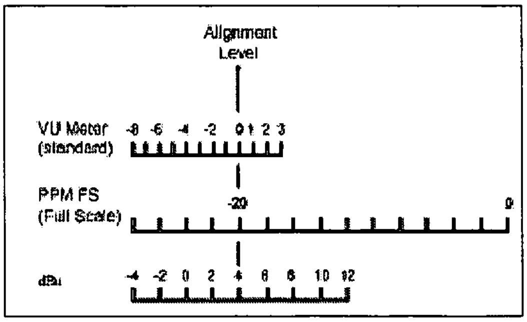


Figure 19: 1 kHz Sine Wave Alignment Level Metering (Dolby Inc. 2003: [4-2]).

Pink Noise Alignment Level

It is known that so-called "pink noise" has an unpredictable, random nature, which has an influence on the accuracy of the Peak Program Meter (PPM)[75] readings. It is suggested that the 1 kHz alignment setting is confirmed and that the alignment level associated with pink noise VU meter reading is then set to 0. The result is that a professional mixing desk would indicate +4 dBu (1.23 Vrms).

4.5.2. Loudspeaker Alignment Level

Bass-Managed System Speaker Alignment

The following suggestions are made to ensure that the loudspeakers demonstrate a flat frequency response:

(1) Supply the centre channel with wideband pink noise at alignment level.

[75] "...which is similar to a VU meter in appearance, but which responds to the peak, or maximum, level of the signal rather than the average level" (White 2003: 260).

(2) By using a Real Time Analyser (RTA)[76], the level of the centre speaker amp (see 'A' in Figure 20) must be set to 85 dBC

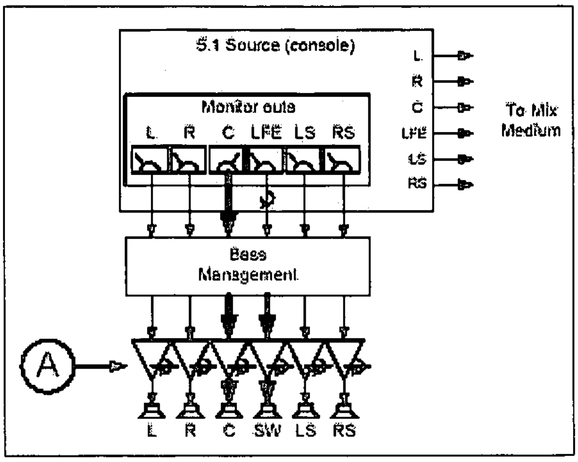


Figure 20: Bassmanaged Loudspeaker Alignment (Dolby Lab.Inc. 2003: [4-3])

(3) To obtain a flat frequency response on the RTA for the centre/sub combination, heighten the level of the subwoofer amp (point A in Figure 20).

It should be noted that point A (Figure 20) should not be changed after a flat frequency response from the bass-managed centre/sub was established. Continue with the remaining four channels (L, R, Ls, Rs), individually changing their relevant amps to a sub level of 85 dBC and a flat response.

LFE Alignment

In terms of bass-managed[77] and non-bass-managed systems, the LFE channel must be calibrated as follows:

(1) Supply the LFE channel with wideband pink noise at alignment level

[76] A special type of spectrum analyzer, which consists of a group of band pass filters, each having a constant percentage bandwidth, such as one octave or one-third octave. All the filter inputs are connected to the input signal, and the outputs of each filter are passed through a detector and thence to an indicating device such as a CRT (White 2002: 275).

[77] Align the LFE channel after changing the subwoofer level (Dolby Lab.Inc. 2003: [4-3]).

(2) The level of the console monitor feed (point B in Figure 21) must be altered to obtain an RTA reading of 10 dB in-band gain (average gain in the region 25 to 120 Hz) relative to the same bandwidth measured for the centre channel (Figure 22).

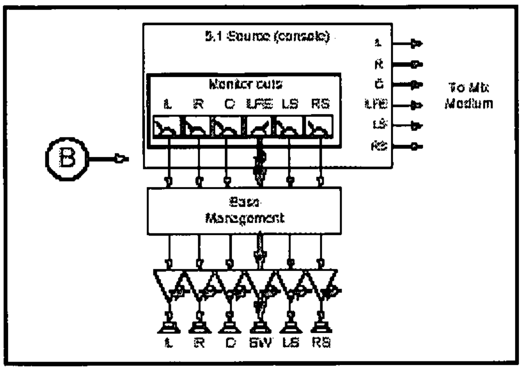


Figure 21: LFE Level Alignment (Dolby Lab.Inc. 2003: [4-3])

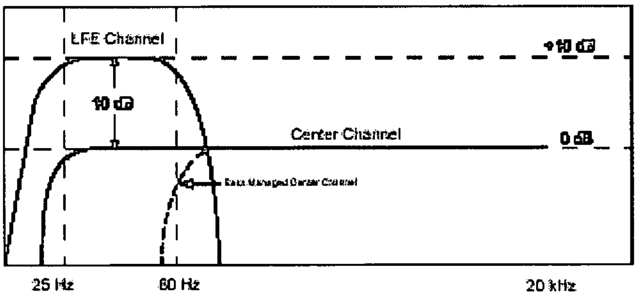


Figure 22: LFE RTA Display (Dolby Lab.Inc. 2003: [4-4])

The level of the LFE channel can also be set by using a SPL meter in the absence of an RTA meter. The SPL reading (C-weighted, slow) must demonstrate a reading of ± 4 to 5 dB above the main channels in the presence of pink noise. Also, the SPL meter gives a wideband measurement (not band-limited to 25–120 Hz) and therefore a lower value, as opposed to a band-limited RTA.

4.5.3. Reference Listening Level

The standard reference listening level, L_{ref}, seems to be lower than the loudspeaker alignment level and is established by:

$$Lref = 85 - 10 \log n \pm 0.25$$

...*n* represents the amount of reproduction channels in the total setup. The LFE channel is non-compulsory for 5.1-channel music mixing, thus *n* equals 5, making the individual channel level 78 dBC. Consequently, the combined five channels will operate at a reference level of 85 dBC.

4.6. Bass management

The present *section* is based on the document, *Listening Conditions and Reproduction Arrangements for Multichannel Stereophony* (Recommended Practice SSF – 01.1-E-2002), which was published in 2003 by the Surround Sound Forum.[78]

4.6.1. Low-frequency extension

In order to ensure the correct reproduction of bass signals in an environment such as a control room, *bass management* is applied. It is important to distinguish between a *low-frequency extension* (LFE) signal, routable through a LFE-channel, and the reproduction of low frequencies through subwoofers (AESTD1001.1.01-10: 5).

---

[78] "An interdisciplinary and supraregional working group that was founded in 1996 at the 19th Tonmeistertagung" [Convention of Sound Design of the Association of German Tonmeister, VDT] (SSF 2002).

The LFE channel is a discrete signal, created in the mix alongside the main channel. This signal is within a restricted frequency range (Dolby Lab, http://www.transtec.nl/download). The Dolby Digital encoder contains a "brick wall" filter[79] at 120 Hz; therefore the signal is bound to work only within the bottom two audible octaves. According to Dolby, limiting the signal to 80 Hz in the console, assure consistency in the delivery of the material.

Concerning the subwoofer signal, the decoder product produces this signal according to the use of the crossover filters chosen in bass management. A subwoofer signal contains bass from any channel or combination of channels. In the absence of the subwoofer, the bass (including the LFE channel if present) will usually be sent to the main stereo pair. In terms of music productions, the following can be said regarding the LFE channel (Dolby Lab 2006: http://www.transtec.nl/download):

> "Since the overall program level may be adjusted to allow for any proportion of bass to be perfectly rendered, the LFE channel might only be an advantage for something like the famous cannon shots in the 1812 Overture."

LFE signal and -channel

A special audio channel assigned a frequency range of roughly 20 to 80-120 Hz was introduced to cater for the extension of low frequency in film sound. ITU-R BS.775-1 stipulates that this applies to the film theatre to a larger degree than to the home theatre environment. Importantly, the absence of a LFE-channel has to be intelligently compensated for. For example, frequencies destined for the LFE-channel could be included in the principal channels, and gain differences should be intelligently compensated for.

Separate low-frequency loudspeakers (subwoofers) within the standard configuration

In addition to the L/C/R/LS/RS speakers, a subwoofer is used for the extension of the lower frequency range. An additional result would be that the loudspeakers' lower cut-off frequency would be approximately 80 Hz higher, if the volume (level) is lowered. A number of individual

[79] A low pass filter containing a very sharp cutoff slope (White 1999: 44).

subwoofers associated with individual channels or a single subwoofer assigned this function for all five main loudspeakers. Figure 23 demonstrates the incorporation of the above loudspeaker system and single subwoofer, through crossover circuits connection.

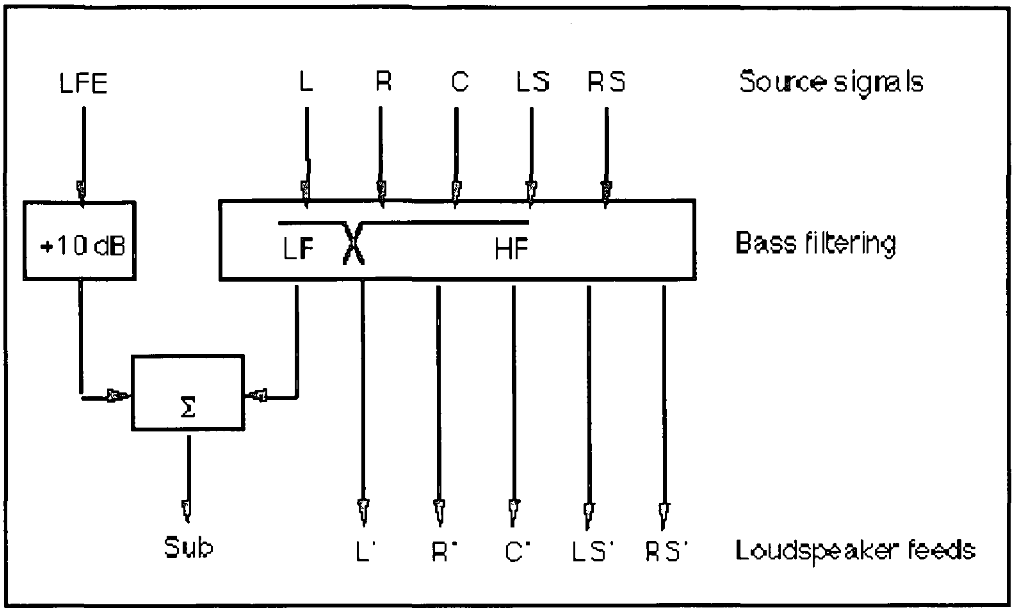


Figure 23: Derivation of combined subwoofer and LFE signals (Rumsey 2001: 92).

Two factors influence the auditory outcome: the relation between loudspeaker and listening position and the nonlinear characteristics of the subwoofer, such as distortions leading to inaccuracies in localization. Firm recommendations regarding subwoofer placement have not been promulgated although locating a single subwoofer in a corner results in an aesthetically pleasing low frequency response.

While optimum subwoofer placement remains a contentious matter, measurements have been published that suggest a corner placement in the case of a single subwoofer as the placement that will secure the most extended, smoothest low-frequency response. One of the negative outcomes could be a pronounced auditory effect as subwoofers not intended for corner placement interact with room modes.

The addition of artificial equalization could be called for in order to ensure a sought-after or preferred flat frequency response at the monitoring "sweet spot". Similarly, manipulation of phase and time information ensures the correct time relationship between subwoofer and the other speakers. In addition, a single subwoofer could be moved in front of a wall while the response of stereo subwoofers depends on placement, construction material and cross-over frequency. Placing stereo subwoofers centrally ensures a non-individual perception.

Multiple drivers convincingly recreate monaural low frequencies when two drivers generate a 90° (out-of-phase) signal.

For home theatre systems, the Recording Academy's Producers & Engineers Wing points to the fact that LFE reliance in badly designed systems may result in low frequency content loss.

Similarly, restrictions applying to so-called "matrixed encoded systems" dictate that sufficient bass frequency content ought to be assigned to the principal loudspeakers in order to eliminate so-called "double bass management".

Filtering the LFE-channel

Finally, it is suggested that for LFE frequencies between 80 and 120 Hz, a 24 db/octave low pass filter be utilized during monitoring. This negates the potential for phase cancellation resulting from the outcome of bass management as discussed in The Recording Academy's Producers & Engineering Wing Recommendations *For Surround Sound*

CHAPTER 5

MIXING IN 5.1 AND ELECTRONIC DANCE MUSIC (EDM)

5.1. Preface to mixing in 5.1 and EDM

It is a well established fact that the term *surround* [and all that it implies] has established itself, and has gained a secure foothold in the home, studio and electronic dance music (EDM) genre (Swenson 2002: online). The implementation of surround sound in the majority of venues still utilizing monophonic sound systems is imminent. The principal reason is the close link between technological change and electronic dance music that has existed. What is particularly exciting is the potential for futuristic implementations such as the "freely definable XY grid" suggested by Tozzoli (EQ-magazine 2006: online). In addition, the fact that initiatives such as a new "Remix Delivery Standard", headed by Brian Transeau (BT), had aimed to release documentation by 2005, provides further support.

It is generally accepted and often discussed informally, that in terms of projection and the experience of and interaction with three-dimensional sound, no boundaries ought to exist. The mind boggles at the potential for participants moving between parallel sound experiences, projecting particular instruments towards particular dance participants, etc.

Brian Transeau, BT, provides a glimpse at the current situation by stating that (M-audio 2006: online):

> "Clubs like Fabric in London already have 5.1. There are a lot of places that are putting 5.1 systems in, and it's amazing. Your first thought might be that you would need to stand in a central space. And it's not like that. I've heard it in a club environment; they did a 5.1 demo in New York at the Billboard conference dance music summit. You could walk anywhere in the room, and you could still tell the changes in spatial location."

Furthermore, his ongoing collaboration with venues such as *Ministry of Sound* has had an influence on the manner in which music for the dance market has been mixed. An important aspect to keep in mind here is the move away from mixing in a front/back-specific way as the audience is in all probability moving, not stationery. Compared to sound reproduction in the home theatre or a motor vehicle, rhythm as musical parameter and low frequency content of a signal demand special attention (The Recording Academy's Producers & Engineerings Wing 2004: [2-10]).

In the final analysis, it can be said that one of the greatest hurdles in a project of this nature is the lack of published reference sources, a problem that the researcher was unable to address in the literature study.

5.2. Definition of Electronic Dance Music (EDM)

The term Electronic Dance Music (EDM) is used to embody a wide range of music styles that has been produced in the last two decades. These styles include Techno[80], House[81], Drum & Bass[82] and Trance[83]. The most unique characteristic of EDM is the application of electronic technologies which include synthesizers, drum machines, sequencers and samplers (Butler 2003: 6-7). Furthermore, EDM can only be defined by investigating the way in which the music is produced, the market and the shared musical characteristics between the different styles in this category.

A new notation form, established into the realm of technology that dominates electronic dance music today, was that notes were replaced with recorded [samples of] notes (Miller 2006: online). This was initiated with the introduction of the *Page R* sequencer in 1982 and the *Fairlight CMI Series II*. These samples[84] were generated in real-time through a computer

---

[80] Developed out of electronic music made in Detroit in the mid-80's and developed into a number of subgenres "...including hardcore, ambient, and jungle" (Bogdanov *et al.* 2001: xiv).

[81] This genre was born in the early "warehouse" parties in Chicago and can be seen as a extension of the Disco era, with a "...heavy 'four on the floor' bass drum pattern... (Verderosa 2002: 35).

[82] Also known as "Jungle", this genre "...is the most rhythmically sophisticated of all forms of Techno. This style incorporates polyrhythm, fast break beats, and deep bass lines, and it is usually entirely instrumental" (Verderosa 2002: 29)

[83] "Trance music has an unusually strong relationship between music, the audience, and the party...Trance styles can be recognized by the geographical area that they come from, such as the Israeli sound (also frequently referred to as Goa) and the more dark European sounds of Break trance, Electro and Emotional Trance" (Verderosa 2002: 39).

[84] "Digital recorded fragments of audio" (Miller 2006: online).

program known as a sequencer. A sequencer makes it possible to change the sound and allow creative application through effects like equalization, volume, panning, cross-fading and digital effects processing. Furthermore, EDM manipulates the complexity of musical events in a more abstract form: the ones and zeroes of digital waveforms (binary code), streaming channels of MIDI data, effects processor parameters and mixing console settings. The main difference between electronic and popular music is summarized by Miller (2006: online) as follows:

> As a purely phonographic form, electronic dance music not only used fragments of recordings (both original and otherwise) as the basic building blocks for larger musical works, but also employed recorded media in its primary mode of 'authentic' live performance: the DJ set. In this sense, 'progressiveness' in electronic dance music genres can generally be attributed to the extent to which *recording* (sampling and sequencing) technology is used to *create* (record) and *recreate* (playback) sounds that are otherwise 'unplayable' by conventional live musicianship.

It is important to point out the different roles (not always distinct) that exist within the production of EDM namely that of the DJ, the dancer and the producer. Because EDM is merely used in this dissertation as a means to demonstrate 5.1, only the role of the producer and the techniques used will be focussed on. Finally, the fact that surround sound is not really applied within the "live" club scene, can be explained by the following statement (Miller 2006: online):

> "..what renders the disco, club or rave environments distinct from the traditional live music environment is also the fact that the sound surrounds you: the speakers are placed everywhere, often facing inward towards the dance floor, so that music appears to be omni-directional – that is, it is everywhere at once, and yet it does not seem to emanate from any one particular point (which is often the reason why dance club sound systems are run monaurally in order to prevent the 'phasing' or 'comb-filtering' effect of stereo signals as they collide into one another). In the traditional live music paradigm, the sound is unidirectional: creativity radiates from the musicians on stage, and therefore the sound system of a live music concert is designed to reinforce this notion."

5.3. Surround Sound Mixing Aesthetics

Firstly, this section begins with a definition of mixing in general, and the difference between stereo and surround mixing. William Moylan (2002: 299) defines mixing as follows:

> Creating an artistic blend of timbres and dynamic levels, and assigned spatial location, distance, and environment qualities (using the skills and concepts of a traditional composer or orchestrator), rehearsing and coordinating the precise changes that will occur to the sound source tracks (functioning similarly to a traditional conductor), and actually performing the changes that were determined and rehearsed above in real time (similar to a performance on a traditional musical instrument).

It is important to distinguish between stereo and surround sound mixing. Barry Rudolph (Mixguides 2003: online) points out the important differences between these two concepts:

> "For surround sound mixing, with higher resolution and six channels to play music over, there is less need to process individual tracks or the mix buses like stereo, so everything fits down a 2-channel pipe. As liberating as this might seem, this quantum leap in sonic and mixing options is not without caveats: Each individual track's sound quality; the cohesiveness and detail of the overall mix; reverb and delay setup and use, and the quality assurance of the final encoding process are critical for the best surround sound."

It can be said that [especially] within the context of EDM, principles in 5.1 mixing are fairly new. Therefore, views on 5.1 mixing are mainly based on opinions of producers and engineers in the industry.

5.4. Digital Signal Processors in *5.1* Mixing

The basic characteristics of a wave front – frequency, amplitude and time – are influenced by the application of signal processors[85]. According to their function, these signal processors are divided into three categories namely: *frequency* (equalizers, filters), *amplitude* (compressors, limiters, expanders, noise gates, de-essers) and *time processors* (delay, reverb) (Moylan 2002:

[85] In this dissertation the focus will mainly be on signal processors within the digital domain.

297). In the digital domain, these processors are applied to the mix in the form of plug-ins. An explanation of these processors will be discussed in the following section because of their relevance within the mixing stage.

5.4.1. Frequency processors (Equalizers, Filters)

Equalizers (EQ)

An equalizer can be defined as follows (White 1991: 120-121):

> "...alters or distorts the relative strength of certain frequency ranges of an audio signal. An equalizer can boost or attenuate a certain frequency band, but in common usage, equalize means to boost."

According to Snoman[86] (2004: 32), the main function of EQ is the prevention of possible frequency masking between instruments. The next section provides a summary of a few important facts pointed out by him in his book *The Dance Music Manual* (2004). It should be noted that merely instruments relevant to Electronic Dance Music, will be discussed.

