T.C
YEDİTEPE UNIVERSITY
GRADUATE INSTITUTE OF SOCIAL SCIENCES

CROSS-MODAL USAGE IN AUDIO VISUAL SYNCHRONIZATION FOR COMMUNICATION DESIGN

by

Mete Yafet

Submitted to the Graduate Institute of Social Sciences
In partial fulfillment of the requirements for the degree of Master of Visual Communication Design.

İSTANBUL, 2008

T.C
YEDİTEPE UNIVERSITY
GRADUATE INSTITUTE OF SOCIAL SCIENCES

CROSS-MODAL USAGE IN AUDIO VISUAL SYNCHRONIZATION FOR COMMUNICATION DESIGN

by

Mete Yafet

Supervisor

Prof. Dr. Emin Doğan AYDIN

Submitted to the Graduate Institute of Social Sciences
In partial fulfillment of the requirements for the degree of Master of Visual Communication Design.

İSTANBUL, 2008

"CROSS-MODAL USAGE IN AUDIO VISUAL SYNCHRONIZATION FOR COMMUNICATION DESIGN"

by

METE YAFET

Approved by:

Prof. Dr. Emin Doğan AYDIN
(Supervisor) ....

Prof. Dr. Ahmet ORKAN ....

Prof. Dr. Melda Cinman ŞİMŞEK ....

Assist. Prof. Dr. Cüneyt BUDAK ....

Assist. Prof. Dr. Simge ESIN ....

Date of Approval by the Administrative Council of the Institude:
___/___/2008

TABLE OF CONTENTS

LIST OF SYMBOLS

Audio-Visual Synchronization .... AV-Sync

Hertz .... Hz

Kilohertz .... kHz

Sign Inverting .... ON

Decibel .... λ / dB

Meter per second .... m/s

Sound Pressure Levels .... SPL

American National Standards Institude .... ANSI

Minimum Audible Angle .... MMA

Inter-aural Time Differences .... ITDs

Inter-aural Level Differences .... ILDs

Millisecond .... ms

Direct Sequence Spread Spectrum .... DSSS

Narometer .... nm

Red Green Blue .... RGB

Cyan Magenta Yellow Black .... CMYK

Short-Medium-Long .... S M L

Lateral Geniculate Nucleus .... LGN

Cycles per degree .... CPD

Millimeter .... mm

Two Dimensional 2-D

Three Dimensional 3-D

Audio Video .... AV

Charge-coupled Device .... CCD

Complementary Metal-oxide Semiconductor .... CMOS

Liquid Crystal Display .... LCD

Digital Light Processing .... DLP

Moving Picture Experts Group .... MPEG

The Advanced Television Systems Committee .... ATSC

Digital Television .... DTV

Transport Video and Audio Standards for Broadcast Television MPEG-2
System Time Clock .... STC
Presentation Time Stamps .... PTS
Network Time Protocol .... NTP
Local Area Network .... LAN
Global Pointing System .... GPS
Audio Codec AC-3
Digital Video Disc .... DVD
Direct Broadcast Satellite .... DBS
Frame per Second .... fps
Internatonal Telecommunications Union .... ITU
National Television System Committee .... NTSC
Digital Video Effect .... DVE
Digital Delay Output .... DDO
Musical Instrument Digital Interface .... MIDI

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First of all, I’m very grateful to my mother Yıldız Köksalan for giving me the opportunity to study and continue my education in the academical field. I would like to thank her for improving me both personally and morally.

I especially want to thank Prof. Dr. Emin Doğan Aydın, who developed my skills and my knowledge in the visual communication area. I appreciate his leadership on my graduate education and my research over the entire course.

I’m very thankful to Assist. Prof. Dr. Simge Esin for helping me so much on arranging my thesis with her guidance and comments. Without her help it would definitely take much longer to complete.

I would like to thank İlkin Sarılgan for supporting me all the time and helping me throughout my research.

And also I would like to thank all my friends who are interested and working in the music production - visual design scene for keeping me updated on the topic and helping me out, solving all the problems I have experienced.

ABSTRACT

In todays modern platform of new visual communication, the usage of audio-visual media is an indispensable union. The viewers and users prefer the remarkable and memorable ones rather than the reasonable and necessary ones.

Now most of the music concert stages include large visual displays, even the smallest banner in a random site has a motion graphic or sound, the motion picture soundtracks are demanding as the movie itself, most of the music artists and bands have a DVD album on the market and potentially developed brands use LCD screen based audio-visual street advertisements rather than the graphics related billboard media that we use to know. All these designs and ideas are prepared with precise methods and in perfect synchronization for keeping the audience interested and keeping the design memorable.

This thesis researches how this union effects the human and the perception, as well as the usage of audio visual synchronization in media and visual communication.

The last section results the topic containing various tests and its effects about perception and response, run on groups of people with the examples of audio&visual elements in sync and out of sync.

Keywords: Sound,Vision, Visual,Communication, Visual Communication, Audial Communication, Synchronization, Asynchrony, Broadcast.

ÖZET

Günümüzün gelişen ve kullanımı her geçen gün daha da artan yeni görsel iletişim platformunda, ses ve görüntü içerikli medya kullanımı vazgeçilmez bir birliktelik olmuştur. İzleyici ve kullanıcılar artık daha fazla etki daha fazla ilgi döngüsü içinde cazip ve uygun olandan çok, çarpıcı ve akılda kalanı tercih etmektedirler. Konserlerde artık dev ekranlarda müziğe uygun hareketli ve renkli görseller kullanılmakta, internet ortamındaki en ufak reklam bile hareketli bir görüntü ve buna bağlı çıkan sesler içermekte, en az DVD ler kadar filmlerin Soundtrack'leride satılmakta, maddi olarak karşılayabilen her sanatçı ve grubun neredeyse bir DVD bazlı görüntülü müzik albümü bulunmakta, büyük markaların sokak reklam kampanyaları ve billboard larını artık slogan ve grafikten çok LCD paneller üzerinde oynayan sesli ve hareketli görüntülü reklamlar oluşturmaktadır. Buna bağlı olarak günümüz ve çevremizde bunun çok daha başka örneklerine rastlamak mümkündür. Bütün bu çalışmalar insanın nasıl çekileceği hedefinden yola çıkarak kusursuz ve akılda kalıcı bir uyum içinde üretilmektedir.

Bu tez çalışması vazgeçilmez bu birlikteliğiye ön ayak olan etkenleri,insanın bu birliktelik dahilinde nasıl ve ne şekilde etkilendiğini, algısının ne yönde değiştiğini, medyada ve görsel iletişimde bu konunun kullanımı ve ortamlarını araştırmaktadır.

Son bölümde ses ve görüntü içerikli medya örnekleri birlikte ve yalnız kullanılarak, senkronizasyonu çeşitli şekillerde düzenli ve düzensiz olarak değiştirilerek insan grupları üzerinde test edilmiş, algılar ve buna bağlı tepkiler incelinerek sonuçlar değerlendirilmiştir.

Anahtar Kelimeler: Ses, Görüntü, Algı, İletişim, Görsel İletişim, Görsel Medya, İşitsel Medya, Senkronizasyon, Asenkronizasyon, Yayın

1. INTRODUCTION

Communication design is a mixed discipline between design and information-development which is concerned with how media intermission such as printed, crafted, electronic media or presentations communicate with people. A communication design approach is more concerned with messages communicated than aesthetics in media. The distinction between communication design and other applied arts is in the motivation:while the communication design process does involve a certain amount of self-expression and creativity, the goals are often those of the commissioning body rather than the artist's, and the parameters set by the commissioning body are often more constraining. The term communication design is often used interchangeably with visual communication and more specifically graphic design, but has an alternate broader meaning that includes auditory communications as well as visual.

Nowadays, increasing variety of multimedia channels requires technical efforts, especially, in sound and visual synchronization. The aim is to show why synchronization is needed to be realized successfully by depending on scientific basis. So, negative effect which can occur on audience disappears and audience watch regularly what they are given.