(1) **Kick drum**: The kick consists out of two essential elements, namely the attack and low-frequency content which create its sound. The attack happens in the frequency range of 3-6 kHz and the low-end impact between 40 and 120 Hz. Certain techniques can be applied to alter and obtain the desired kick sound. If the kick sounds very low without an evident attack stage, a high Q[87] and large gain reduction should be applied. The latter will create a notch filter[88], wherein the frequency control can be used to sweep within the frequency range of 3-6 kHz. Furthermore, by placing a cut underneath the attack will result in increase of frequencies just above the cut-off. A prominent attack without a 'punch' can be attributed to a lack of low-frequency content. Boosting frequencies between 40-120 Hz, and applying a compressor could solve this problem, but the use of

---

86 Snoman is "A UK dance-music veteran who has been in the business for nearly 20 years. Rick Snoman has released various white labels and remixed more than 30 artists for the dance floor, not to mention led remixing and production seminars across the UK" (Remix 2004: Online).

87 "Q" resembles "quality factor", and can be defined as "...a measure of the sharpness of the resonant peak in the frequency response...the higher the Q, the higher and more well-defined the peak in the response" (White 2002: 269).

88 According to White (2002: 223) this filter rejects a narrow band of frequencies.

a different kick and therefore different timbre is suggested instead. Because a compressor holds down the transient[89] and therefore reduces the high frequency content, the perception of the mix could be changed severely.

(2) **Snare drum:** A snare drum has a lot of low frequency content and therefore makes the mix sound unclear. This should be avoided by utilizing a high-pass[90] filter that will remove all frequencies below 150 Hz. The 'snap' of most snares exists between 2-10 kHz, whereas the main body is situated between 400 Hz and 1 kHz.

(3) **Hi-Hats and Cymbals**: Low frequency information in hi-hats and cymbals plays no significant role towards the sound of these instruments and will only blur frequencies in the midrange. Therefore, a high-pass filter should be applied to remove frequencies under 300 Hz. These instruments are present in the frequency range of 1-6 kHz. The presence of these instruments lies between 1 and 6 kHz, while their brightness can reside as high as 8-12 kHz. A shelving filter can be used to boost all frequencies above 8 kHz that will result in more brightness. At the same time, it is suggested to *roll off*[91] all frequencies above 15 kHz, which will stop possible "hissing" from breaking through into the track. Furthermore, boosts in dB with a Q of about an octave, at 6 Hz, should add some presences in the sound quality of these instruments.

(4) **Toms and Congas**: These instruments contain frequencies under 100 Hz which can blur the low and low midrange of other instruments in the mix. Therefore, it is suggested to shelve off all frequencies under 200 Hz. Furthermore, the significance of these instruments can be heightened by applying a Q of half an octave between 300-800 Hz.

(5) **Bass:** Boosting frequencies between 60-80 Hz (with a thin Q), will make the bass sound more prominent in relation to the kick with regards to synthetic instruments. The use of a Q of half an octave whilst implementing boost, cut or sweeping between 120-180 Hz,

---

[89] This refers to an alteration in the amplitude (Campbell and Greated 1987: 12). Furthermore, a transient peak, describes the loudest piece in the attack or onset transient (Campbell and Greated 1987: 157).

[90] In this filter, a signal is passed above a certain frequency, namely the cut-off frequency. At the latter frequency, the response of the filter is 3 dB below the nominal response (White 1999: 160).

[91] A filter where a rise in frequency will result in the decrease of the filters output (White 1999: 286).

should add to a more prominent punch. Another problem with bass timbre is concerned with volume in relation to the kick. While the bass timbre may sound fine, there may not be enough volume for the bass to be prominent in the mix. By applying small amounts of controlled distortion, it should supply the bass with just enough volume to pull it to the front of the mix. By compressing the kick drum with a short attack stage, so that the transient is captured by the compressor, will also have optimum results.

(6) **Synthesizers/piano/guitars:** Finally, the remaining instruments mostly have their fundamental frequencies within the mid range. A general guide to the frequencies of most sounds that sit in the mid range along with the frequencies that contribute to the sound, can be found in Snoman (2004: 334).

5.4.2. Amplitude processors (compressors, limiters, expanders, noise gates, de-essers)

Compressors

The primary role of a compressor is to decrease the dynamic range of a performance (Snoman 2004: 92). According to Katz (2002: 308), this dynamic range can be defined as follows:

> "The range in decibels between the highest level which can be encoded and the lowest level which can be heard. Since this is a perceptual (or ear-based) determination, it is an approximate number. In a properly dithered system, available dynamic range can be greater than its measured signal-to-noise ratio."

During the recording process, the optimum signal must be obtained to avoid the artificial boosting of the gain. The latter will result in an increase in volume as well as background noise. Additionally, the difficulty lies within the huge dynamic range of vocals and 'real' instruments where the difference between clipping [92] and a nominal level can be as much as 20 dB.

[92] "A type of distortion that occurs when a signal is recorded too loud int he digital domein. A clipped waveform exhibits a crunchy or harsh sound, as the top of the waveform is removed from the signal" (Snoman 2004: 485).

In EDM, the action of the compressor is further applied for more creative possibilities. This can be explained by the following (Snoman 2004: 93):

> "Since the signals that exceed the threshold are reduced in gain, the parts that do not exceed the threshold aren't touched, so they remain at the same volume as they were before compression. In other words, the difference in volume between the loudest and quitest parts of the recording are reduced, which means that any compressed signals will become louder relative to the compressed parts. This effectively boosts the average signals level, which in turn not only allows yout o push the volume up further, but also makes it sound louder."

5.4.3. Parameters within compression

- ***Threshold***

This parameter determines the signal level (dB) where the compressor will be activated and consequently reduce the incoming signal. This signal level works in conjunction with a gain reduction meter to demonstrate the amount of compression affecting the incoming signal.

- ***Ratio***

This control is used to control the amount of gain reduction that takes place after the signal has passed the threshold. The ratio determines the dynamic range that will be affected by the compressor. It also identifies the difference between the incoming signals that exceeds the threshold, to that of the outgoing levels. An incoming signal that surpasses the threshold with a ratio of 4:1 shows that the incoming signal exceeds the threshold by 4 dB and the compressor will compress the signal that will result in a 1 dB increase at the output of the compressor.

- ***Attack/Release***

The attack parameter defines the time it takes for the compressor to reach the point where maximum gain is attained. The release parameter determines the time the compressor will wait before it stops processing, after the signal has dropped below the threshold.

- ***Soft/hard knee***

These two parameters prescribe the shape of the envelopes and therefore the characteristic of the compressor's behaviour when a signal moves towards the threshold. Hard knee compression can be explained as the instant reaction where the compressor squashes the signal when it exceeds the threshold. Furthermore, by continually measuring the incoming signal, soft knee[93] compression apply gain reduction steadily when the incoming signal approaches 3-14 dB towards the current threshold.

- ***Side-chaining***

By routing an audio signal into the side-chain inputs of a compressor, the envelope[94] of a sound can be used to control the action of the signal being sent through the normal inputs.

The practical application of compression

A ratio of 4:1 and a gain reduction meter with a reading of between -8 and -10 dB, can function as a starting point for how compression is applied. Subsequently, it is suggested that the attack is set to its fastest speed with a release time of ± 500 ms. Finally, instruments with a wide dynamic range, demands a higher ratio and lower threshold settings. The settings in the Table below may be used as a guideline; though naturally only representing starting points during mixing.

Table 11: Compression settings (Snoman 2004: 98).

Compression settings	Ratio	Attack parameter (ms)	Release parameter (ms)	Gain reduction	Knee
Starting settings	5:1 to 10:1	1-10	40-100	-5 to -15	Hard
Drum loop	5:1 to 10:1	1-10	40-100	-5 to -15	Hard
Bass	4:1 to 12:1	1-10	20 or auto	-6 to -13	Hard
Leads	2:1 to 8:1	3-10	40 or auto	-8 to -10	Hard
Vocals	2:1 to 7:1	1-7	50 or auto	-3 to -13	Soft
Brass instruments	4:1 to 10:1	1-7	30 or auto	-8 to -13	Hard
Electric guitars	8:1 to 10:1	2-7	50 or auto	-5 to -12	Hard
Acoustic guitars	5:1 to 9:1	5-20	40 or auto	-5 to -12	Hard

[93] Soft knees of **6-9 dB**, are considered to offer the most natural compression for instruments (Snoman 2003: 96).
[94] Commonly refer to as ADSR. The latter is short for "...attack decay sustain release, time constants associated with signals generated by electronic music SYNTHESIZERS (White 1999: 9).

The creative application of compression

The obvious use of the compressor plays an important role in creating EDM. Because the greatest energy within a drum loop is drawn from the kick drum, this can be creatively manipulated (2004: 99):

> "...if a compressor is inserted across this particular drum loop and the threshold is set just below the peak level of the loudest part, each consecutive kick will activate the compressor. If the release time is set short, the compressor will activate on the kick, and quickly release when the kick drops below the threshold. This result in rapid changes in volume, producing a pumping effect as the compressor activates and deactivates on each kick. "

This effect is known as '*gain pumping*' that is frequently used within EDM to extend the dynamic feeling of a mix. Furthermore, this effect can be developed by adding more instruments to the previous mentioned loop and compressing it again. Because the kick is still controlling the compressor, its appearance will result in a drop of volume concerning the other instruments. This results in a more powerful accentuation in the overall rhythm of the mix.

Gain pumping can also be applied across the entire mix, after the individual compression of instruments but must be applied with caution. The following settings should be used to create a powerful mix: an attack of 20-30 ms, a release of 250 ms as well as a low threshold and ratio to reduce the range by 2 dB. The tone colour of individual instruments can also be altered by for example, the use of heavy compression (a low threshold and high ratio) on an isolated snare drum. This results in a 'thwacks' sound where the snare's attack avoids compression but the decay is squashed, which brings it up to the attack's gain level.

Multiband compression

Katz (2002: 125) defines this type of compression as follows:

> "...a multiband processor splits information into two, three or more frequency bands, so that the compression action in the one band will not cause another band to be affected."

Furthermore, Katz (2002: 125) points out circumstances that is favourable for the use of multiband compression:

> "When there is heavy and somewhat isolated bass drum and bass, splitting the processing into two bands prevents the drumbeats from modulating the rest, or vice versa.
>
> When you want to let transients (percussive sounds) through while still punching the sustain of the sub accents or the continuous sounds. Transients contain more high frequency energy than continuous sounds, so splitting the processing into a low and a high band permits using gentler compression or no compression at high frequencies (e.g., higher threshold, lower ratio).
>
> When there is too much sibilance. Sibilance can be controlled by using selective compression in the 3 through 9 kHz range (the actual frequency has to be tuned by listening to the vocalist). Try a very fast attack and medium release and a narrow bandwidth for the active band."

Limiters

Limiters are used as a precaution to compress the possibility of transients that was not captured by the compressor. Instead of compressing a signal according to a ratio, they immediately stop the signal of exceeding the threshold and will therefore never go against the current threshold setting.

Noise Gates

This processor can attenuate or remove any signal that are below the threshold setting and is applied to remove unnecessary noise during a silent passage.

5.4.4. Time processors (Delay, reverb)

Reverberation

Reverberation plays a important role in the mix, since most samplers and synthesizers do not generate natural reverberations and therefore must by applied artificially to add depth to the mix. Because of the importance of this processor, some of its parameters will be discussed in the following section:

- ***Ratio/Mix***

This parameter determines the ratio between the direct sound and the amount of reverberation to which this sound is subjected.

- ***Pre-delay time***

The pre-delay refers to the time difference that exists between the direct sound and the first reflection that reaches our ears. This makes it possible to set the amount of time between the unaffected sound and the beginning of the first reflection.

- ***Early reflections***

This parameter allows you to determine the nature of the surface from which the sound is reflected. Because of the complexity of these reflections, they can only be found within high-end processors.

- ***Diffusion***

The extend to which early reflections are spread across the stereo image, is measured by this parameter. Furthermore, the position of the sound source determines the relation between the reflections and the amount of stereo width. A sound source in the far field will result in the spreading of the stereo width of the reverb, but there will be more reverberation than if it was upfront. in contrast, sound sources in the near field will result in monophonic reverberation.

- ***Reverb fade away time***

Snoman (2004: 111) defines this concept as follows:

> "The amount of time it takes for a reverb to fade away (after the original sound has stopped) is measured by how long it takes the sound pressure level to decay to one millionth of its original value. Since one-millionth equates to a 60 dB reduction, reverb decay time is often referred to as RT60 time."

- ***HF- and LF damping***

The further [away] reflections have to travel, the more high frequency content is absorbed by the air. By reducing these high frequency content, and therefore the decay time, the idea of a small enclose area with soft furnishings[95] can be created. Additionally, by increasing the decay time as well as the elimination of high frequency content can make the sound source seem further away. Finally, by enhancing the lower frequency damping, a large open space can be simulated.

- ***Chorus***

Phase cancellations are introduced by this effect, since it emulate the sound of two or more instruments playing at the same time. These phase cancellations can be explained by the fact that no two instrumentalists can play exactly in time with each other. By constantly changing, the amount of delay and amplitude to which the incoming signal is subjected, the chorus unit will achieve the desired effect. This modulation (controlled by three parameters), works on the same principle as the Low Frequency Oscillator (LFO) on a synthesizer.

- ***Phasers and flangers***

By utilizing an LFO, these effects either change the phase shifting of the phaser or the time delay of the flanger. Because the original and delayed signals are out of phase, a series of phase cancellations is therefore stimulated. As a result, phasers produce a range of harmonically related notches within the audio file, whereas flangers have a constantly different frequency because they use a time delay circuit.

- ***Digital Delay***

Digital delay is the most important effect in dance music and is often referred to as *Digital Delay Line* (DDL). The unit will allow you to delay the incoming audio signal by a predetermined time. The amount of delays produced by this unit can be controlled by increasing the feedback setting that allows you to produce more than one repeat from a single sound.

[95] Responsible for the absorption of high frequencies (Snoman 2004: 111).

5.5. Surround Mixing

Due to the lack of published documentation regarding Surround Mixing, this section is mainly based on *The Recommendations For Surround Sound Production* provided by The Recording Academy's Producers and Engineers Wing (2004: online).

5.5.1. Imaging and Panning

The main principle when mixing in stereo is to create a "phantom" centre by sending equal amount of signal into both the left and right channel. The resulting image that appears in the middle of the two speakers tends to move to one or the other speaker when the listener moves out of the *sweet spot.*[96] Furthermore, because of the presence of Comb filtering, the phantom centre can demonstrate a change in amplitude and frequency response. The two signals that reach our ears from these loudspeakers differ, and thus the resulting offsets can lead to possible cancellation of certain sound components. The latter situation can be avoided by the placement of a true centre channel in surround sound that will provide a steady signal with a consistent frequency response. Within the context of surround sound mixing, both the phantom image as well as this centre channel, is constructed by the sound engineer. An important parameter in this context is the *divergence control.* This control determine the relation between the amount of centre-panned signal send to the centre channel versus signals routed to the left and right channel in equal amounts.

Within Surround mixing, it's possible to create phantom images between both the left-centre as well as right-centre speaker that intensify localization. The latter placements are less effective because of the shorter distance in which these 'in-between' phantom images are created. The same principle can be used to create a phantom rear centre by sending the same amount of signal to the left and right speakers. It should be noted that phantom images between the front-

[96] White defines the sweet spot as "The listener position in front of stereophonic loudspeakers that provides the optimum listening conditions for tonal balance, stereo separation, etc. (2002: 324).

left and rear-left as well as front-right and rare-right are very unstable due to the presence of the Head Related Transfer Function (HRTF) caused by the size, shape and density of the human head. As previously discussed, HRTF is mathematical formulas that demonstrate how easy the listener can distinguish between the panning of sound sources between left, centre and right. Furthermore, HRTF's indicate that it is impossible for the listener to detect uniform panning of sound sources from the front to the rear speakers. The latter can be explained in terms of the difference in the ear's frequency response between sounds emanating from behind the ear versus directly in front.

It should be noted that because the size of the head is constant, despite the fact that different frequency components generate different wavelengths, localization is mainly frequency dependant. Furthermore, the localization of higher frequencies is more successful than low-frequencies, and therefore a single subwoofer is applied to work with very low-frequency sounds.

5.5.2. Use of the Centre- and Rear channels

Because of differences in speakers that exist in various playback systems, mixers of surround sound music try to avoid the overuse of this centre channel. Instruments send to this channel (bass, snare drum, kick drum and instrument solos) are therefore routed to the other two front speakers as well, which will in turn enhance localization within the "front wall". Creative psychoacoustic effects can be generated by placing an instrument in the centre channel as well as in one of the rare speakers. Most surround mixers avoid the use of reverberation in the centre channel, while others add a small amount of early reflections and reverb with short delays.

The rear channels are mainly used for ambient effects, but recently we see surround mixers placing a large amount of musical content within Ls and Rs, and therefore creating a bigger sound field. "Elements of Surprise" is also introduced to the mix by routing brief events to the rear channels, for example percussive accents or sound effects. The latter concept is also known as the "exit sign effect" (2004: [4-6]). Furthermore, an interesting phantom image can be created where the sound appears to move just behind the left or right shoulder of the listener. This is done by routing the signal to the centre channel and one (not both) of the rear speakers with careful balancing. This can be more effective by applying delays or phase offsets. Finally, multi-channel bus compression and equalization tools available in the market today, makes it possible to create a more coherent sound field.

5.6. Producers mixing in 5.1

Well-documented research in EDM concerned with 5.1 mixing, is scarce in the industry. The application of Surround Sound in EDM is a fairly knew concept, and is still in its experimental stage. Therefore, it was decided too just briefly point out a few opinions on 5.1 mixing, made in interviews, with the main focus on Brain Transeau. This world famous producer in EDM was chosen, based on his productions in 5.1, as well as his appointment as the head of a new 'Remix Delivery Standards' committee (A part of 'The Producers & Engineers Wing of The Recording Academy) in the last half of 2005 (Mix online 2005: online). Finally, it should be noted that a more subjective and creative approach will be used to understand 5.1 mixing. The researcher will accomplish this by means of a more practical approach to 5.1 mixing (See Chapter 6).

5.6.1. Brian Transeau (BT)

Brian Transeau – better known as BT – is a world famous Producer, Arranger and Programmer within the EDM context. Originally, BT became famous for his pioneering work in Trance[97] music with his avant garde use of sonic landscapes. BT also creates film scores, which include *Monster*, as well as *The Fast and the Furious* and *Under Suspicion*, and for the TV series

[97] Although difficult to define, this subgenre of EDM can be generalized as "...the only form of dance music that's constructed around glamorous melodies which are either incredibly vigorous, laidback or pretty much anywhere in between." (Snoman 2004: 219)

Tommy Lee Goes to College (Sunier 2006: Online). Furthermore, BT is one of few artists that experiments with 5.1 – a generally new concept within EDM. One of his latest projects, *This Binary Universe*, is a 5.1 channel DTS[98] surround that includes seven examples where his unique approach to electronic music and film scoring is combined. The primary quality of the listening experience is a cosmic, changing sonic 360 ° sound field around the listener.

Concerned with BT's view on 5.1, he made the following comment after the release of the Academy Award winning drama "Monster". In this Monster Score Album, BT makes use of multi-channel monitoring for the first time (M-audio 2006: online):

> "Clubs like Fabric, in London, already have 5.1. There are a lot of places that are putting 5.1 systems in, and it's amazing. Your first thought might be that you would need to stand in a central space. And it's not like that. I've heard it in a club environment; they did a 5.1 demo in New York at the Billboard conference dance music summit. You could walk anywhere in the room, and you could still tell the changes in spatial location.

Concerning mixing in general, BT suggest a few interesting techniques that can be applied. Shelving everything at 120 Hz, excluding your kick drum and bass line, will result in more space for everything in the mix (Interview done by Kissel 2005: 2). If you experience problems with a sound in a mid-wide notch, by removing 500 Hz should obtain a better result. Furthermore, BT demonstrates his use of side-band compression by referring to William Orbit, a groundbreaker in the field of compression (Kissel 2005: 3):

> "What they did... is, it's a straight eighth-note bass line, and it's moving in octaves, so it's like the low note first, and then the octave above on the "ands". Straight four-on-the-floor kick drum, right? So the bass note is emphasizing the kick drums. The sound is a really buzzy saw sound, like 16 Saw oscillators. Obviously an analog synth? But then underneath it they've also put a sine wave, which is like the old drum-and-bass trick, which is another trick that I poach, where you've got a high-pass filtered bass line that's really buzzy or distorted sound-like high-pass filter or even band-pass filter, when they sweep the band pass-but you've got a high-pass filter on it at, you know, 120 Hz, and then you get all the bass end in the sound from the sine wave.