This thesis is constructed with 3 main sections. In second section of the research physical topics, which are related to synchronization process, are explained under the title of main concepts. Audio visual synchronization is investigated in the third section with its basic components, specifications and technological problems. Fourth section contains experiments which are done in audio visual synchronization topic or in areas related closely to this topic.

The methods used for this study are literature survey on topics; Audio in Media and Broadcasting, also documental research on internet for the topics; Perception and Synchronization. The experiments are tested through the same group of individuals for understanding the effects of audio visual synchronization and usage of them.

2. AUDIO AND VISUAL SENSING OF HUMAN BEINGS IN TERMS OF PERCEPTION AND COMMUNICATION

2.1. The Fields of Communication

Communication is the exchange and flow of information and ideas from one person to another. It involves a sender transmitting an idea to a receiver. Effective communication occurs only if the receiver understands the exact information or idea that the sender intended to transmit.

> "Organizational communication establishes a pattern of formal communication channels to carry information vertically and horizontally. To ensure efficient and effective accomplishment of objectives, information is exchanged. Information is passed upward from employees to supervisors and laterally to adjacent departments. Instructions relating to the performance of the department and policies for conducting business are conveyed downward from supervisors to employees. The organization carries information from within the department back up to top management. Management furnishes information about how things are going, notifies the supervisor of what the problems are, and provides requests for clarification and help. Supervisors, in turn, keep their employees informed and render assistance. Supervisors continually facilitate the process of gaining necessary clarification and problem solving; both up and down the organization. Also, supervisors communicate with sources outside the organization, such as vendors and customers." (Allen, 1998)

2.1.1 The Process of Communication

Thought: First, information exists in the mind of the sender. This can be a concept, idea, information, or feelings.

Encoding: Next, a message is sent to a receiver in words or other symbols.

Decoding: Lastly, the receiver translates the words or symbols into a concept or information that he or she can understand.

> "During the transmitting of the message, two processes will be received by the receiver: content and context. Content is the actual words or symbols of the message which is known as language the spoken and written words combined into phrases that make grammatical and semantic sense.We all use and interpret the meanings of words differently, so even simple messages can be misunderstood. And many words have different meanings to confuse the issue even more. Context is the way the message is delivered and is known as Paralanguage - it includes the tone of voice, the look in the sender's eye's, body language, hand gestures, and state of emotions(anger, fear, uncertainty, confidence, etc.) that can be detected. Although paralanguage or context often causes messages to be misunderstood as we believe what we see more than what we hear; they are powerful communicators that help us to understand each other. Indeed, we often trust the accuracy of non-verbal behaviors more than verbal behaviors." (Clark, 2005)

Communication is the process of passing information and understanding from one person to another. The communication process involves six basic elements: sender (encoder), message, channel, receiver (decoder), noise, and feedback..Communication can break down at any one of these elements.

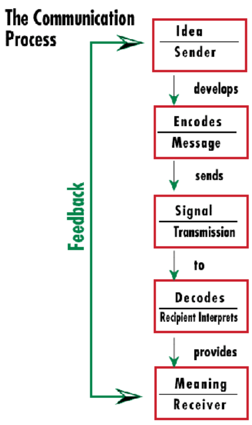

Figure 2.1 Communication process

(Allen, 1998)

2.1.2 Models of the Communication

i)Shannon's Model of the Communcation Process:

"Shannon's (1948) model of the communication process is, in important ways, the beginning of the modern field. It provided, for the first time, a general model of the communication process that could be treated as the common ground of such diverse disciplines as journalism, rhetoric, linguistics, and speech and hearing sciences. Part of its success is due to its structuralist reduction of communication to a set of basic constituents that not only explain how communication happens, but why communication sometimes fails." (Foulger, 2004)

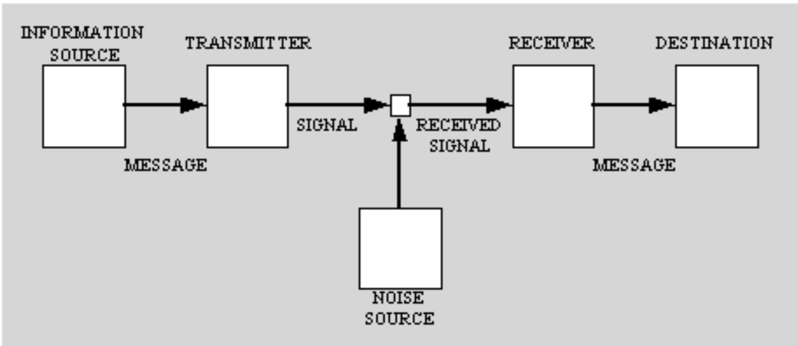

Figure 2.2 Shannon's (1948) Model of the communication process
(Foulger, 2004)

Shannon's model, as shown in Figure 2.2, breaks the process of communication down into eight discrete components:

"1)An information source. Presumably a person who creates a message.

2)The message, which is both sent by the information source and received by the destination.

3)A transmitter. For Shannon's immediate purpose a telephone instrument that captures an audio signal, converts it into an electronic signal, and amplifies it for transmission through the telephone network. Transmission is readily generalized within Shannon's information theory to encompass a wide range of transmitters. The simplest transmission system, that associated with face-to-face communication, has at least two layers of transmission. The first, the mouth (sound) and body (gesture), create and modulate a signal. The second layer, which might also be described as a channel, is built of the air (sound) and light (gesture) that enable the transmission of those signals from one person to another.

4)The signal, which flows through a channel. There may be multiple parallel signals, as is the case in face-to-face interaction where sound and gesture involve different signal systems that depend on different channels and modes of transmission. There may be multiple serial signals, with sound and/or gesture turned into electronic signals, radio waves, or words and pictures in a book.

5)A carrier or channel, which is represented by the small unlabeled box in the middle of the model. The most commonly used channels include air, light, electricity, radio waves, paper, and postal systems.

6)Noise, in the form of secondary signals that obscure or confuse the signal carried. Given Shannon's focus on telephone transmission, carriers, and reception, it should not be surprising that noise is restricted to noise that obscures or obliterates some portion of the signal within the channel. This is a fairly restrictive notion of noise, by current standards, and a somewhat misleading one. Today we have at least some media which are so noise free that compressed signals are constructed with an absolutely minimal amount information and little likelihood of signal loss.

7)A receiver. In Shannon's conception, the receiving telephone instrument. In face to face communication a set of ears (sound) and eyes (gesture). In television, several layers of receiver, including an antenna and a television set.

8)A destination. Presumably a person who consumes and processes the message."
(Foulger, 2004)

ii)A New Model of Communication Process: The ecological model of communication, shown in Figure 2.3, attempts to provide a platform on which these issues can be explored. It asserts that communication occurs in the intersection of four fundamental constructs: communication between people is mediated by messages which are created using language within media; consumed from media and interpreted using language.

> "This model is, in many ways, a more detailed elaboration of Lasswell's (1948) classic outline of the study of communication: "Who ... says what ... in which channel ... to whom ... with what effect". In the ecological model , the "who" are the creators of messages, the "says what" are the messages, the "in which channel" is elaborated into languages (which are the content of channels) and media (which channels are a component of), the "to whom" are the consumers of messages, and the effects are found in various relationships between the primitives, including relationships, perspectives, attributions, interpretations, and the continuing evolution of languages and media." (Foulger, 2004)