[98] " An acronym for Digital Theatre Systems. It commonly refers to the company's glossy encoding scheme, which greatly reduces the size of digital audio data by discarding information that, in theory, cannot be perceived" (The Recording Academy's P & E Wing 2004: [G-3]).

> Anyway, they side-band compressed it to the kick drum. So you get this incredible, sucking sound on the "and"...

Reverberation, a very important feature within EDM where different reverb plug-ins have different results. BT's use of particular plug-ins (GRM EQ post reverb) is evident in the following statement (Alberts 2004: online).

> "I set up an aux send in Pro Tools and place the GRM EQ post-reverb. On the aux fader I'll then put something like a D-Verb and put the EQ through it, then automate the moves in time using the random curve pencil tool in Pro Tools. I'll put the Superslider on so that it's jumping around the different EQ curves in 16th notes and in tempo. "

The panning of instruments in 5.1 is very controversial subject and difficult to pin down. According to BT, the centre channel should be ignored in 5.1 (Interview by Mike Levine 2004: online), because of its destructive effect on the stereo image in the front. Nevertheless, BT employs the centre channel for creative effects:

> ...like having a bell in the center channel and then throwing a reverb with different predelays in the back speakers.

5.6.2. General Discussion on 5.1 mixing

Rich Tozzoli[99] (EQ-magazine 2006: online) states that many engineers will use the centre channel to "lock in" onto the image of a vocal, kick drum, snare drum, and bass guitar. Additionally, a certain amount of this will be mixed into the left and right stereo channels. Tozzoli uses the centre channel to create a strong mono image in regards to the centre channel. This is also true for the kick, snare, hats, and bass. Another technique used by Tozzoli for centre channel imaging, is delay on vocals in the surrounds were he suggests the following:

---

[99] Well known mixer, producer and composer (Crane 2003: online) and author of the book, "Pro Tools Surround Sound Mixing".

"Try creating a short (50 ms or under) stereo delay on a stereo aux send and pan it to the left and right surround speakers. Then, as you gradually increase the send from the mono vocal, the image will begin to envelope you. If you dampen the frequency of the delay' returns, the overall effect becomes less noticeable. When adding in the rest of the mix, it creates a nice sense of vocal space.

The above technique can be applied to any instrument. Furthermore, Tozzoli (EQ-magazine 2006: Online) suggest 10 surround sound production techniques:

1. Think surround in pre-production: Planning for a surround sound project is critical. First and foremost, make sure your client is ready for the extra time and budget needed. Try to get the best sounding room possible and check that the surround monitors work before you arrive.

2. You don't need multi-channel mics to record surround: Yes, higher quality mics are certainly better, but you can record with a bunch of SM-57s if you really need to. Try a quad array of decent omnis or cardioids around a drum kit, and pan each into the four corners of your mix. It'll sound like you're sitting on the kit.

3. Use the surround mics for stereo reverb: Take those two rear surround mics behind the kit in the above example, and use them in your stereo mix. They can provide "real" reverb, and add a nice depth to your overall sound.

4. Center channel – Don't rely on it: Some engineers don't use it at all. However, the consumer may feel cheated or that something is wrong if they don't hear anything in the center channel. Try using divergence or Center % (for Pro Tools users), which will then spread the signal across adjacent channels. With a vocal, for example, it will be heard in the Center, Left, and Right Front channels. The amount in each is your choice, but I always check my mixes with the center channel muted to see if it'll remain punchy. You never know if that consumer at home — the listener we're ultimately mixing for — has that channel in the wrong place, or worse, not connected at all.

5. Use bass management: Bass management, on the consumer level, allows those small satellite speakers to sound large — by routing the bass into the subwoofer where it can be reproduced properly. Usually in the 80–120Hz area (often selectable), the filter in the receiver cuts off any frequencies below that and sends the rest to the speakers. As surround producers, we should take this into consideration by checking our mixes as such. You can use hardware bass management systems that plug in before your studio monitors, or use software that works with your DAW — such as the Waves M360 Surround Manager. Whatever you choose, it's good to think like a consumer but mix like a professional.

6. Vocal delays in the surrounds: Try taking a mono or stereo delay, placing it into the Left/Right Surrounds and sending some front positioned vocal to it. This will help "pull" the vocal out of the front of the mix, creating additional depth and clarity. Try filtering the high frequencies in the delay to help reduce any sibilance and add warmth. Push the vocal delay send up to the point where it's audible, then back it down a pinch. When you mute the delay, you'll miss it — that means you've got it just right.

7. Use a piece of string to keep your speakers equal: Good Surround Monitoring is critical and ideally the signal from all five (or more) speakers should arrive at the same time. While the angles and heights of the speakers may change with music or post-oriented mixing, the distance from you to the tweeter should remain the same.

8. Capture your finals at the highest resolution possible: Whether the end product is DVD-Video, DVD-Audio, SACD, a video game or a TV broadcast, it's best to work with the highest resolution possible. Remember, it's easier to downconvert and lose a bit of fidelity than to upconvert, which can never sound better than the original source. Plan for the future now.

9. Prepare for the recall: Recalls by the artist, producer and/or record company are a daily occurrence in the production world. With surround sound, things can get even more complicated. Whether you mix "in the box" or with a full-blown console, document your work thoroughly. The variables of a surround sound mix are enormous, from multichannel outboard gear to subwoofer crossover levels — so be prepared for that total recall when it happens, which it will.

10. Go for it: Since there are "no rules", enjoy it. Take chances. Do something nobody has done before you. If it's wild enough, we'll all hear about it and you'll end up with a Grammy for Best Surround Mix. Hey, it could happen.

Also, Wilkens (2006: online) comments on the use of stereo processors in surround. Because of its clutter effect, the use of stereo reverbs must be restricted and a multi-channel reverb must be used instead:

"The only time it is okay to use a stereo reverb in 5.1 is when you are placing it into a stereo child buss within the main surround buss, i.e. you place the 'verb' into just Ls/Rs."

The use of reverb to create artificial depth to the mix is not of real importance in 5.1. Therefore, reverb is applied to a lesser extend in 5.1 mixes against stereo. The latter statement applies to equalization as well. Mainly in 24 bit, the application of 5.1 compressors is critical but must be use with diligence. According to Wilkens (2006: online), the primary factor for the perfect 5.1 mix, is the correct placement, and therefore 'panners' need to be properly learned.

CHAPTER 6

PRACTICAL: MIXING IN 5.1

6.1. Introduction to the practical

The following section consists of a demonstration of an EDM composition, *Bulgaria* (see the DVD accompanying this dissertation), which was composed and programmed by the researcher. The purpose of this section is to demonstrate aspects of the functionality of the 5.1 surround sound environment. The execution of this demonstration proceeds along the following lines: (1) attention is paid to hard- and software issues, specifically limitations and implementations; then (2) a discussion of the musical and technical facets of *Bulgaria* follows.

6.2. Studio set-up, hardware and software

The principal guideline determining the studio set-up was the ITU-R BS.775-1 recommendation for 5.1. However, the asymmetrical and particular shape of the control room made it difficult to follow the specifications with regards to the placement of the loudspeakers. Although the ITU-R BS.775-1 documents specify the use of full range monitors, the implementation thereof was not possible. This was mostly due to the current commercial and academic activity in the studio that made it impossible to employ full range monitors. Therefore, two Genelec 1031A Nearfield[100] speakers where implemented as the front-left (L) and front-right (R). As for the centre- (C), rear speakers (Ls; Rs) and subwoofer, Rocket-powered-"5", -"6" and -"10" speakers (from KRK Systems) were employed.

[100] "Studio speakers that are designed to be placed on top of a mixing console meter bridge in order to aid their low-frequency response" (P&W Wing 2003: [G-3]).

With regards to hard- and software, the control room is equipped with Digidesign's Pro Tools HD. The particular hardware version is the 192 I/O Pro Tools HD audio interface. This audio interface maintains up to 16 channels of analog and digital input and output (I/O). 192 I/O features a wide range of analog and digital I/O options to choose from, including 8 channels of high-definition I/O, 8 channels of AES/EBU, 8 channels of TDIF, 16 channels of ADAT, and 2 additional channels of AES/EBU or S/PDIF digital I/O (Pluginz.com 2006: on-line).

6.3. Pro Tools Session Setup (Surround and busses)

There are three primary standards in use for the track layout of the individual channels in a 5.1 mix. For the purpose of this thesis, use was made of the SMTE/ITU format for Dolby Digital (AC3)[101], with the surround channels routed to outputs 1-6 of the audio interface of Pro Tools HD (Digidesign 2005: 599). Pro Tools provides a default I/O setup path configuration when a new session is created according to your chosen Surround Sound format. The Surround mix I/O Settings files provide output and bus paths and consist of the following (Digidesign 2005: 612):

Default 5.1 Output Paths:

1. One 5.1 main output path, with sub-paths for center, left/right, LCR, and 5.1 (no LFE), and

2. One stereo main path with two mono subpaths

Default 5.1 Bus Paths

1. One 5.1 main bus path, with sub-paths for 5.0 (no LFE), left/right (stereo), LCR, and center (mono), and

2. One stereo main path with two mono subpaths.

[101] The other two formats are 1) Film (pro Tools default) or DTS with Pro Control (Digidesign 2005: 599).

A visual representation of the I/O Audio paths used for this project can be viewed in Figure 24.

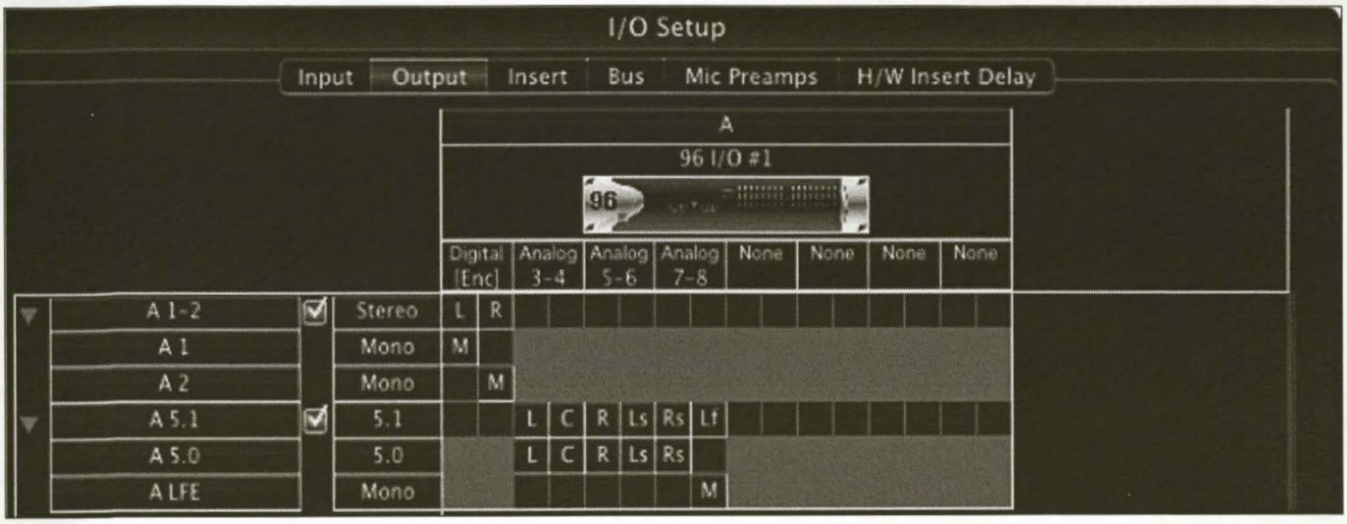


Figure 24: The I/O Setup (Output paths) for 5.1.

In essence, the I/O set-up of *Bulgaria* follows the same structure as the one represented above. It should be noted that the sub-paths of the Bus 5.1 are divided into a *Bus 5.0* and a *Bus LFE* (Auxiliary tracks). The latter method is used to ensure that if the subwoofer does not have its own Low Pass Filter, the cross-over frequency manipulation having been done with Pro Tools plug-ins. The low frequencies above 80 Hz were filtered by inserting a low pass filter plug-in (EQ3, 24 dB/octave at 80 Hz) on the Bus LFE channel. Furthermore, all frequencies below 80 Hz on Bus 5.0 were filtered with high pass filter plug-in (EQ3, 24 dB/octave at 80 Hz). This means that only frequencies below 80 Hz are sent to the subwoofer and only frequencies above 80 Hz are sent to the 5.0 speakers.

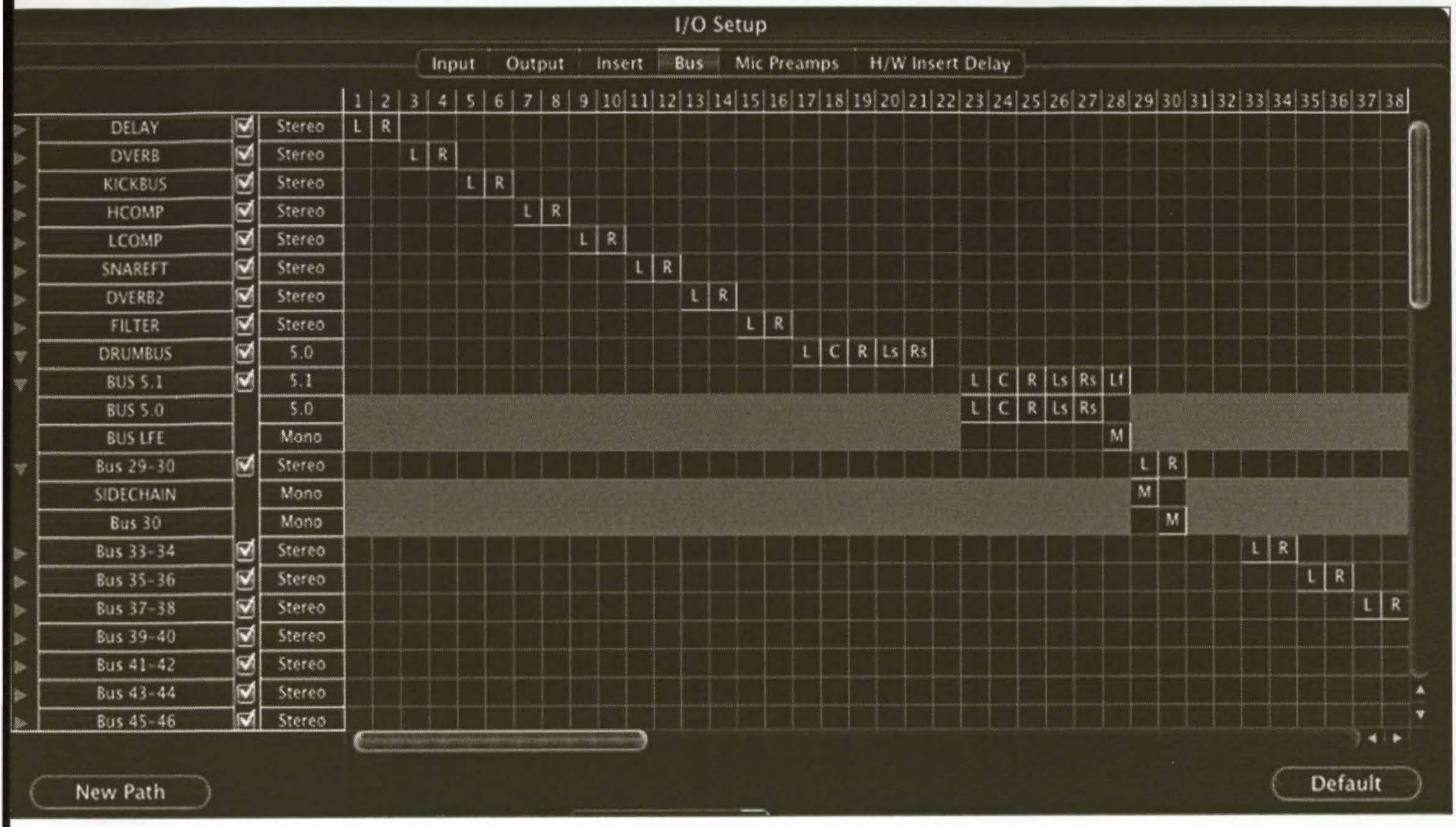


Figure 25: The I/O Setup (Bus paths) for 5.1.

6.4. Structure of Composition

Bulgaria can be categorised as belonging to the Progressive House[102] EDM genre. The composition is approximately 6:30 minutes in length (216 bars) with a tempo of 133 bpm. Although debatable, EDM compositions follow a definite structure with musical content and events consciously and constantly manipulated, added and removed. The ultimate aim in this is to arrive at a so-called audio "mix". In order to arrive at a convincing "mix" and "groove" or main rhythm section, tension-building and relaxing musical mechanisms are utilised. An

[102] "House music grew out of the post-disco dance club culture of the early '80s. After disco became popular, certain urban DJs...altered the music to make it less pop-oriented. The beat became more mechanical and the bass grooves became deeper, while elements of electronic synth-pop, Latin soul, dub reggae, rap, and jazz were grafted over the music's insistent, unvarying four-four beat." (Bogdanov 2001: xiv).

excellent way to represent the emotion and tension this song creates is to present it on a song map (see Table 11).

Table 12: Song Map

Introduction	1st break	1st Buildup	1st Climax	2nd Theme	2nd break	2nd Buildup	2nd Climax	Exit

6.4.1. Instruments and processors used:

A visual representation showing some of the instruments as they appear in the Pro Tools Edit Window, can be viewed in Figure 26 with a Mix Window representation in Figure 27.

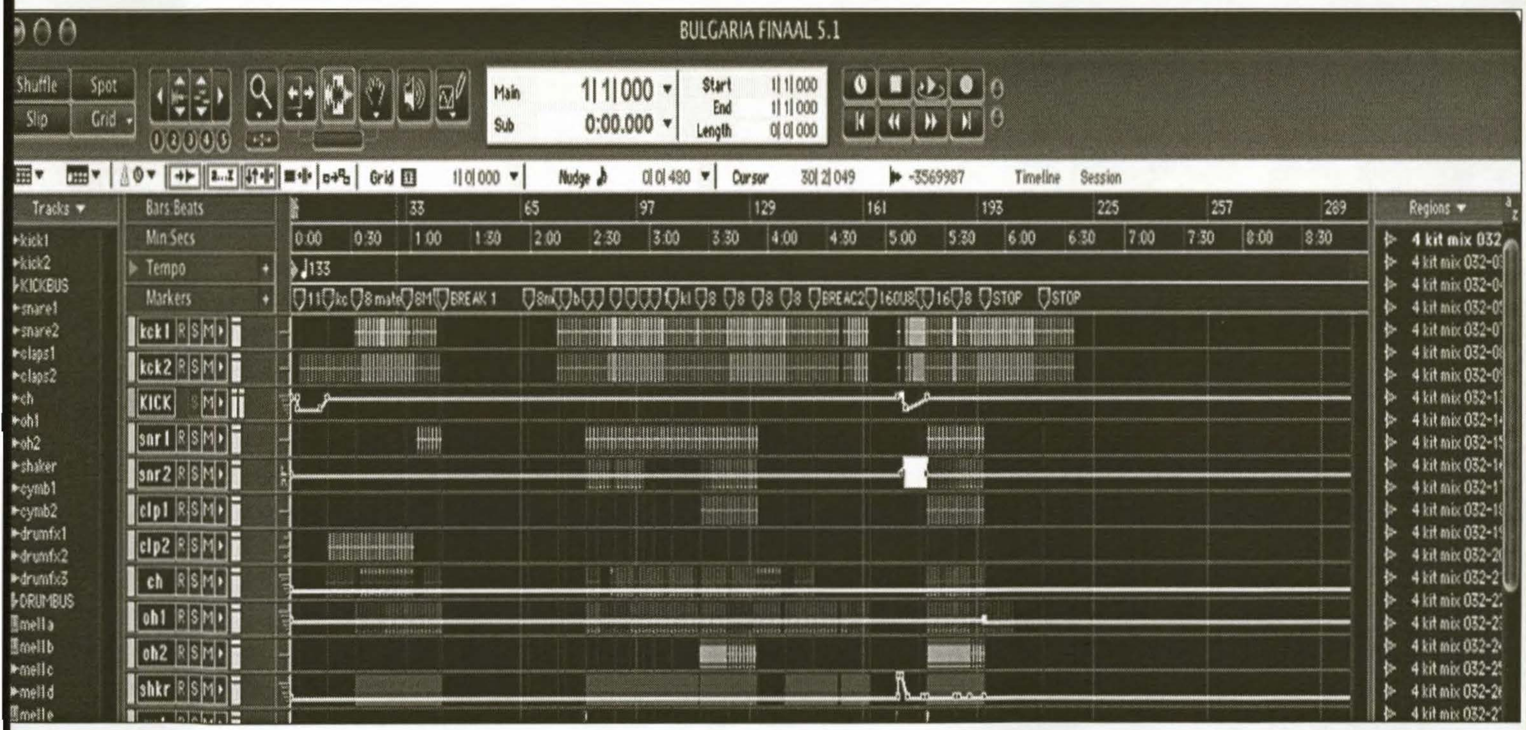


Figure 26: Edit Window

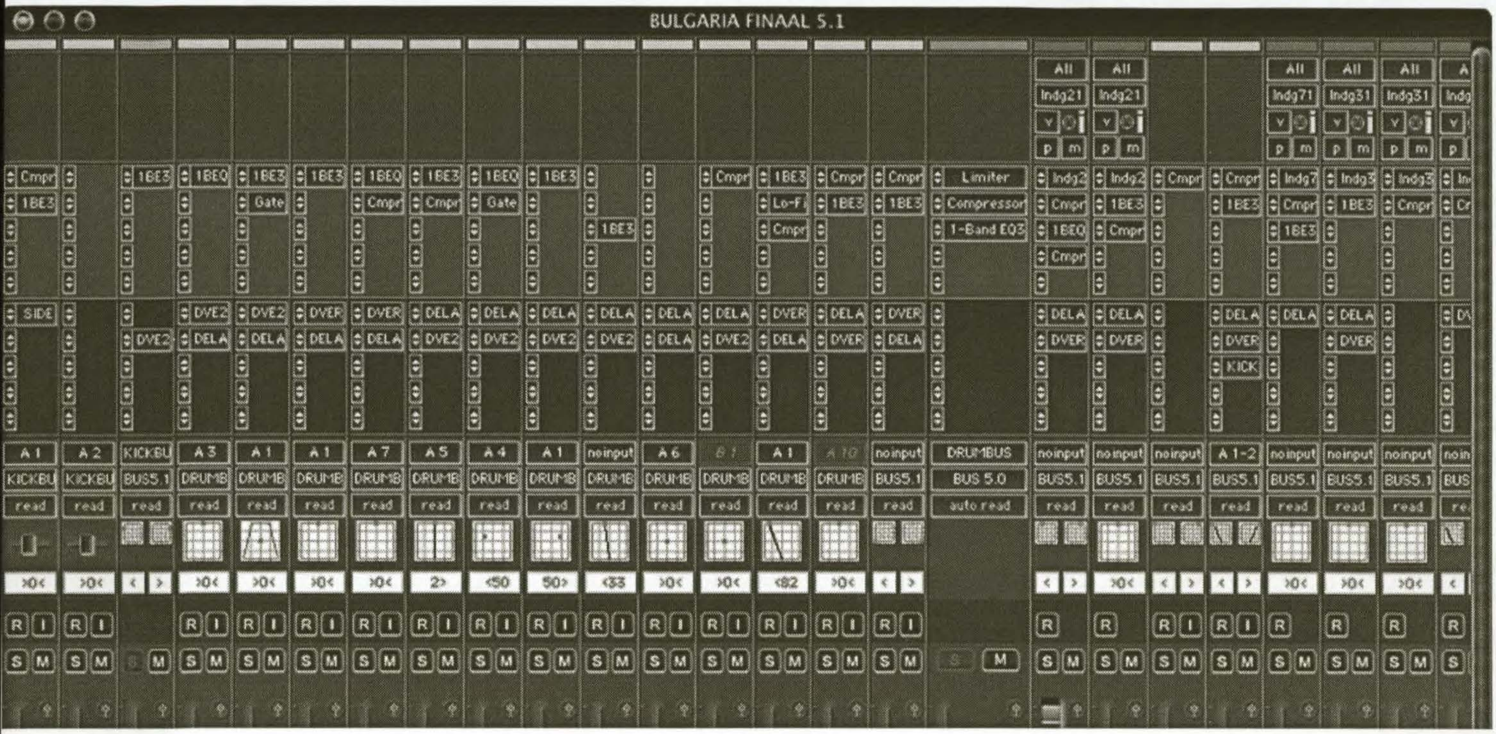


Figure 27: Mix Window

The main digital instrument or tone generator used in this composition is a software digital synthesizer, the *Access Virus Indigo TDM Plug-in.* It falls under the category software plug-ins for the Pro Tools environment. In Pro Tools version 7.0 the *Access Virus Indigo TDM Plug-in* allows (Performance Audio 2006: on-line):

> "...for building dense, layered textures with that distinctive Virus "virtual analog" sound. Virus Indigo is capable of powering entire compositions with up to 160 voices (on Pro Tools HD systems at 48 kHz; 128 voices are available with expanded Pro Tools|24 MIX systems)."

The two main instruments or instrumental sounds in an EDM composition are the kick-drum and open hi-hat sounds, together forming the core of the rhythm section. Essentially, the kick drum determines the main contour of an EDM composition in that it generates, defines and drives the tension in the introduction, build-up, climax and exit sections. Usually, the kick drum is responsible for defining a "groove" and plays an important role in the build-up leading to the climax. Interestingly, a composition can exist only with the kick drum and open hi-hats playing, a situation that creates a particular "feel" or "vibe". The absence of the kick drum usually indicates the beginning of a so-called "break" section.

The frequency content of the kick drum is manipulated as it is routed through a high-pass filter, which gradually "opens" over the duration of the first 8 bars. By side-chaining the kick drum through an audio compressor, a "pumping" effect is generated. The rest of the rhythm section (snare, closed hi hats, shaker and suspended cymbals) is introduced gradually until the entire percussion section is playing. This usually indicates that the *climax* of the song has been reached.

The application of *Drum-effects* adds creative ambience as well as a degree of tension to the mix. An example of these effects is the use of reverse cymbal, mostly to announce the beginning of a new section. Another example is the use of sixteenth note snare samples that is applied in the build-up. The latter is further manipulated by adding delays and filtering effects. Drum-effects also announce the beginning of a section. A shaker (16 beat) is employed to complete the rhythm section, adding to the movement of the song.

Two melodies (*mel1* and *mel2*) are systematically introduced throughout the course of the composition. Both these melodies consist of motives which, though also being introduced systematically and stepwise, gradually establish the two principal melodies. These motives are introduced sequentially with *mel1*'s motive 1a followed by 1b, 1c, 1d and finally 1e, and *mel2*'s motive 2a followed by 2b.

The melodies in this composition display certain qualities which are brought about through the use of automation. The latter is responsible for altering the attack and release times of melody notes as well as for altering of the *env-amt* or "envelope amount" parameter. Furthermore, by altering the oscillators, filters and modulation parameters on the virus, the timbres of the melodies are manipulated to create interesting effects. When sounding together, melodic notes are also used as chords in progressions that thicken the texture of the music. The complete *mel1* appears in the first break and the complete *mel2* in bar 96. In the main body, both of these melodies, as well as derivative developments, are polyphonically interleaved in a process of free imitation.

Another important element that is employed in this context is the *bass* and *bass-arpeggiator*. The bass adds to the rhythmic structure of the composition and emphasizes the harmonic root position of the mix.

The main signal processing procedures that were used amount to equalisation, reverberation, delay, gating and compression. Various permutations and the use of side-chaining combine to add to the spaciousness and multi-dimensionality of the final audio mixdown. Each track was subjected to compression in order to control the dynamic range of a particular instrument associated with that track. Finally, the following signal processors were added to the master fader: *ff d2+6 Band low-pass filter,* *C_4 Multiband Parametric Processor* and an *L1 Ultramaximizer*.

6.4.2. Panning of instruments in the mix

A visual representation of the output window can be seen in Figure 28. The surround positioning associated with a particular channel is represented graphically with a green dot in the middle of a square indicating that the sound is distributed to all loudspeakers. For example, if positioned hard-left-middle or hard-right-middle, it creates a phantom image between the front and rear speakers. The movement of the green dot and therefore the surround panning can be automated in order to generate dynamic movement in the audio mixdown.

Firstly, it is important that the LFE-fader should be on 0 dB with regards to all the tracks (audio- and instrument [MIDI]), unless the low frequency content of a specific track is not wanted on the subwoofer. This ensures that the signal is sent in its full capacity to the LFE channel. This LFE fader refers to the one presented in the multi-channel output window where the panning of instrument takes place.

The rear surround sound speakers are reserved for delay and reverberation whereas the main melodies are routed to the front speakers. These delays create a surround effect on their own and therefore are sufficient to create envelopment of the sound over the listener. Because of the timbre of the main melodies, it seemed unnecessary to pan and automate these instruments to the rear-surrounds because they seem to travel on their own, caused by the delays. Seeing that low frequencies do not travel, it was decided that only instruments with higher frequencies qualify for panning.

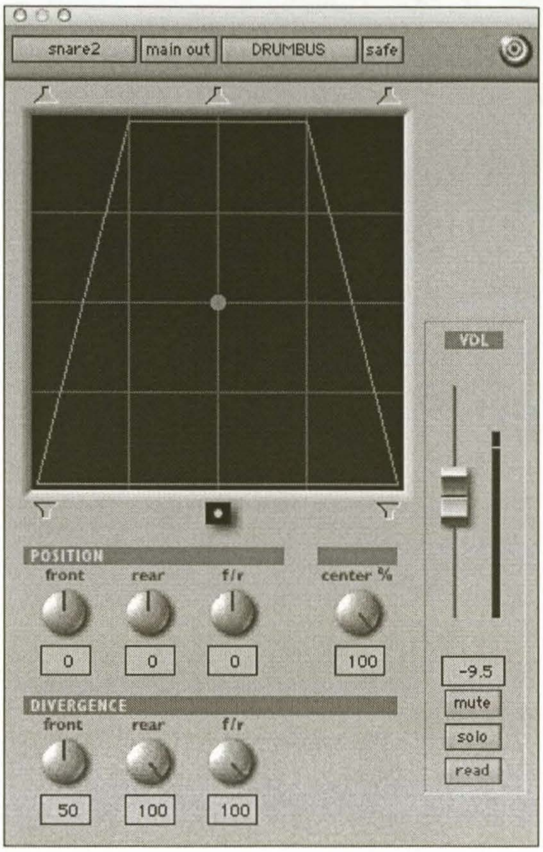


Figure 28: Output Window

An interesting effect is created by panning the snare drum in the build-up (in main climax) from centre middle to both Ls and Rs, whilst applying the *divergence knob*. The latter function is applied to create a more focused sound within a specified area. It should be noted that because full range monitors were not used, the timbre of instruments changes when automating sound sources from the front to the back. The latter can be ascribed to the use of two types of loudspeakers, and therefore not much of automation is being applied.

The final mixing was done by Mario Cronje after which it was mastered by Tim Lengfeld (TL Mastering 2006: on-line) into the final DVD[103] product. It should be noted that this project was composed and programmed in stereo. Therefore, the first step was to *up-mix* the session into 5.1. Although implemented, the latter process was quite lengthy and a detailed discussion is beyond the scope of this dissertation.

[103] Although DVD-A or SACD disc is the standard for only-music material, DVD-V was chosen as the delivery format.

CONCLUSION

Through the research done for the present dissertation it was discovered that, although there is great interest in the application of 5.1 surround in EDM, the implementation of this approach to mixing EDM is still relatively rare. While the practical chapter provided above illustrates that 5.1 can be successfully implemented within the context of EDM, there remains much scope for research on and exploration of the practical application of 5.1 in this field. The growing emphasis placed on 5.1 mixing in studios around the world is most assuredly set to render the necessity for research in this field more urgent.

REFERENCES

Acoustic Engineering (2006) *Wallace Clement Sabine*. [on-line]. Availible: http://www.acoustics-engineering.com/sabin/wcsabine.htm [August 18, 2006].

AES Technical Document AESTD 1001 (2001) *Multi-channel surround sound systems and operations*. [on-line]. Available: http://www.aes.org/technical/documents/AESTD1001.pdf. [August 23, 2004].

AES Workshop (2006) *W-7 Hands-on workshop on surround mixing techniques (Abstract)*. [on-line]. Available: http://www.aes.org/events/121/workshops/session.cfm?code=W7.

Alberts, R. (2004) *BT: Getting All Emotional about Technology*. [on-line]. Available: http://www.grmtools.org/userstories/bt.html [August 4, 2006].

Barbour, Jim (2002) *Applying Aural Research: the aesthetics of 5.1 surround.* School of Social and Behavioural Sciences, Swinburne University. [on-line]. Available: http://www.iii.rmit.edu.qu/sonolog/acmc/ACM2002_Web_Proceedings/017-026_Barbourpdf [March 4, 2004].

Beranek, L. (1996) *Concert and Opera Halls: How They Sound.* Woodbury, NY: Acoustical Society of America, 1996.

Blauert, J. (1969/1970) *Sound Localization in the Median Plane*, in: *Acustica 22*.

Blauert, J. (1999) *Spatial Hearing: The psychophysics of human sound localization*. England: The MIT Press.

Blauert, J. (2003) *Spatial Hearing*. [on-line]. Available: http:// www.uni-koern.de/philfak/mmuwi/fricke/103blauert.pdf [August 23, 2006].

Bogdanov, V., Busch, J., Erlewine, S.T. and Woodstra, C. (2002) *All Music Guide to Electronica: The Definitive Guide to Electronic Music*. San Francisco: Backbeat Books.

Borwick, J. (1997) *Sound Recording Practice 4th Edition*. Oxford/New York: Oxford University Press.

Bradley, J.S. and Soulodre, G.A. (1994) *The Influence of Late Arriving Energy on Spatial Imperssion,* in: *Journal of the Acoustical Society of America 97 (4).*

Bradley, J.S., Norcross, S.G. and Reich, R.D. (2000) *On the Combined Effects of Early and Late arriving Sound on Spatial Impression in Concert Halls*, in: *Journal of the Acoustical Society of America 108 (2).*

Bregman, A. (1990) *Auditory scene analysis: the perceptual organisation of sound.* Cambridge, Massachussets: MIT Press.

Bunish, Christine (2003) *Surround sound: for some apps-DVD, Cinema spots and films-this format is taking off, but what about TV?*. [on-line]. Available: http://www.postmagazine.com/ME2/Default.asp. [August 3, 2006].

Butler, David. (1992) *The Musicians's Guide to Perception and Cognition.* New York: Schirmer Books, A division of Macmillian, Inc.

Butler, Mark J. (2003) *Unlocking the Groove: Rhythm, Meter, and Musical Design in Electronic Dance Music.* Ann Arbor, USA: ProQuest Information and Learning Company.

Chang, Peter Sou-Kong. (2004) *Exploration of Behavioral, Physiological, and Computational Approaches to Auditory Scene Analysis.* Degree Master of Science in the Graduate School of the Ohio State University.
[on-line] Availible:http://www.theagnostic.com/research/writings/Chang_Peter_MSThesis.pdf [March 27, 2006].

Chomyszyn, J. (1992) *Loudness as a cue in distance perception*, in: *Proceedings of the 1992. International Computer Music Conference*, San José State University, California, Oct. 14-18.

Correira, M.J. (2002) *Haircells.* University of Texas Medical School, Galveston, Texas, USA Encyclopedia of Life Sciences. USA: Macmillan Publishers Ltd, Nature Publishing Group . [on-line]. Available: http://www.els.net [August 10, 2006].

Crane, D. (2003) *Pro Tools Power User Rich Tozzoli*, in: *Universial Audio Webzine, Volume 1, Number 5.* [on-line]. Available: http://www.uaudio.com/webzine/2003/august/text/content7.html [August 3, 2006].

Dana, P.H. (1995) *Coordinate Systems Overview.* [on-line]. Available: http://www.colorado.edu/geography/gcraft/notes/coordsys/coordsys.html [September 6, 2005].

Digidesign. (2005) *Pro Tools Reference Guide (Version 7.0).* Digidesign, A Division of AVID Technology.

Digidesign: A Division of Avid (2006a) *What's New: Pro Tools HD 7.2.* Digidesign, A Division of AVID Technology. [on-line]. Available: http://akmedia.digidesign.com/support/docs/New_in_PTHD_72_29745.pdf. [June 2, 2005].

Digidesign: A Division of Avid. (2006b) *Dolby Surround Tools.* [on-line]. Available: http://www.digidesign.com/index.cfm?&key=/products/details.cfm&itemid=1042 [August 3, 2006].

Dolby Laboratories Information (2006) *Some guidelines for producing music in 5.1 channel surround.* [on-line]. Available: http://www.transtec.nl/download.php?id=189&file=Pro%2Fdolby%2F5.1mixing.pdf [August 2, 2006].

Dolby Laboratories, Inc. (2003) *Dolby 5.1-Channel Music Production Guidelines, Issue 2.* [on-line]. Available: http://www.dolby.com/assets/pdf/tech_library/4_Multichannel_Music_Mixing.pdf [May 10, 2004].

Dolby News, Professional Audio Edition (2004) *Dolby Model 585 Breaks Boundaries in Pitch Shifting*. [on-line]. Available: www.dolby.com/assets/pdf/newsletters/pa_nw_0409_AES2004Newsletter.pdf [June 2, 2006].

Durant, J.D. and Levrinic, J.H. (1995) *Bases of Hearing Science (2nd edition).* Baltimore: Williams and Wilkens.

Faller, C. and Merimaa, J. (2004) *Source localization in complex listening situations: Selection of binaural cues based on interaural coherence*, in: *Journal of the Acoustical Society of America, Vol. 116, No. 5.*

Ferl, R.J, Sanders, G.P. and Wallace, R.J. (1996) *Biology: The Science of Life (4th Edition).* New York: Harper Collins College Publishers.

Grey, H. (1918) *Anatomy of the Human Body.* Philadelphia: Lea & Febiger. [on-line]. Available: http://www.bartleby.com/107/pages/page1033.html [June 4, 2006].

Illingworth, V. (1998) *The Penguine Dictionary of Electronics (3rd edition).* London: Penguin Books Ltd.

ITU-R Recommendation BS.775-1. (1994) *Multichannel Stereophonic Sound System with and without accompanying Picture.*
[on-line]. Available: http://www.itu.int/rec/R-REC-BS.775/ [Augustus 30, 2004].

JBL Proffesional. (1999) *Sound System Design Reference Manual*. [on-line]. Available: http://www.jblpro.com/pub/manuals/pssdm_1.pdf. [June 3, 2006].

Jenkin, R.M., Kapralos, B. and Milios, E. (2003) *Auditory Perception and Spatial (3D) Auditory Systems*. Technical Report CS-2003-07. Department of Computer Science, University of York, Canada. [on-line]. Available: http://www.cse.yorku.ca/techreports/2003/CS-2003-07.pdf [March 2, 2004].

Katz, B. (2002) *Mastering Audio: the art and the science*. Canada: Focal Press.

Kissel, T.B. (2005) *Positively BT*. EQ Online Edition. [on-line]. Available: http://www.eqmag.com/story.asp?storycode=9813 [July 18, 2006].

Lamancusa, J.S. (2000) *Room acoustics: Acoustics of rooms and enclosures*. [on-line]. Available: http://www.mne.psu.edu/lamancusa/me458/8_rooms.pdf. [December 1, 2004].

Levine, M. (2004) *Sonic Surgeon.* [on-line] Available: http://emusician.com/ar/emusic_sonic_surgeon/index.htm [August 2, 2006].

Mackensen, P. (2004) *Auditive Localization: Head movements, an additional cue in Localization.* Doktor der Philosophie, Berlin.

Macmusic. (2006) *Pro Tools HD*. [on-line]. Available: http://www.macmusic.org/software/view.php/lang/en/id/1213/ [August 3, 2006].