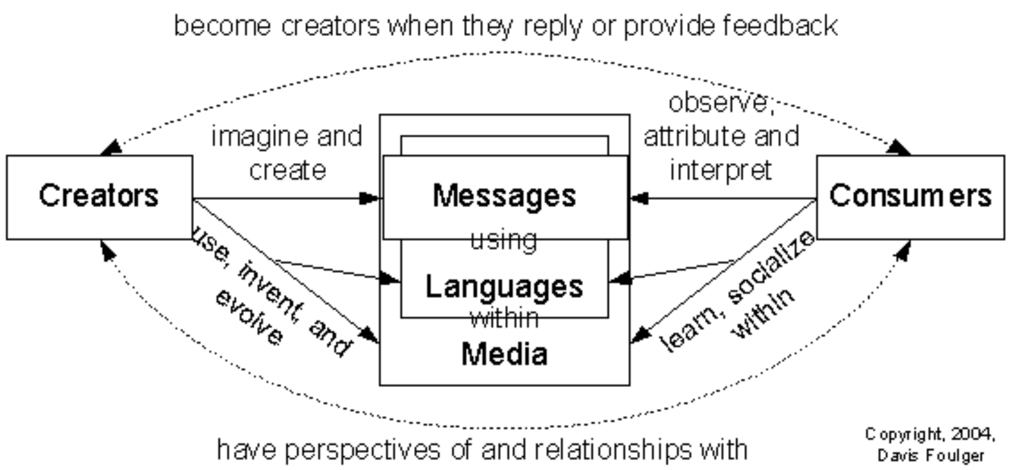

Figure 2.3 Ecological Model of Communication
(Foulger, 2004)

A number of relationships are described in this model as Foulger(2004) refers:

“1)Messages are created and consumed using language

2)Language occurs within the context of media

3)Messages are constructed and consumed within the context of media

4)The roles of consumer and creator are reflexive. People become creators when they reply or supply feedback to other people. Creators become consumers when they make use of feedback to adapt their messages to message consumers. People learn how to create messages through the act of consuming other peoples messages.

5)The roles of consumer and creator are introspective. Creators of messages create messages within the context of their perspectives of and relationships with anticipated consumers of messages. Creators optimize their messages to their target audiences. Consumers of messages interpret those messages within the context of their perspectives of, and relationships with, creators of messages. Consumers make attributions of meaning based on their opinion of the message creator. People form these perspectives and relationships as a function of their communication.

6)The messages creators of messages construct are necessarily imperfect representations of the meaning they imagine. Messages are created within the expressive limitations of the medium selected and the meaning representation space provided by the language used. The message created is almost always a partial and imperfect representation of what the creator would like to say.

7)A consumers interpretation of a messages necessarily attributes meaning imperfectly. Consumers intepret messages within the limits of the languages used and the media those languages are used in. A consumers interpretation of a message may be very different than what the creator of a message imagined.

8)People learn language by through the experience of encountering language being used within media. The languages they learn will almost always be the languages when communicating with people who already know and use those languages. That communication always occurs within a medium that enables those languages.

9)People learn media by using media. The media they learn will necessarilly be the media used by the people they communicate with.

10)People invent and evolve languages. While some behavior expressions (a baby's cry) occur naturally and some aspects of language structure may mirror the ways in which the brain structures ideas, language does not occur naturally. People invent new language when there is no language that they can be socialized into. People evolve language when they need to communicate ideas that existing language is not sufficient to.

11)People invent and evolve media While some of the modalities and channels associated with communication are naturally occurring, the media we use to communicate are not."

2.2. Perception of Human Beings

In psychology and the cognitive sciences, perception is the process of attaining awareness or understanding of sensory information.Perception is one of the oldest fields in psychology.

"The word perception comes from the Latin perception, percepio, meaning "receiving, collecting, action of taking possession, apprehension with the mind or senses."(OED, 2008) Psychophysics uses physical stimuli to provide quantitative measures of perceptual experience. Sensitivity is usually estimated in a threshold, the stimulus level at which a subject's responses shift from one perceptual experience to another, but these are subject to bias." (Mather, 2006)

As you know even sharing the same experience or learning the same data can show varieties of understandings and reactions in different people. These individual differences are based several factors. Experimental professor of psychology, Mather G (2006) explains these as:

> “Several factors may produce consistent individual differences in perception, including age, sex, culture, and expertise. Most perceptual functions improve rapidly early in life and then decline gradually in later years. This pattern mirrors the change in brain weight over the lifespan. Female advantages for tone sensitivity, taste sensitivity, odour recognition, and touch recognition have been reported, although males consistently outperform females on tasks of spatial vision. There is considerable variability and overlap in the performance of the two sexes. The influence of culture on perceptual capabilities has been investigated using pictorial competence tests and illusion studies. However, studies have not controlled for sources of bias such as experimental instructions, so evidence for cultural differences in perception is mostly equivocal. Formal training in music and art has a measurable effect on perception. Practice has also been shown to result in considerable improvements even in simple sensory tasks. Expertise can result in physical changes in brain structure.When the effects of age, gender, and expertise are controlled for, there may still be idiosyncratic differences in perception between individuals.”