Malham D. (1998) *Spatial Hearing mechanism and Sound Reproduction.* University of York. [on-line]. Available: http://www.york.ac.uk/inst/mustech/3d_audio/ambis2.htm [January 6, 2004].

M-Audio: A Part of Avid. (2006) *News: M-Powered Artist BT: Staying Ahead of the Curve.* [on-line]. Available: http://www.m-audio.com/news/en_us-235.html [August 2, 2006].

Miller, G. (2006) *The Work of Art in the Age of Mechanical Reproduction*, in: *Stylusmagazine.* [on-line]. Available: http://www.stylusmagazine.com/articles/weekly_article/the-real-deal-toward-an-aesthetic-of-authentic-live-electronic-dance-music.htm [August 28, 2006].

Mixonline. (2005) *Recording Academy's P&E Wing Launches New Initiatives.* [on-line]. Available: http://www.mixonline.com/news/headline/naras-pe-wing-0602051/index.html [February 3, 2006].

Moore, B.C.J. (1989) *An Introduction to the psychology of Hearing (3rd Edition).* London: Academic Press Limited.

Moylan, W. (2002) *The Art of Recording, Understanding and crafting the mix.* Focal Press, United States of America.

Munkstedt, S. (2006) *Subjectively prefferd sound representation for different listening situations.* Master Thesis in Hearing Technology, Department of Speech, Music and Hearing Royal Institute of Technology, S-100 44 Stockholm. [on-line]. Available: http://www.speech.kth.se/publications/masterprojects/2006/SamuelMunkstedt.pdf [February 12, 2006].

Neher, T. (2004) *Towards a Spatial Ear Trainer.* Department of Music & Sound Recording, School of Arts, University of Surrey, Submitted for the degree of Doctor of Philosophy. [on-line]. Available: http://www.surrey.ac.uk/soundrec/pdf/NeherThesis.pdf [March 26, 2006].

Nielsen, S.H. (1993) *Auditory distance perception in different rooms*, in: *Journal of the Audio Engineering Society, Volume 41, no. 10.*

Nisbett, A. (1995) *The Sound Studio 6th Edition.* Oxford: Focal Press.

Nsamba, C. (1979) *Ernst Reissner 1824-1878*, in: *Archives of Otolaryngology-Head and Neck Surgery, Volume 105, No. 7.* [on-line]. Available: http://archotol.ama-ssn.org/cgi/content/abstract/105/7/434 [August 13, 2006].

O'Connor, J.J. and Robertson, E.F. (2003) *John William Strutt.* The MacTutor History of Mathematics archive. University St.Andrews, Scotland. [on-line]. Available: http://www-history.mcs.st-and.ac.uk/Biographies/Rayleigh.html [June 3, 2006].

Palmer, S and Rock, I. (1990) *The legacy of Gestalt psychology,* in: *Scientific American. 263.*

Parncutt, R. (2004) *Psychoacoustics and music perception.* Submitted to Bruhn, Kopiez, Lehmann, Oerter: Musikpsychologie-das neue Handbuch. [on-line]. Availible:http://www-gewi.uni-graz.at/staff/parncutt/PSYCHOACOUSTICS.pdf [April 4, 2006].

Performance Audio. (2006) *Digidesign : Access Virus Indigo World-Class Synthesizer Plug-In for Pro Tools TDM Systems.* [on-line]. Available: http://www.performanceaudio.com/cgi/pages.cgi?buildPage=item&ProductID=001250 [June 3, 2006].

Pickles, J.O. (1982) *An Introduction to the Physiology of Hearing*. New York: Academic Press.

Pluginz.com: Digital Tools for Digital Artists. (2006) *Dolby Surround Tools*. [on-line]. Available: http://www.pluginz.com/product/10472?&sessionhash=7f0235afab72a13598db165a492add18 [August 20, 2006].

ProAV. (2006) *ProAV*. [on-line]. Available: http://www.bnoack.com/index.html?http&&&www.bnoack.com/data/A-weighting.html [August 15, 2006].

Raichel, D.R. (2000) *The science and applications of Acoustics*. Chapter 10, 11. New York, Inc: AIP Press.

REMIX. (2006) *Reading for the Dancefloor*. [on-line]. Available: http://remixmag.com/mag/remix_reading_dancefloor/index.html. [June 4, 2006].

Rudolph, B. (2003) *Signal Processing and Methods in Surround Mixing: Proffesional Tips and Tricks*, in: *Mixguides*. [on-line]. Available: http://mixguides.com/consoles/tips_and_techniques/signal-methods-surround-0303/ [January 22, 2006].

Rumsey, F. (2001) *Spatial Audio*. Great Britian: Focal Press.

Savioja, L., Huopaniemi, J., Lokki, T., Väänänen, R. (1999) *Creating interactive virtual acoustic environments*, in: *Journal of the Audio Engineering Society*, vol. 47, no. 9.

Sunier, J. (2006) *BT-This Binary Universe-DTS Entertainment*, in: Dual Disc Reviews, *Audiophile Audition*. [on-line]. Available: http://www.audaud.com/article.php?ArticleID=1742 [August 2, 2006].

Surround Sound Forum. (2002). *Recommended Practice SSF-02.1-E-2002. Multichannel Recording in s3/2 Format: Parameters for Program Interchange and Archiving, Alignment of Reproduction Equipment*. [on-line. Available: http://www.tonmeister.d/foren/surround/texte/SSF_02_F_2002.pdf [February 23, 2004].

Scavone, G.P. (2006) *Online course notes for 'MUMT 614: Computational Acoustic Modeling* Winter 2006' at the Department of Music Technology, McGill University. Montreal, Quebec, Canada, revise. [on-line]. Availible: http://www.music.mcgill.ca/~gary/614/week3/room_acoustics.html#SECTION00012000000000000000 [August 10, 2006].

Science Clarified (2006) *Science Clarified*. [on-line]. Available: http://www.scienceclarified.com/index.html [August 2, 2006].

Shinn-Cunningham B. (2001) *Localizing Sound in Rooms*, in: *Proceedings of the ACM SIGGRAPH and EUROGRAPHICS Campfire: Acoustic Rendering for Virtual Environments,Snowbird, Utah.*

Smaragdis, Paris (2001) *Redundancy Reduction for Computational Audition, a Unifying Approach*. Program in Media Arts and Sciences, Doctor of Philosophy in Perceptual Computing. [on-line].

Availible: http://web.media.mit.edu/~tristan/Classes/.../Technical/Smaragdis_Thesis.pdf [June 4, 2006].

Snoman, R. (2004) *The Dance Music Manuel: Tools, Toys and Techniques*. Great Britian: Focal Press.

Snyder, B. (2000) *Music and Memory: An Introduction.* Cambridge, Massachusetts, London, England: The MIT Press.

Sone, T. (1968) *On the difference between localization and lateralisation*, in: Reports of the 6. ICA, Tokyo, No. A-3-6.

Sound Retrievel Systems [SRS]. (1998) *SRS Technical Description.* SRS Laboratories, Incorporated. [on-line]. Available: http://www.paysan.co.uk/SRS Technology.pdf [January 23, 2004].

Swenson, K. (2002) *Awakening Spirits*, in: Remix Magazine. [on-line]. Available: http://remixmag.com/index59.html [January 2, 2006].

The Columbia Electronic Encyclopedia. (2002) *The Columbia Electronic Encyclopedia.* Columbia University Press. [on-line]. Available: http://www.columbia.edu/cu/cup/cee/cee.html [August 21, 2006].

The Recording Academy's Producers & Engineerings Wing. (2004) *Recommendations For Surround Sound Production.* The National Academy of Recording Arts & Sciences, Inc. [on-line]. Availible:http://www.grammy.com/PDFs/Recording_Academy/Producers_And_Engineers/5_1_Rec.pdf [August 3, 2006].

Theile, G. (1980) *Uber die Lokalisation im ¨uberlagerten Schallfeld* [On Localization in a superimposed sound field]. Ph.D. thesis, Technische Universtit¨at Berlin, 1980. [on-line]. Available: http://www.irt.de/wittek/hauptmikrofon/theile/ON_THE_LOCALISATION_englishpdf [December 15, 2006].

Tim Lengfeld. (2006) *TL Mastering Services: Comprehensive Mastering*. [on-line]. Available: http://www.tl-mastering.com/services.htm [August 30, 2006].

Tozzoli, R. (2006) *Top 10 Surround Sound Production Tricks*, in: EQ Online Edition. [on-line]. Available: http://www.eqmagazine.com [August 24, 2006].

Verderosa, T. (2002) *The Techno Primer: the essential reference for loop-based music styles.* Milwaukee: Hall Leonard Corporation.

Von Békésy, G. (1960) *Experiments in Hearing* (ed. E.G. Wever). New York: McGraw-Hill.

Watkinson, J. (1999) *MPEG-2.* Great Britian: Focal Press.

West, James R. (1998) *Five-Channel Panning Laws: An Analytical and Experimental Comparison.* Masters of Science in Music Engineering Technology, University of Miami. [on-line]. Available: http://www.music.miami.edu/programs/mue/Research/jwest/Master.html [April 3, 2005].

White, G.D. (1991) *The Audio Dictionary 2nd Edition.* Seattle: University of Washington Press.

Wolfe, J. (2006) *What is Acoustic Impedance and Why is it Important*. Music Acoustics. University of New South Wales. [On-line]. Available: http://www.phys.unsw.edu.au/~jw/z.html [August 13, 2006].

Wrigley, S.N. (2002) *A Theory and Computational Model of Auditory Selective Attention*. Doctor of Philosophy in Computer Science, Department of Computer Science, University of Sheffield. [on-line]. Available: http:// www.dcs.shef.ac.uk/~stu/pdf/phd.pdf [April 2, 2006].

APPENDIX

DOLBY®

5.1-Channel Music Production Guidelines

Issue 3

Dolby Laboratories, Inc.

USA

Dolby Laboratories, Inc.
100 Potrero Avenue
San Francisco, CA 94103-4813
Telephone 415-558-0200
Fax 415-863-1373
www.dolby.com

Dolby Laboratories, Inc.
3601 West Alameda Avenue
Burbank, CA 91505-5300
Telephone 818-823-2800
Fax 818-823-2833

Dolby Laboratories, Inc.
1350 Avenue of the Americas
New York, NY 10019-4703
Telephone 212-767-1700
Fax 212-767-1705

UK

Dolby Laboratories, Inc.
Wootton Bassett
Wiltshire SN4 8QJ England
Telephone (44) 1793-842100
Fax (44) 1793-842101

Issue 3
S05/14926/15996

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

With the advent of high-quality multichannel consumer music formats, 5.1-channel music producers may benefit from a standardized set of production practices. This document guides the multichannel music professional, working in small- to medium-sized control rooms, in the production of high-quality, 5.1-channel music intended for playback in consumer environments. If followed, these guidelines will facilitate interchangeable critical listening judgments between various locations.

Please note that production practices for 5.1-channel music intended for playback in large cinemas are well-documented elsewhere, and that there are methods by which a home playback system can be made to emulate the cinematic playback experience. Neither of these topics is discussed in this document, which focuses on the 5.1-channel music experience afforded by the expanded artistic flexibility and palate of a 5.1-channel soundscape, as well as its delivery from the production studio directly to the home.

Many aspects of 5.1-channel music mixing are covered in this document, including monitoring, recording levels, practical setup guidelines, and program interchange standards. Subjects such as room design are discussed with recommended target parameter values.

This manual can be used as a quick setup reference (starting in Section 3.3) or as an in-depth sourcebook for the multichannel music creator.

This manual is written from the perspective that:

- 5.1-channel music intended for home enjoyment is a unique art with its own requirements, but these requirements can co-exist with 5.1-channel production practices already established for cinema or broadcast applications.
- It would be harmful to the advancement of 5.1-channel music if every industry using it made conflicting demands on the consumer playback environment.
- A universal multichannel sound system applicable to music, cinema, and broadcasting would be beneficial to the listener.
- Flexibility within these guidelines may be necessary to ensure that the system is as universal and as practical as possible.

Every effort has been made to utilize existing international standards: please see the reference section for a complete list of sources.

Chapter 2
Historical Perspective

2.1 Cinema Sound

5.1-channel audio was first developed for cinema applications. Cinema sound has an advantage over other consumer playback experiences: it is mixed in an environment that is extremely similar to that in which it is enjoyed. Film dubbing stages largely use the same speaker/crossover/amp systems as commercial theatres and both follow the same guidelines for room equalization. All aspects of the sound, such as the recording levels on the film soundtrack, program equalization, and the overall monitor levels during playback have been standardized and calibrated so that what the mixers create on the dubbing stage matches what is heard in the cinema.

This is not the case with 5.1-channel music. Every 5.1-channel music engineer/producer has their favorite monitors to mix on, just as every mastering engineer has their own approach and monitoring system. Because most consumers don't have systems remotely similar to what the music was mixed and mastered on, the chances of successfully translating the studio experience to the home are currently less then optimal.

Many consumers listen to new 5.1-channel music mixes on their home theater systems, originally bought to watch movies. Also, because tens of thousands of movie titles have audio referenced a particular way, it makes sense to adopt some of the production practices for reference of film calibration, so that the consumer experience does not vary wildly between watching a film and listening to 5.1-channel music on the same system.

These issues present particular challenges to the 5.1-channel music mixer. This manual recommends production practices in an effort to approach a level of standardization that will help ensure that the consumer hears what the 5.1-channel music producer intended.

2.2 The LFE Channel

There are two distinct purposes for a subwoofer in a 5.1-channel music system. One entails the reproduction of the low-frequency effects (LFE) channel information and is discussed here in light of its cinematic origin. The other entails the reproduction of the bass content from other channels (bass management) and is discussed in Section 4.4.

Even though the continuous evolution of power amplifier and loudspeaker technologies have made it possible to reproduce a higher quality of sound in the cinema, it is still not easy to deliver or reproduce deep or accurate bass with a uniform response in large rooms filled with people. The best soundtracks of the late 1970s (70 mm magnetic analog) had reached their maximum recording capability, so it was impossible to increase the bass content without causing overload. Moreover, often playback systems are inefficient at reproducing low frequency information. Even today, the main screen speakers used in cinemas typically do not reproduce below 30 Hz, so if the soundtrack carries additional low frequency information to the amplifiers, it is not necessarily reproduced.

To illustrate how adding low-frequency content can tax a system, consider the relative sensitivity of the human ear to different frequencies at different monitoring levels. As seen in Figure 2-1, at a reference level of 80 dB, it takes an additional 10 dB of 65 Hz to equal the perceived loudness of 1 kHz. Adding this kind of level to the main channels often causes overload and decreases the available dynamic range.

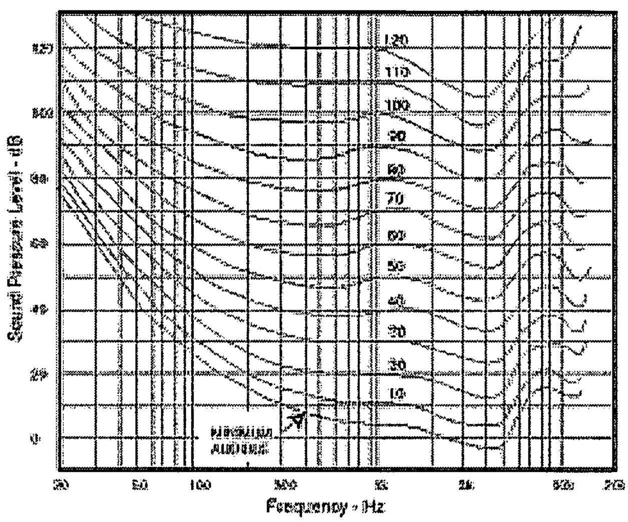


Figure 2-1 Fletcher-Munson Equal Loudness Curve

Subwoofers were installed to increase low-frequency playback capabilities and increase the overall dynamic range of cinema systems. A separate channel, LFE, was added to the soundtrack to provide an additional bass signal to the subwoofers. The LFE channel handles bass created specifically for special subwoofer effects.

Current 5.1-channel music delivery formats such as DVD-Audio allow each channel in a 5.1-channel mix to carry bass content. So why is there an LFE channel in a consumer audio delivery format? Quite simply, it allows movie soundtracks to be transcribed directly without alteration to the home video format. However, this scenario does not dictate the use of the LFE channel for multichannel music. It suggests that the LFE channel may not be the only or the best way to provide loud, deep bass, which becomes more apparent when one mixes multichannel audio using a properly configured and calibrated studio monitor system.

The use of a separate loudspeaker for reproducing the lowest part of the frequency spectrum in 5.1-channel music applications does have several advantages, including:

- Freedom to locate the bass source optimally in relation to room mode pressure distributions
- Reductions in the size (and sometimes cost) of the main loudspeakers
- Reduction in distortion because the main loudspeaker driver displacements can be reduced.

In short, if you really want the widest dynamic range and frequency response but feel that putting additional bass content in the five main channels will cause overload or decrease the overall dynamic range of those channels, the LFE channel may be just the tool to accomplish the goal.

2.3 Current Multichannel Sound Systems

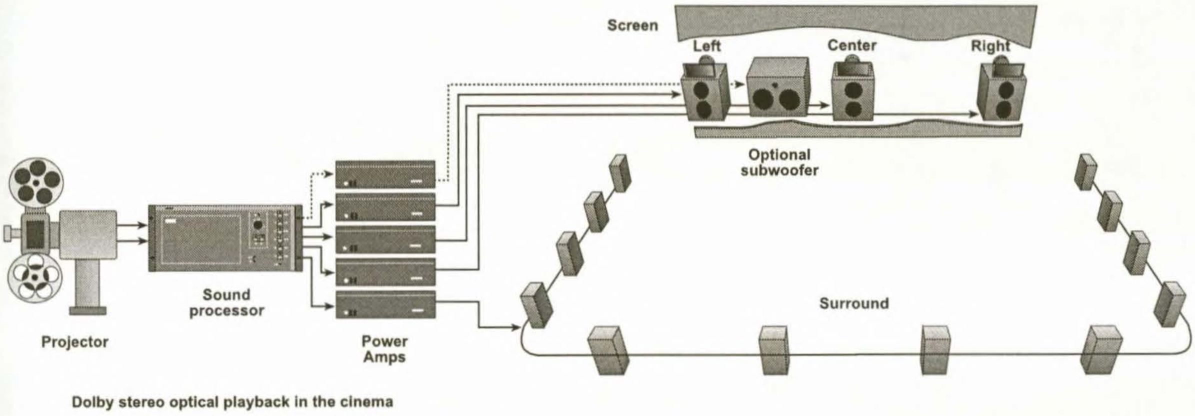


Figure 2-2 Dolby® Stereo (Motion Picture Matrix Four-Channel [4:2:4]) Setup for Theatres

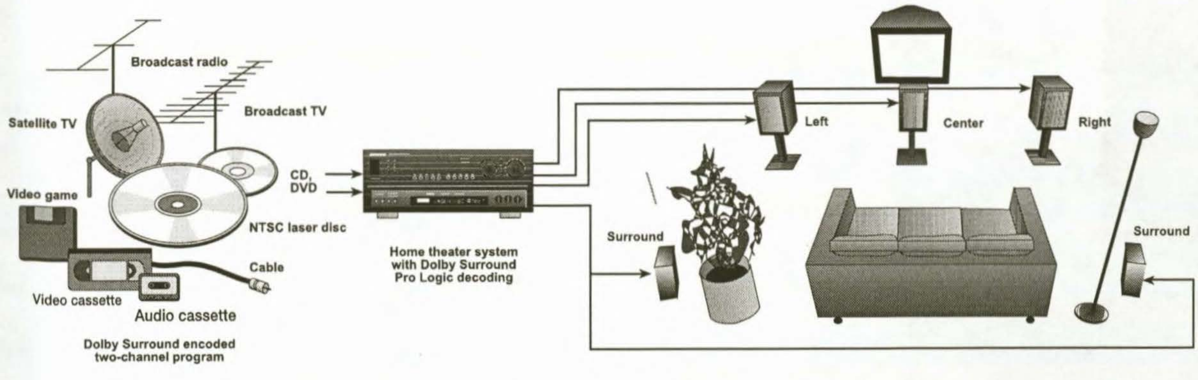


Figure 2-3 Dolby Surround (Matrix Four-Channel [Pro Logic®]) Setup for the Consumer

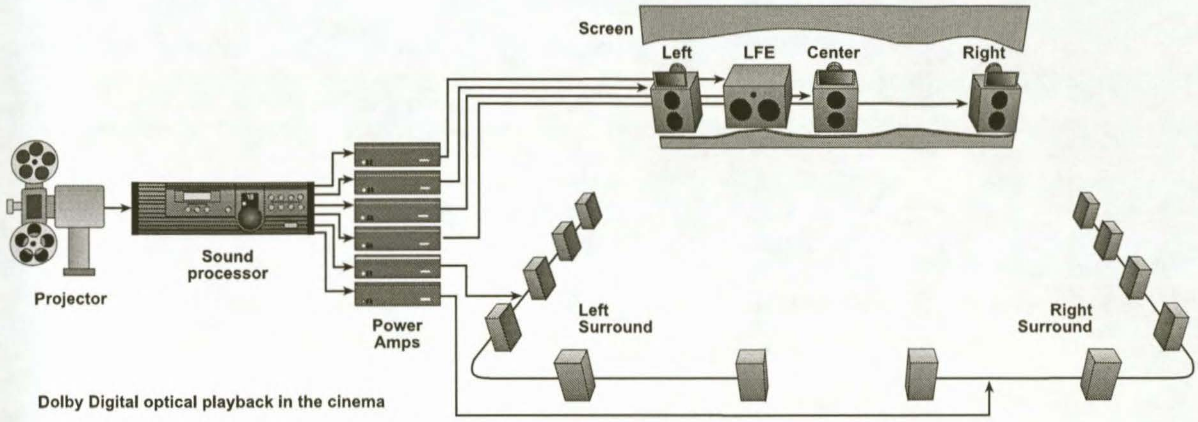


Figure 2-4 5.1-Channel Setup for Theatres

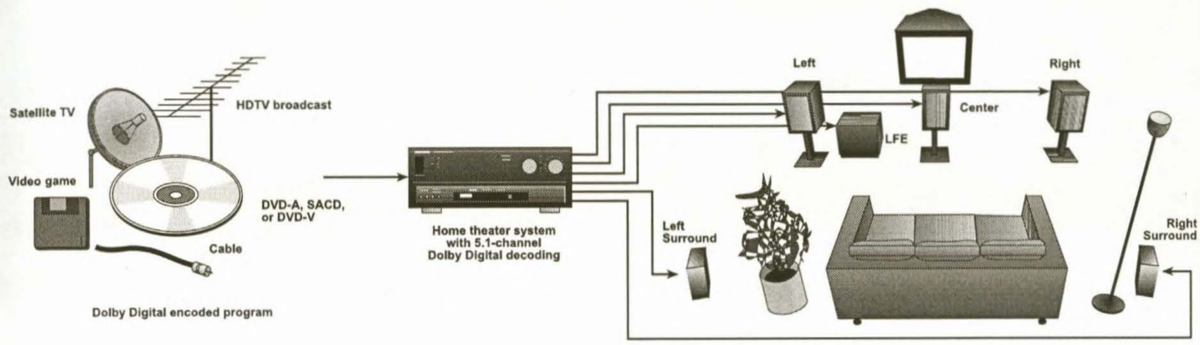


Figure 2-5 5.1-Channel Setup for the Consumer

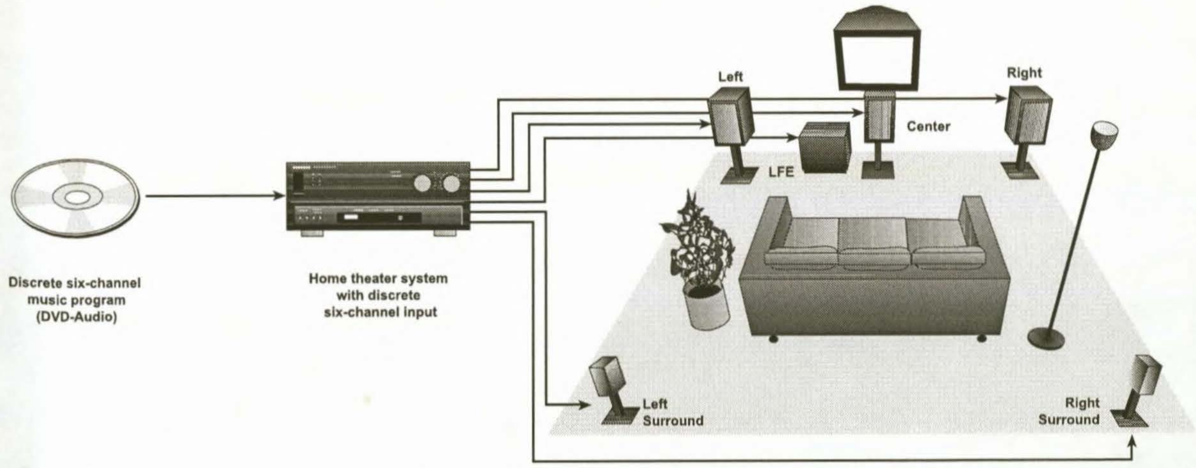


Figure 2-6 Alternate 5.1-Channel Setup for Consumer Music Playback

Chapter 3
The 5.1-Channel Music Mixing Environment

3.1 Room Design

Significant differences exist between stereo and multichannel production environments. Multiple speakers firing in different directions affect such basic factors as the optimum room size and geometry, equipment needs, construction methods, wiring, HVAC, lighting, power, and ergonomics. The addition and placement of equipment necessary for multichannel production often affects room acoustics as well. Whether designing a new facility or planning to retrofit an existing one, consulting a professional acoustician and architect familiar with building critical audio monitoring environments is always recommended.

The primary aim of a reference listening room is to facilitate interchangeable judgments between locations. Please note that these guidelines relate primarily to small- and medium-sized rooms.

3.1.1 Dimensions

The parameters given in Table 3-1 are meant as general guidelines and cannot completely describe the optimal sounding room; however, they are based on existing international standards for reference listening conditions for small- to medium-sized control rooms and provide a good starting point.

While many room shapes may work, the ideal room is symmetrical along the line between the center speaker and the reference listening position. An environment with no parallel walls (including the floor and ceiling) helps prevent the buildup of low-frequency standing waves.

A minimum height of 3 meters (9 feet) is desirable.

Table 3-1 Room Dimensions

Parameter	Units/Conditions	Value
Room Floor Area		>30 m^2 (320 ft^2)
Room Volume		<300 m^3 (10,500 ft^3)
Room Proportions	L = Length (larger dimension, irrespective of orientation) W = Width (shorter dimension, irrespective of orientation) H = Height	1.1W/H ≤ L/H ≤ 4.5W/H –4 with L/H <3 and W/H <3 No ratios of L, W, and H within ±5% of an integer value

For stereo listening, the speakers are traditionally set up on the short dimension of the room firing into the long dimension. Many new production rooms designed for 5.1-channel applications now place the front speakers on the long dimension, firing into the short dimension, which usually provides the most efficient and symmetrical use of the listening space. Multiuse production rooms that have been set up with alternate speaker orientations do not need to be reoriented. If there is an opportunity to redesign a room, reorientation should be considered.

3.1.2 Acoustics

Early Reflections

Any early reflections (within 15 ms) should be at least 10 dB below the level of the direct sound for all frequencies in the range 1 kHz to 8 kHz [6].

Reverberation Field

Reverberation time is frequency-dependent. The nominal value, T_m, is the average of the measured reverberation times in the 1/3-octave bands from 200 Hz to 4 kHz and should lie in the range: $0.2 < T_m < 0.4$ s. T_m should increase with the size of the room; the formula in Table 3-2 is a guide.

Table 3-2 Reverberation Values

Parameter		Units/Conditions	Value
Reflected Sound	Early Reflections	0–15 ms (in region 1–8 kHz)	< –10 dB relative to direct sound
	Reverberation Time	T_m[s] = nominal value in region of 200 Hz to 4 kHz V = listening room volume V_0 = reference room volume (100 m^3 [1075 ft^2])	$\approx 0.25(V/V_0)^{1/3}$

The reverberation time T, measured in 1/3-octave bands over the frequency range from 63 Hz to 8 kHz, should conform to the tolerance mask shown.

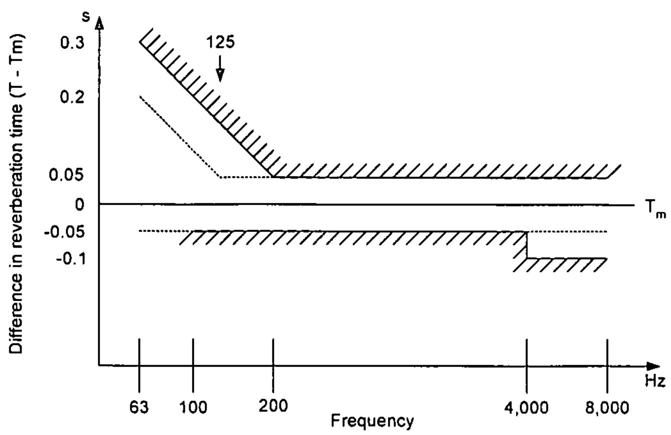


Figure 3-1 Tolerance Limits for Room Reverberation Time
(dotted lines = tighter tolerances proposed by the AES, see [1])

Reflective and Absorbent Surfaces

Large flat reflective surfaces should be avoided in the mixing environment. Placement of doors, control room windows, and equipment should be considered with speaker placement and aiming in mind. A combination of diffuse reflectors and absorptive materials should be used to achieve a smooth RT decay time within the specified range shown in Figure 3-1.

Again, it is recognized that these values may not be achievable in some installations, but is recommended that the room be measured using a real-time analyzer and that architectural solutions (wall treatments, bass traps, room reorientation, and so on) be utilized first to achieve the recommended values. A mixture of diffuse reflective and absorbent surfaces, applied evenly to the whole room, aids in creating an acceptable reference listening condition [12].

Only after considerable effort has been made using architectural solutions to smooth the room response should equalizers be introduced into the monitor chain. See Section 4.2 for more information on room equalization.

Background Noise

The listening area should ideally achieve an NC rating of 10 or below with the equipment off, measured at the reference position. A studio with equipment such as video projectors, video monitors, and other ancillary equipment powered on should achieve a rating of ≤ NC 15.

Any background noise should not be perceptibly impulsive, cyclical, or tonal in nature.

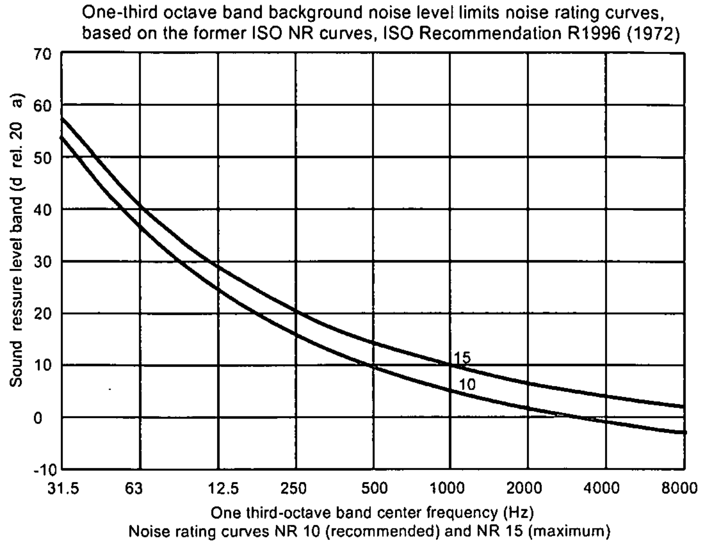


Figure 3-2 Noise Rating Curves

NR 10 or NR 15 may be hard to realize in a practical manner in some installations, in which case, every effort should be made to identify the loudest noise sources and correct as appropriate. The most common noise sources and possible remedies include:

- **HVAC systems**: Increase the surface area of the supply air vent. Separate or float all mechanical connections between high velocity or rumbling motors and ducts and the listening room.
- **Equipment**: Contain computers and other equipment with loud fan noise in noise attenuating, ventilated cabinets.
- **Doors and windows**: Make sure all the doors and windows are aligned properly and form a seal when closed. Adding a second window or door, with air space between it and the original, can reduce unwanted noise considerably.

Other sources of problem noise may need to be addressed. Every effort should be made to approach the recommended values shown in Figure 3-2.

3.1.3 Console Placement

In most cases, the console should be placed equidistant from the listening studio sidewalls. Lateral placement is discussed in Section 3.3.

3.2 Monitoring

3.2.1 Reference Monitors

Note: All five loudspeakers (L, R, C, Ls, Rs) should be identical.

One of the main differences between 5.1-channel setups for music and those for home theater playback is the type of speakers used for the surround positions. The goal of surround reproduction in the cinema (accomplished using multiple speaker arrays) is to provide surround playback to large audiences.

Surround effects are often very diffuse, ambient soundscapes. Dipole speakers are sometimes used in the home environment (rarely used in movie theatres) at the surround locations to help create the wide wash of sound created using an array of speakers in the cinema.

However, in 5.1-channel music production, as well as in films, the surrounds are sometimes used for distinct placement of featured performers. Accurate reproduction in this case requires the use of direct-firing speakers that match the overall characteristics of the front speakers. Use of matched direct radiator or monopole speakers in 5.1-channel music production is recommended for achieving the greatest control of level, timbre, and image location. Due to their dependence on null spot positioning, reflective front and rear listening room walls, and preference of a diffuse surround field, dipole speaker monitoring is not ideal for critical 5.1-channel music production.

It is recommended that all five speakers (L, C, R, Ls, Rs) be identical in all of the parameters listed in Table 3-3.

Table 3-3 Reference Loudspeaker Specifications

Parameter	Units/Conditions	Value
Amplitude/frequency response	20 Hz to 20 kHz* on axis (0°)	4 dB
	±10°	Deviation to 0°, 3 dB
	Horizontal ±30°	Deviation to 0°, 4 dB
Difference between speakers	In the range >250 Hz to 2 kHz	.5 dB
4 Directivity Index	250 Hz to 16 kHz	8 dB ±2 dB
Nonlinear distortion attenuation (SPL = 96 dB)	<100 Hz	–30 dB (=3%)
	>100 Hz	–40 dB (=1%)
5 Transient fidelity Decay time t_s for reduction to a level of 1/e	t_s[s]	<5/f [Hz] (preferably 2.5/f)
6 Time delay	δt	≤10 µs
7 System dynamic range Maximum operating level (per IEC 60268, § 17.2, referred to 1 m distance)	$L_{\text{eff max}}$	>112 dB (at IEC 60268 program simulation noise or special condition)
Noise level	L_{noise}	≤10 dBA

* 20 kHz is a minimum value. Some delivery formats contain content up to 96 kHz. Choice of speakers may depend on the production format in use.

Source: Modified from AESTD1001.1.01-10

Not every speaker meeting these criteria may be useful for your purposes: conduct careful evaluations to determine the suitability of a particular speaker system.

For the pre-selection of loudspeakers, the frequency response curve over the range 20 Hz to 20 kHz, measured in one-third octave bands using pink noise on the main axis (directional angle = 0°), should preferably fall within a tolerance band of 4 dB. Frequency response curves measured at directional angles ±10° should not differ from the main axis frequency response by more than 3 dB, and at directional angles ±30° (in the horizontal plane only) by more than 4 dB.

The frequency response of different loudspeakers should be matched. The differences should preferably not exceed the value of 1.0 dB in the frequency range of at least 250 Hz to 2 kHz. [11]

The amplifier/speaker system should be capable of reproducing 120 dB without significant distortion.

3.2.2 Subwoofers

The frequency response of the subwoofer should be flat, ±3 dB between 20 Hz and 120 Hz.

Lower crossover frequencies will result in more freedom in the choice of subwoofer placement but with less benefit in the reduction in size of the main loudspeakers, as discussed in Section 2.2.

For almost complete freedom in the choice of location in a typical-sized room, the bass management (see Section 4.4) crossover frequency (when used) should be as follows:

Table 3-4 Reference Subwoofer Specification

Parameter	Value
Crossover frequency	80 Hz
Out-of-band harmonic distortion levels	≤–50 dB (0.3%)
Filter order	Fourth

Because the LFE channel for cinema applications contains content up to 120 Hz, however, the subwoofer itself should be capable of reproducing up to 120 Hz.

Parametric equalization may be needed to flatten the response of the subwoofer in a particular environment. See Section 4.2.

3.2.3 Power Levels

As a rule, the power amp should be able to provide 3 dB more power than the loudspeaker peak rating. The loudspeaker peak rating should be 3 dB higher than the peak allowed by the medium being mixed to.

3.2.4 Monitor Source Selection

Often, it is necessary to A/B between six-channel sources. If the console does not offer this feature, obtain an outboard device with this capability.

3.2.5 Multichannel Ganged Fader

If a multichannel mix is to be done on a console with fewer then six main monitor outputs, an outboard device offering six ganged attenuators (a six-channel volume knob) should be obtained. All six channels should track within a .5 dB tolerance over their full range.

3.3 Reference Positions

The reference position of the mixer's head is typically:

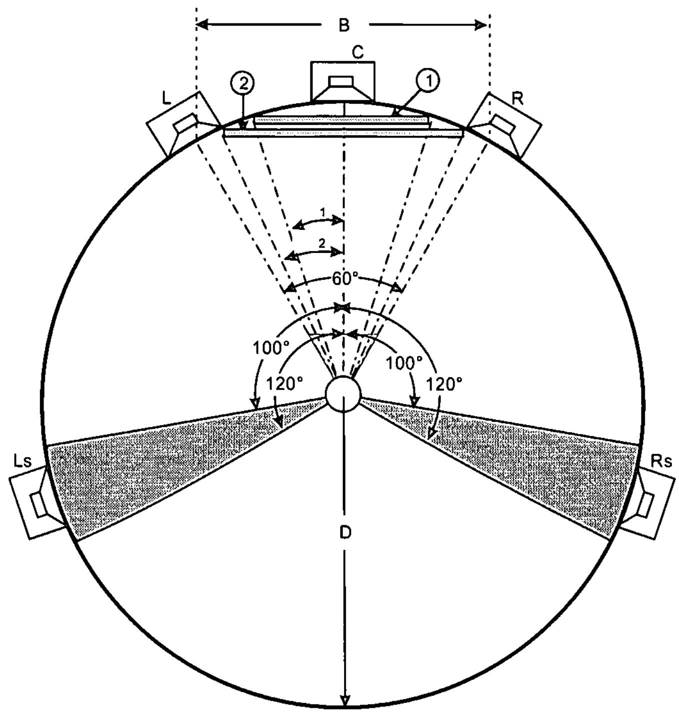
- At the center of the console
- Equidistant from the sidewalls
- Directly above the rear edge (arm rest) of the mixing console
- 1.2 meters (≈ 48 inches) off the floor

This imaginary point is the reference point for all speaker distances and aiming angles.

3.4 Speaker Placement

The speakers should be placed as shown in Figure 3-3. Final placement may depend on uncontrollable conditions, such as the physical dimensions or other constrictions of the facility and/or equipment. Note that these guidelines contain ranges of options, particularly with respect to the surround and subwoofer placement, which it is often necessary to test to obtain the most accurate and pleasing monitoring environment.

Regardless of particular system constrictions, please remember that the goal is a balanced multichannel monitoring system that properly images and facilitates interchangeable critical listening judgments between various locations.



H: height of screen
B: loudspeaker base width
D: listening distance
Screen 1 HDTV: reference distance = 3H (2 $_1$ = 33°)
Screen 2 = 3H (2 $_2$ = 48°)

Loudspeaker	Horizontal Angle from Center (degrees)	Height (meters/feet)	Inclination (degrees)
C	0	1.2/4	0
L, R	30	1.2/4	0
Ls, Rs	100–120	≥1.2/4	0–15 down

Figure 3-3 Reference Loudspeaker Placement

3.4.1 Front-Speaker Placement

The three front speakers (L, C, R) should be the same distance from the reference position. The center channel should be directly to the front of the reference position. With reference to the line formed between the center speaker and the reference position, L and R should be ±30 degrees horizontally.

Each front speaker should be the same height as the reference position (1.2 m/4 ft). All front speakers must be ±0 degrees vertically referenced to each other unless the center speaker needs to be positioned above or below a video monitor, forcing the acoustic centers of the three front speakers out of alignment. If this occurs, attempt to situate the speakers so the tweeters are in as close to a straight, horizontal line as possible. This may require either an inverted or lateral orientation of the center speaker, as well as rotating the center tweeter (when possible) to maintain the proper dispersion characteristic. In any case, keep the speakers equidistant from and directed to the reference position.

Table 3-5 Reference Position

Parameter	Units/Conditions	Value
Base Width	B [m]	2–3 m (6.5–10 ft)
Basis Angle	[°] referred to L/R	60°
Listening Distance	D [m]	= B

3.4.2 Surround Speaker Placement

The surround speakers should also be the same distance away from the reference position as the front speakers and located 110° ±10° from the reference line. They may be elevated to a position not to exceed 15° above the reference position, as long as they remain equidistant from and directed to the reference position.

3.4.3 Subwoofer Placement

Positioning the subwoofer(s) can often be an arduous task and the relative locations are not the same for all rooms. Expect a certain amount of experimentation, particularly when retrofitting an existing production room. The main requirement is that the location of the subwoofer not be audibly apparent.

One method for locating an optimal position is to place the subwoofer(s) near the listening position and play program material with significant low-frequency content. Then, listen at likely subwoofer locations around the room and choose the location that delivers the smoothest bass response. This location is apt to be the best choice for final subwoofer placement. Remember that the signal to the subwoofer is band-limited anywhere from 80 to 120 Hz and that as the crossover frequency rises, the ability to localize the loudspeaker position increases. Because the LFE channel may have content up to 120 Hz, it is recommended that the sub crossover be set at 120 or bypassed (as the bass-management filter will provide the necessary rolloff). However, keeping the crossover frequency for the bass-managed channels low (80 Hz) provides the greatest flexibility in positioning the subwoofer.

As recommended previously, control rooms are often set up in a symmetrical design, making it tempting to locate the sub in an equally symmetrical location (for instance, along the center line under the front speaker). However, a symmetrical placement in a symmetrical room often creates symmetrical standing waves and thus, an uneven room response. Placing the sub slightly asymmetrically may produce a more satisfactory result. Using a second sub can also help smooth out uneven room response problems.

Chapter 4
The 5.1-Channel Music Mixing Signal Path

4.1 Level Calibration

4.1.1 Dynamic High-Frequency Warning

In multichannel music, calibration and reference are vitally important to the outcome of the project. Over the years in broadcast audio production, AES, SMPTE, and ITU standards have been adhered to for consistent and predictable audio performance. Within the music industry, however, mixing levels and practices have been somewhat arbitrary. Consistency in playback of various media, from movies to music, is highly desirable to consumers.

Given today's higher resolution multichannel audio, mixers need to be aware of specific sonic scenarios related to monitoring, mixing, and mastering. High-resolution audio formats have the potential to contain greater dynamic range, span a wider frequency spectrum, and are capable of reproducing loud high-frequency content not realized or reproduced in earlier audio formats. Because such content above 20 kHz can affect the high-frequency drivers of some loudspeakers, and potentially, human hearing, caution with respect to high-frequency distortion is highly recommended when monitoring.

When working with high-resolution audio, make sure to analyze the extended frequency spectrum of the program to identify anomalies that may be harmful to equipment or the listener.

The frequency spectrum analysis, as well as an overall system alignment check, should be done at the start of each project and periodically throughout the duration of the project.

4.1.2 Alignment Signal Level

1 kHz Sine Wave Alignment Level

Considerations:

- The goal of this manual is to produce repeatable reference listening experiences in different listening environments.
- AES, EBU, ITU and SMPTE standards differ regarding alignment levels.

Use –20 dBFS as the 1 kHz sine wave alignment signal level, per current multichannel DVD production standards.

The RMS level of all test signals (see Section 5.6) will be at **Alignment Signal Level**, that is, –20 dB with respect to dB full-scale (FS) digital level in digital devices [3]

A 1 kHz sine wave at this level typically produces +4 dBu (= 1.23 Vrms relative to 0 VU [8]) from professional consoles.

In the case of digital devices, the alignment level must be a 1 kHz sine wave 20 dB below the maximum possible coding level of the particular digital system, irrespective of the total number of bits available. The alignment levels for 16-bit audio systems are shown in Table 4-1.

Table 4-1 Digital Codes for 1 kHz Sine Wave Alignment Levels

Number of Bits	Audio Alignment Level	
	Negative Peaks	Positive Peaks
16	F333	0CCD

The values in Table 4-1 produce the indications shown in Figure 4-1 for various types of program meters.

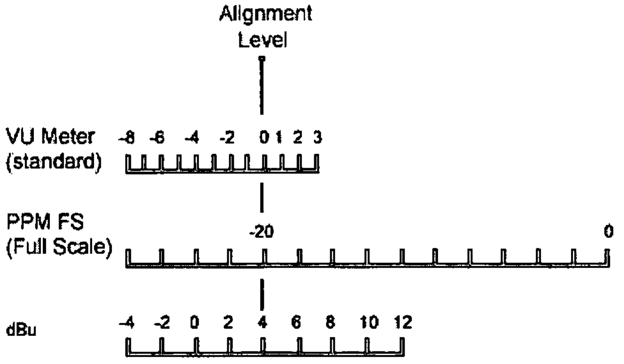


Figure 4-1 1 kHz Sine Wave Alignment Level Metering

Pink Noise Alignment Level

Because of the random nature of pink noise, peak program meter (PPM) readings can vary. The pink noise alignment signal level should be set to 0 on a VU meter after confirming the 1 kHz alignment settings. Again, this value usually produces +4 dBu (1.23 Vrms) on professional consoles.

4.1.3 Loudspeaker Alignment Level

Bass-Managed System Speaker Alignment

Because the subwoofer handles the lower frequencies for every channel in a bass-managed system, care must be taken to ensure a flat frequency response for the combined sub/satellite unit.

1. Feed wideband pink noise to the center channel at alignment level.
2. Adjust the level of the center speaker amp (point A in Figure 4-2) to 85 dBC in its operating range using an RTA.
3. Slowly raise the level of the subwoofer amp (point A in Figure 4-2) to achieve a flat frequency response on the RTA for the center/sub combination.

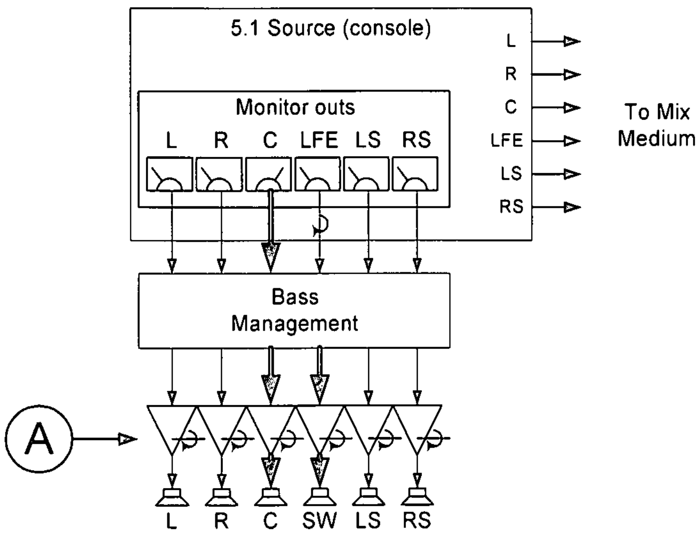


Figure 4-2 Bass-Managed Loudspeaker Alignment

After obtaining a flat frequency response from the bass-managed center/sub combination, the subwoofer and center amp settings (point A) should not be changed. Proceed with the remaining four channels (L, R, Ls, Rs) individually, adjusting their respective amps to a satellite/sub level of 85 dBC and a flat response.

LFE Alignment

The LFE channel should be calibrated in the following way for both bass-managed and non-bass-managed systems. If the system is bass-managed, however, align the LFE channel after adjusting the subwoofer level as described above.

1. Feed wideband pink noise to the LFE channel at alignment level.

2. Adjust the level of the console monitor feed (point B in Figure 4-3) to obtain an RTA reading of 10 dB in-band gain (average gain in the region 25 to 120 Hz) in relation to the same bandwidth measured for the center channel. See Figure 4-4.

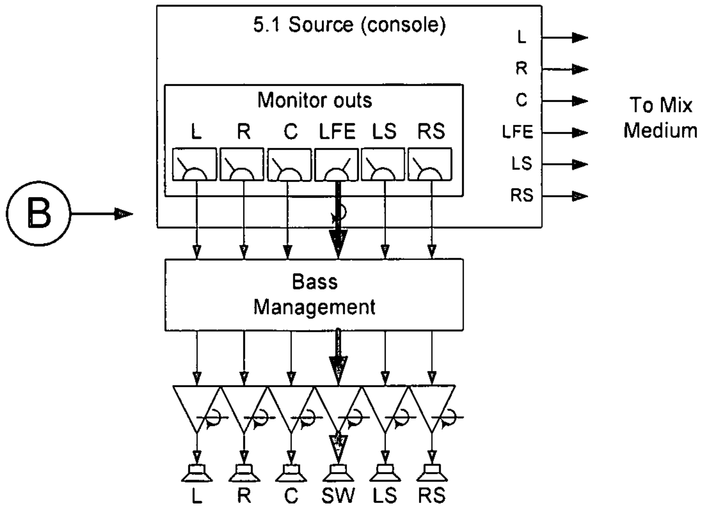


Figure 4-3 LFE Level Alignment

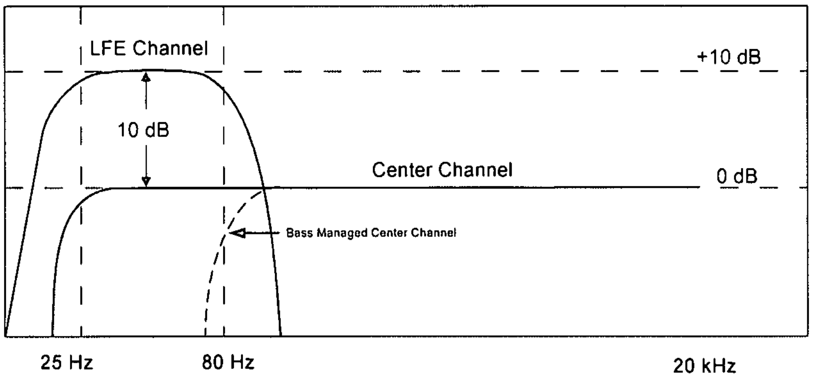


Figure 4-4 LFE RTA Display

Using this method to align the LFE channel ensures the accurate reproduction of the LFE channel commonly used for movie sound effects, and such.

If an RTA is not available, the LFE channel level can be set using the more commonly available SPL meter. Generally, when using wideband pink noise, the SPL reading (C-weighted, slow) should be approximately 4 to 5 dB above the main channels (that is, if the Center channel reads 85 dBC, then the LFE channel should read 89–90 dBC). Note that the SPL meter is taking a wideband measurement (not band-limited to 25–120 Hz) and therefore returns a lower value than the same measurement on a band-limited RTA.

If an RTA is available for the alignment procedure, record the SPL meter readings for that room for future alignments when an RTA is not available.

Alternatively, both the main speakers and LFE channel alignments can be done using specially prepared highpass and lowpass frequency-limited pink noise in which the highpass frequency-limited pink noise is used to adjust the main speaker levels and the lowpass frequency-limited pink noise is used to adjust the subwoofer level. In this case, the SPL will read the same 85 dBC for the LFE as it does for the main channel.

As discussed in Section 2.2, bass content can easily overload a recording system. This is the historical reason for the creation of the LFE channel and the practice of recording to tape 10 dB below reference level and boosting the playback 10 dB to achieve a flat response. In this way, plenty of bass can be included in the recording without causing tape saturation or distortion.

Because lower frequencies require more energy than higher frequencies to be perceived as the same level of loudness, recording energy-sapping deep bass on a separate track at a lower level to allow for a wider dynamic range is still good practice. For these reasons, plus the added bonus of legacy cinematic recording playback compatibility (both in the studio and the home), the +10 dB practice is still recommended.

4.1.4 Reference Listening Level

While Section 4.1.3 describes the correct loudspeaker alignment procedure, many listeners find the resultant level too loud for prolonged listening periods.

Several international listening standards dictate a reference listening level, L_{ref}, which is slightly lower than the loudspeaker alignment level and is determined by:

$$L_{ref} = 85 - 10 \log n \pm 0.25$$

where n is the number of reproduction channels in the total setup. The LFE channel is optional for 5.1-channel music mixing; therefore, n equals 5, making the individual channel level 78 dBC.

Using this level keeps the combined five channels, operating at reference level, at 85 dBC, which is well within the safety guidelines for occupational noise exposure [15].

4.2 Room Equalization

The biggest obstacle to obtaining a smooth room response in small- to medium-sized rooms is, by definition, the limited dimension in height, length, and width available to address bass nodes. Achieving the tolerances for proper room response curves shown in Section 3.1.2 can be challenging, especially in the low-frequency areas.

Use of an RTA is highly recommended to identify anomalies in the room response. Often, architectural or acoustical treatment of the room and/or adjustments to the loudspeaker and listening positions can solve many of the irregularities.

Equalization may be necessary when the linearity of the operational room response curve cannot be achieved by architectural means.

To avoid degrading the quality of reproduction, electrical equalization should be used carefully, and if possible, in the low-frequency range only.

4.3 Delay

The surround and front-channel signals should arrive at the listening position at the same time (coincident arrival); therefore,

- If the front loudspeakers need to be placed on a straight line base (moving the center speaker closer to the reference position), compensating time delay must be introduced in the signal feed of the center loudspeaker.
- If the surround loudspeakers are placed closer to the reference position than the front speakers, compensating time delays must be introduced in the signal feed of the Ls and Rs loudspeakers.

When necessary, delay should be added according to the following rule:

For each meter (foot) short of the reference distance, add 2.94 ms (.9 ms) of delay.

4.4 Bass Management

Note: Always check your mix using a bass-managed system.

In cinematic sound, explosions, earthquakes, and other high-energy, low-frequency special effects dictate the need for a dedicated subwoofer fed by the sixth, ".1" or LFE signal channel (see Section 2.2). Multichannel music may or may not make use of the LFE channel. Some music engineers feel that there is no need for the LFE channel at all. Because many of the popular consumer 5.1-channel speaker packages direct all of the bass from the main channels to the sub, however, combining it with any LFE information that may be present, it is important to at least check the effect bass management has on a mix while still in the studio. For example, in one known case, the combined bass from the main speakers was out of phase with the LFE channel, virtually eliminating the low end of the program when played back through a consumer, bass-managed system.

The goal of bass management is accurate bass reproduction in a given room—hearing all the bass frequencies in a smooth response in combination with the high-frequency drivers—and must be approached with caution and complete dedication to accurate calibration and monitoring.

As mentioned in Section 3.4.3, a low bass-management crossover frequency (80 Hz) gives more flexibility in subwoofer placement and helps keep its location audibly invisible.

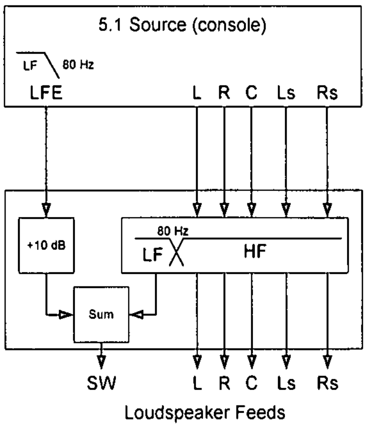


Figure 4-5 Bass Management

Bass-management allows the user to redirect low-frequency information from any of the five main speakers to the subwoofer. This is important since the five main speakers in home theater systems (satellite/sub speaker arrangements) are typically not designed to reproduce frequencies much below 80 Hz . Even though bass management is not required when monitoring in a studio with full-range speakers and a subwoofer, it is useful for checking how redirected low frequencies from any of the main channels may interact with the LFE-channel information. Remember that the consumer is likely to use some form of bass management, so proper bass management is necessary to emulate a consumer home theater system.

Historically, the LFE channel for cinema applications has a range that extends up to 120 Hz. Some less expensive consumer receivers offer only a fixed bass-management crossover frequency (often at 80 Hz) using one filter after the summation of the LFE and the main channel information (see Figure 4-6). This results in a situation where bass content between 120 and 80 Hz in the LFE channel is lost.

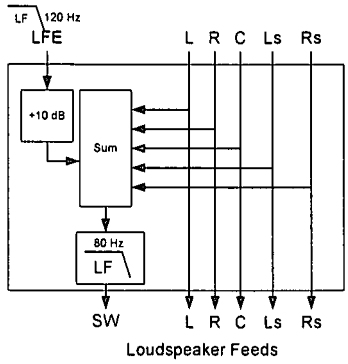


Figure 4-6 Worst-Case Consumer Bass Management

For this reason, it is recommended that the LFE channel be rolled off at 80 Hz during 5.1-channel music mixing.

The bass management feature on some consumer processors offers other combinations of redirection. For example, when the consumer does not have a subwoofer in the system, the LFE content can often be sent to the L and R main speakers. The 5.1-channel music producer may want to monitor the effects of these processes while still in the studio to ensure a desirable result.

Again, the goal in production and home playback is to hear all the bass accurately, regardless of which speaker reproduces it. When using full-range speakers with good bass response, there is theoretically no need for bass management, even when using a sub for monitoring the LFE channel. However, the track should still be listened to at some point with bass management engaged, to emulate what the consumer will hear at home.

4.5 Downmixing

The performance of a multichannel system under the conditions of two-channel playback should be tested using a reference downmix. Although the use of a fixed downmix may seem restricting, it covers the worst-case real-world scenarios that may be encountered.

In addition, time and budget may limit the ability to create a separate stereo mix, leaving the downmixed 5.1-channel as the only stereo option. In some circumstances, only a downmix is available (portable players, broadcasting, and so on); therefore, it is important to monitor the results while still in the studio. The equations for the reference downmix [10] are:

$$\mathrm{Lo} = 1.00\,\mathrm{L} + 0.71\,\mathrm{C} + 0.71\,\mathrm{Ls}$$
$$\mathrm{Ro} = 1.00\,\mathrm{R} + 0.71\,\mathrm{C} + 0.71\,\mathrm{Rs}$$

These downmix equations should not be confused with the more flexible systems offered by particular music formats, but as a quality check, the mix should be monitored as Lo, Ro.

Please note that since the equations are additive, it may be necessary to lower the overall level of the resultant downmix. Also, notice that the LFE channel may be disregarded in some circumstances, which should be kept in mind when deciding what content to include in the LFE channel during mixing.

4.6 Timecode

Timecode plays two important roles in preparing multichannel mixes. First, it is common to use some form of SMPTE or MIDI timecode for synchronizing recording machines and digital editors while recording and mixing material. It is important to know at the beginning of a project what the final timecode delivery format will be. Working in that format will save time later and prevent possible errors in frame rate and synchronization. This is especially true when working with video. With the introduction of high-definition video formats worldwide, even more varying frame rates and timecode modes (including drop-frame) now exist.

Second, if the material is going to accompany video at some point in the future, timecode is needed to enable a time stamp that is used to synchronize audio with the video, essential for proper audio/video synchronization.

Having the correct timecode that matches the picture is essential to proper audio and video synchronization. In addition, it is also very important that the timecode be stable and uninterrupted. Always use the timecode generated from a digital source such as a digital VTR. If unsure of the timecode source, generate clean timecode from a synchronizer or use a quality timecode regenerator.

In all cases, the associated audio and video equipment should all be clocked to a single master clock (see Section 5.3) to avoid timing drifts, mutes, and clicks and pops in the signal path and resulting master.

4.7 Consoles

When deciding on a console, it is wise to consider both current project demands and future 5.1-channel production needs. The requirements of 5.1-channel mixing consoles differ significantly from those of two-channel stereo. Fortunately, with the increasing flexibility of analog and digital consoles, there are now many options for surround mixing. Film-style four-bus (Left, Center, Right, and Surround) consoles have been in production for many years.

The fundamental requirement, however, of a 5.1-channel mixing console is a minimum of six discrete output buses (Left, Center, Right, Left Surround, Right

Surround, and LFE) per input/output channel. As with four-channel production, the six-bus console must also provide a means of panning audio. A console with film-style panning between the five main channels (L, C, R, Ls, and Rs) and routing to an LFE channel offers the greatest flexibility of sound placement in the surround field. While most manufacturers provide channel bus and pan features within the console, third-party developers have created add-on outboard devices with mixing controls to properly route multichannel audio for 5.1-channel mixes. In addition, whether onboard or third-party, console parameter automation is an asset when completing complex multichannel mixes.

Analog mixing consoles with six bus outputs and automation can offer a capable solution for multichannel mixing but require an analog-to-digital conversion stage, since 5.1-channel audio is delivered in the digital domain

In contrast, digital mixing consoles offer a direct path to today's multichannel delivery systems. Many digital consoles provide format conversion and internal signal processing. This all-in-one design delivers greater efficiency and flexibility during production. However, unwanted signal delays created by linking and chaining digital effects in digital audio consoles can cause timing errors between output channels. For instance, Left and Right audio channels may be configured to pass with relatively little processing along their signal paths. If the Center channel signal were to undergo processing such as digital EQ or dynamic compression, a measurable output signal delay could occur between the Center and Left/Right channels. Acknowledging this, digital console designers and manufacturers have created a variety of solutions to these timing problems.

4.8 Small Format Consoles

There has been an explosion in the marketplace of small professional consoles with enough buses to handle 5.1-channel mixing. While the implementation of multi-bus panning may be different, the functionality and setup of analog consoles and the newer consumer digital consoles are basically the same. Instead of panning in stereo, consoles with four or more auxiliary sends (in addition to the stereo bus prefader) can be used to place sounds anywhere in the 5.1-channel soundfield.

Currently, there are several excellent small digital consoles on the market. These consoles handle five-channel panning in different ways, with different features on the various models. Key features of some of these consoles include:

- Center mix level: allows adjustable panning through the center channel.
- Surround pattern editor: allows the path of the surround pan to be changed. Select the size and shape of a circle, arc, or line.
- Jog wheel speed manipulation: allows use of the wheel to change the speed of a pan.
- Multiple surround formats: 2+2, 3+1, 3+2+1.

- Master fader can be made into six-channel ganged fader.
- Divergence control on L/C/R and Front/Rear: these controls focus or spread the sound by controlling the bleed to adjacent channels during a pan.
- Surround Bus Assignment: in 5.1-channel mode.
- LFE level: controls the amount of signal going into the LFE channel.
- Surround Bus Isolation: Surround buses 1–8 are automatically solo-isolated so that they are not muted when any (surround-pan enabled) channel is in solo mode.
- Software control includes morph function that can interpolate between two points.
- On-screen editing of surround panning.
- Copy control of surround pans between channels.

Most console types put the surround mix through either the output bus path, or, in some cases, through auxiliary bus paths without separate multichannel monitoring paths. To have constant levels to the multitrack recorder as well as adjustable monitoring volume, it is necessary to bring the outputs of the multitrack back onto six grouped faders on the console and out a separate path to the amps and speakers. For the monitor outputs, use either auxiliary sends or a separate I/O card.

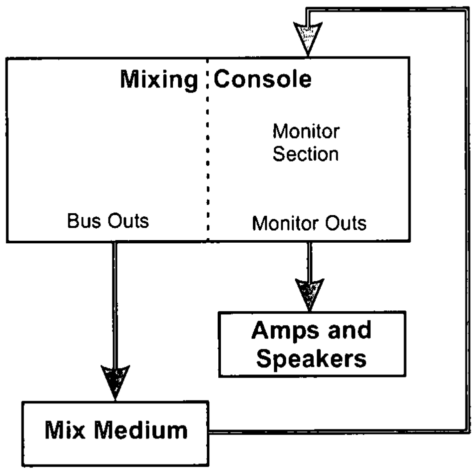


Figure 4-7 Console Interconnect Example

In Figure 4-7, arrows refer to six-channel buses, and should be connected using the following channel assignments, if possible: (1) Left, (2) Right, (3) Center, (4) LFE, (5) Left Surround, and (6) Right Surround. Choice of mix outputs (Bus, Aux, Monitor, and so on) depends on the console.

Currently, mixing-surface remotes that act as interfaces for software driven systems are getting more popular (Figure 4-8). Given the streamlining of this technique, there is less need to have additional digital I/O, sample rate converters, and so forth, while maintaining data integrity and audio fidelity in one host workstation.

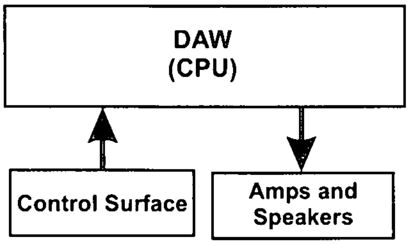


Figure 4-8 Digital Audio Workstation Interconnect Example

Note that some interface boxes for these types of digital audio workstations (DAWs), as well as some consoles, offer outputs at both +4 dBu and –10 dBV. Make sure that all of the monitor outputs are of the same level.

Chapter 5
Tips for Mixing 5.1-Channel Music

5.1 About 5.1-Channel Mixing

5.1-channel mixing offers an unparalleled level of creative flexibility. Use of the center and rear speakers, however, continues to be a topic of debate among artists, producers, engineers, and other serious listeners.

It is expected that genres of perspectives will develop over time and that mixes may someday be grouped into categories such as "in the band," "in the audience," or other real or imaginary listening perspectives.

Knowing the history of 5.1-channel audio will help in the future creation of a successful mix.

5.1.1 The Center Channel

If you want an image in the front center, use the Center channel. This is not to imply that hard panning or bussing to the Center channel only is recommended, but rather that it has been demonstrated [13] that a phantom center image delivers less clarity and a frequency response different from that delivered by a hard Center speaker. Using a combination of phantom and hard Center generally delivers the smoothest frontal image. However, when mixing to a combination of speakers, consider the phase relationships imposed by signal processors in use in some channels but not in others. If these signals are ever to be downmixed, unintended signal cancellations or comb filtering may result.

5.1.2 The LFE Channel

Since the reproduction of the LFE channel can sometimes be considered optional, essential low-frequency information should not be mixed exclusively to the LFE channel. In fact, in most downmixing situations, the LFE is completely disregarded. Conversely, in some rare cases, there may be a consumer playback system with small speakers, incapable of deep bass reproduction, yet lacking the bass management to direct the main channel bass to the subwoofer. In such cases, the content of the LFE channel is all the bass that will be heard.

When mixing to the LFE channel, it is important to band-limit the content at 100 Hz (see Section 4.4).

5.1.3 The Surround Channels

Even though it is recommended that five identical speakers be used for 5.1-channel music production, there are many installed consumer systems that use dipoles for the surround channels. Dipoles should be set up, calibrated, and used to listen to mixes, just to experience what consumers with these systems hear.

5.2 Metadata

Most new audio delivery formats allow the inclusion of information that describes, and in some cases, controls many aspects of the reproduction. The following are just some of the recommended fields that should be noted during various stages of production and, if possible, included in the metadata channel of the particular delivery format.

5.2.1 Informational Metadata

Informational metadata describes the content. Just a small sampling may include:

- The reference mixing level
- Room type/size
- Copyright information
- Artistic and production credits
- Identifying numbers (International Standard Recording Code [ISRC], etc.)
- Album title, track title
- Track number, disc number
- Label
- Genre
- Beats per minute

5.2.2 Control Metadata

Control metadata contains parameters that can be acted upon by a system capable of reading the information. Control metadata allows the production team to take control of and optimize how its audio program will be reproduced in different home listening configurations and environments. Examples of control metadata include:

- **Level normalization**: This is a measurement of the average loudness over time of a typical section of the program and can be specified as a dB level below full-scale digital. The film industry has extensive experience with this parameter. It is newer to the broadcast and music industries, but should be noted as a tool for matching program levels.

- **Downmixing**: As mentioned in Section 4.5, many formats enable automatic creation of downmixed versions of 5.1-channel material based on parameters set by the producer. Some systems offer a wide range of controls, including flags that limit downmixing capability.
- **Dynamic range**: For circumstances that require playback with varying amounts of limited dynamic range, this parameter offers the producer a chance to specify how the music should be controlled.

It is important to monitor the effects of various control metadata values while still in the studio to confirm the desired effect.

Obviously, each delivery format has its own metadata system and capabilities, but it is highly recommended that metadata of both informational and control natures is captured and monitored as early as possible—ideally, at the mixing stage—to help enable a more accurate and efficient production flow.

5.3 Master Clock

In all instances of multichannel digital audio production, every piece of gear should be synchronized to a master clock source. The absence of a master clock during sampling/recording, playback, editing, transferring, and so forth can lead to a number of disasters including phase issues, jitter and clocking errors, noises, pops, and ticks.

5.4 Sample Rate Conversion

Currently, music is released in a wide variety of formats including Red Book CD, DVD-Video, DVD-Audio, and low bandwidth codecs. These formats support a wide variety of sample rates and bit depths including 48/96/192 kHz and 44.1/88.2/176.4 kHz at 16-, 20-, or 24-bit resolution. Given that a number of projects require sample rate conversion, it is recommended that, in the interest of data integrity and fidelity, integer-based sample rate conversion be utilized when SRC is necessary. For a higher resolution project that is also slated for release on CD, the lowest common denominator is CD at 44.1 kHz at 16 bits. Ideally, the master recording exists at 176.4 kHz (24 bits) or 88.2 kHz (24 bits) allowing for precision integer down sampling to 44.1 kHz (for example, 88.2 kHz ÷ 2 = 44.1 kHz in conjunction with dithering 24-bit to 16-bit audio). Additionally, the same 176.4-kHz or 88.2-kHz master can be further down sampled to 44.1-kHz or 22.050-kHz and dithered as needed for low bandwidth codec delivery.

The converse, however, is discouraged. That is, because a number of 44.1-kHz and 48-kHz sampled digital masters currently exist, it is tempting to zero pad odd samples and integer upsample to be able to claim 88.2-kHz/176.4-kHz or 96-kHz/192-kHz output, respectively. Because the resultant audio stream does not contain additional audio data and is not of higher quality then the original, it is recommended that the music simply be released in its original sample rate (fs) and bit depth.

There are special circumstances where upsampling does produce a higher quality end result. For instance, upsampling a multitrack **before** digitally mixing has the advantage that any processing and/or effects added in the digital domain will be at the higher resolution. In this case, the end result **is** better then the original and is encouraged.

5.5 Documentation

Complete, clear, and accurate documentation should always accompany the source delivery master. This information is important not only when the master is in use but also as a reference, once it is archived. Dolby has created *Mix Data* and *Mastering Information* sheets to facilitate proper documentation or to use as a guide for creating similar documents. These sheets are available in Appendix A and at www.dolby.com under Technical Information. The *Mix Data* sheet provides concise information about the source media to all the engineers on a project. Typically, it includes information on sampling frequency, bit resolution, timecode, track assignment, titles, and program start and stop times. The *Mastering Information* sheet provides documentation relevant to the mastering engineer or authoring facility on source media, timing, and encoder settings, as well as general notes.

5.6 Test Signals

A 30-second, 1-kHz alignment signal at –20 dBFS, and a 30-second, 100 Hz alignment signal also at –20 dBFS should appear on all channels at the beginning of the source delivery master prior to program start. Two minutes of wideband pink noise, also at –20 dBFS (see Section 4.1.2) should be included on all channels to allow for the reference listening level alignment. The finished master should contain at least 30 seconds of digital black after the alignment signals and before each subsequent program. If appropriate, each title should begin with at least two seconds of encoded digital black.

For proper transport control in DVD-V and DVD-A, identification of intended index points/start times of the song for interpretation in the DVD authoring software (I and P frames) is required. See Appendix A.

Channel IDs and a polarity check mechanism would also be beneficial.

5.6.1 Mix Data Sheet

The *Mix Data* sheet provides production and mastering engineers with concise yet thorough media layout information. The information contained in the *Mix Data* sheet should be created during or after the multichannel master media is mixed, prior to mastering and encoding.

The *Mix Data* sheet should be used in conjunction with the *Mastering Information* sheet to provide necessary parameters prior to encoding for final delivery medium (that is, DVD-Video, DVD-Audio, DTV broadcast, etc.). All mix data information should be duplicated and placed on the master media as well. While recording and production media types may vary, accurate labeling of the media with *Mix Data* sheet information provides additional engineers with the proper origin knowledge.

5.6.2 Mastering Information Sheet

The *Mastering Information* sheet provides the mastering engineer or the digital authoring specialist technical information on media layout, timing information, and encoder specifics for each of the delivery mediums (that is, DVD-Video, DVD-Audio, DTV broadcast, etc.). If there is more than one delivery medium to address, each should receive a *Mastering Information* sheet. The information contained in the *Mastering Information* sheet is created both before the mastering process (RECOMMENDATION status indicated) and duplicated during the authoring/mastering/creation (FINAL MASTER status indicated) of the final delivery medium, (for example, Dolby® Digital or MLP Lossless™ bitstream for use on DVD-Audio) to confirm final selection of parameters.

Additional documentation, such as production notes, is invaluable in completing a project. Notes provide engineers with an explanation for key actions with relation to time, level, error, artistic consideration, and downmix-specific parameters. In addition to hard copy for each output medium, all documentation should be duplicated and affixed to the appropriate master delivery media (i.e., DLT tape, DVD-R, etc.).

5.7 Program Interchange

5.7.1 Channel-to-Track Allocation

Dolby encourages the adoption of channel-to-track allocation described in ITU-R BR.1384 Recommendation, *Parameters for International Exchange of Multi-channel Sound Recordings* and SMPTE 320M-1999, *Channel assignments and levels on multichannel audio media(Standard Assignment A).*

Track layouts depend on channel complement, although tracks 1, 2, and 3 are always channels Left (L), Right (R), and Center (C), respectively. Table 5-1 shows possible configurations. When the LFE channel is not used, track 4 may contain a mono Surround (MS) signal. Additionally, tracks 7 and 8 can be utilized for corresponding Lt/Rt or Lo/Ro stereo material. Alternative practices exist within various industries, so it is imperative to check the source and accompanying documentation.

Table 5-1 Channel/Track Allocation

	Track							
Format (channels)	1	2	3	4	5	6	7	8
3/1 (four-track)	L	R	C	Ms				
3/1 (alt)	L	R	C		Ms (–3dB)	Ms (–3dB)		
3/2.1 (5.1)	L	R	C	LFE	Ls	Rs		
3/2.1 (5.1) + 2/0 [14]	L	R	C	LFE	Ls	Rs	Lt or Lo	Rt or Ro

Where:

L	= Left channel	A	= 2/0 Left channel
R	= Right channel	B	= 2/0 Right channel
C	= Center channel	Lt	= Matrixed Left
LFE	= Low-Frequency channel	Rt	= Matrixed Right
Ls	= Left Surround channel	Lo	= Left only
Rs	= Right Surround channel	Ro	= Right only
Ms	= Mono Surround		

When recording additional multichannel programs onto a carrier with more than eight tracks (such as a 24-track recorder), groups of eight (that is, tracks 9–16 or tracks 17–24) should be preserved, as shown in Table 5.1.

Appendix A
Mix and Mastering Data Sheets

Mix Data

Date ____/____/________

Project ______________________ **Project #** ______________________

Client ______________________ **Producer** ______________________

Studio ______________________ **Engineer** ______________________

Sampling Frequency ☐ 44.1 kHz ☐ 48 kHz ☐ ×2 ☐ ×4

Bit Resolution ☐ 16-bit ☐ 20-bit ☐ 24-bit

Timecode Format ☐ 25 fps ☐ 29.97 fps ☐ 30 fps ☐ DF ☐ NDF

Media Format ☐ ______________

Surround Level SPL Calibration ☐ Equal to Front ☐ –3 dB to Front ______ dB SPL/channel

Tones ☐ 1 kHz @______dBFS ☐ 100 Hz @____dBFS

Downmix data ☐ See Notes ☐ 100 Hz @____dBFS

Channel Configuration

CH1	CH2	CH3	CH4	CH5	CH6	CH7	CH8

Timecode	Program
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	

Notes: ____

____

____

____

Mastering Information

Date ____ / ____ / ________

Project	________	**Project #**	________
Client	________	**Producer**	________
Studio	________	**Engineer**	________

Mastering Status	☐ Recommendation	☐ Final Master			
Sampling Frequency (fs)	☐ 44.1 kHz	☐ 48 kHz	☐ ×2	☐ ×4	
Bit Resolution	☐ 16-bit	☐ 20-bit	☐ 24-bit		
Timecode Format	☐ 25 fps	☐ 29.97 fps	☐ 30 fps	☐ DF	☐ NDF
Media Format	☐ ________				

Tones	☐ 1 kHz @ ______dBFS		☐ 100 Hz @______dBFS
Coding	☐ MLP Lossless™		☐ Dolby® Digital
Timecode Control	☐ No	☐ Yes	☐ Notes: In/Out/Duration
Downmix data	☐ See Notes		

Dolby Digital Encoding Information	**Bitstream Information**	**Processing**
Metadata Parameters LFE Filter ☐ ON ☐ OFF Dialogue Level: _____ Mix Level: _____ Data Rate: _____ Channel Mode: ______ Encoder Used: _______ Software Version: _____	Bitstream Mode: __________ RF Overmod Protection: _____ ☐ Copyright Bit Center Downmix Level:_____ Surround Downmix Level: _____ ☐ Extended Bitstream ☐ Dolby Surround EX Preferred Downmix _______ Lt/Rt C Downmix Lvl: ______ Lt/Rt S Downmix Lvl: _______ Lo/Ro C Downmix Lvl: _______ Lo/Ro S Downmix Lvl: _______ ☐ HDCD converter	☐ Digital De-Emphasis ☐ DC Highpass Filter ☐ Bandwidth Lowpass ☐ LFE Lowpass Filter
Surround Channel Processing ☐ 90-Degree Phase Shift ☐ 3 dB Attenuation	**Line Mode/RF Mode Preferences** ☐ None ☐ Speech Film: ☐ Standard ☐ Light Music: ☐ Standard ☐ Light	

Channel Configuration

Mode	CH1	CH2	CH3	CH4	CH5	CH6	CH7	CH8

Timecode	Program
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	
: : ;	

Appendix B
References

1. *AES Technical Council Document AESTD1001.1.01-10, Multichannel surround sound systems and operations*, Audio Engineering Society, New York, NY

2. *EBU Technical Recommendation R22-1999, Listening conditions for the assessment of sound program material*, European Broadcasting Union, Geneva, Switzerland, 1999

3. *EBU Technical Recommendation R68-2000, Alignment level in digital audio production equipment and in digital audio recorders*, European Broadcasting Union, Geneva, Switzerland, 2000

4. *EBU Technical Recommendation R91-1998, Track allocations and recording levels for the exchange of multichannel audio signals*, European Broadcasting Union, Geneva, Switzerland, 1998

5. *EBU Technical Recommendation R96-1999, Formats for production and delivery of multichannel audio programs*, European Broadcasting Union, Geneva, Switzerland, 1999

6. *EBU Tech. 3276, Listening conditions for the assessment of sound program material: monophonic and two-channel stereophonic*, European Broadcasting Union, Geneva, Switzerland, 1998

7. *EBU Tech. 3276-E Supplement 1, Multichannel sound*, European Broadcasting Union, Geneva, Switzerland, 1999

8. *SMPTE RP 155-1997, Audio Levels for Digital Audio Records on Digital Television Tape Recorders*, Society of Motion Picture and Television Engineers, White Plains, NY, 1997

9. *ITU-R BR.1384, Parameters for international exchange of multichannel sound recordings*, International Telecommunication Union, Geneva, Switzerland

10. *ITU-R BS.775-1, Multichannel stereophonic sound system with and without accompanying picture*, International Telecommunication Union, Geneva, Switzerland, 1994

11. *ITU-R BS.1116-1, Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, International Telecommunication Union, Geneva, Switzerland, 1997

12. Mick Sawaguchi and Akira Fukada, NHK, *Multichannel Sound Mixing Practice for Broadcasting*.

13. *SMPTE RP 173-2002, Loudspeaker Placements for Audio Monitoring in High-Definition Electronic Production*, Society of Motion Picture and Television Engineers, White Plains, NY, 2002

14. *SMPTE 320M-1999, Channel Assignments and Levels on Multichannel Audio Media*, Society of Motion Picture and Television Engineers, White Plains, NY, 1999

15. *DHHS (NIOSH) Publication No. 98-126, Occupational noise exposure*, National Institute for Occupational Safety and Health, Washington, D.C., 1998