2.3. Sound and Audial Perception

2.3.1 Ear and Its Anatomy

The ear is the sense organ that detects sounds. It not only acts as a receiver for sound, but plays a major role in the sense of balance and body position. The ear has 3 main parts. These are: Outer Ear, Middle Ear and Inner Ear as shown in Figure 2.4 given below:

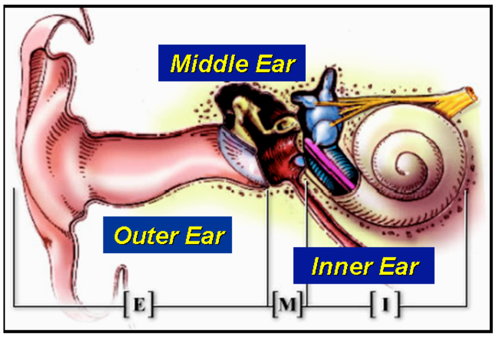

Figure 2.4 Anatomy of Ear
(Anatomy of Ear and Eye, 2008)

i)Outer Ear: 1- Pinna (the feather) matches ear canal to outside world. 2- Meatus (the passageway) conducts sound into head. 3- Tympanium (the drum) transforms pressure fluctuations into displacement.

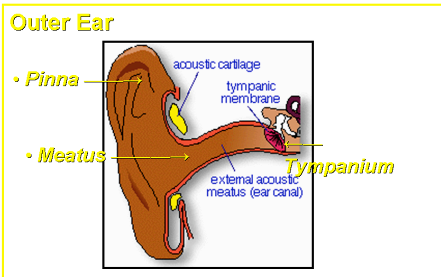

Figure 2.5 Outer Ear
(Anatomy of Ear and Eye, 2008)

ii)Middle Ear: 1-Malleus (the hammer) transmits the sound vibrations from the eardrum to the incus. 2- Incus (the anvil) supported by ligaments that protect against loud. 3- Stapes (the stirrup) transmits the sound vibrations from the incus to the membrane of the inner ear.

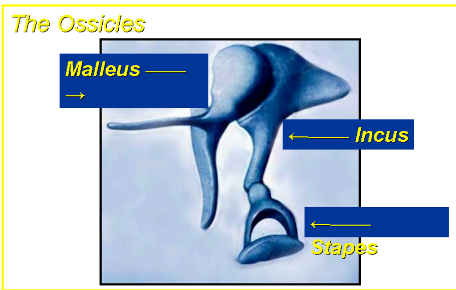

Figure 2.6 Middle Ear

(Anatomy of Ear and Eye, 2008)

iii)Inner Ear: 1-Cochlea (the snail) converts displacement into neural impulses. 2-Auditory Nerve, neural impulses to brain. 3- Semicircular canals detect motion and orientation.

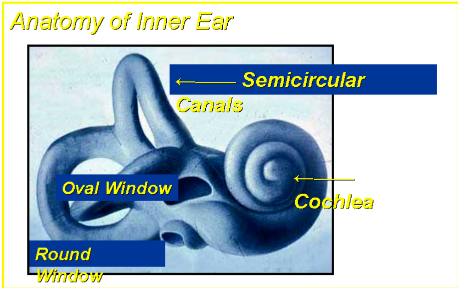

Figure 2.7 Inner Ear

(Anatomy of Ear and Eye, 2008)

“The outer part of the ear collects sound.That sound pressure is amplified through the middle portion of the ear and, in land animals, passed from the medium of air into a liquid medium.The change from air to liquid occurs because air surrounds the head and is contained in the ear canal and middle ear, but not in the inner ear. The inner ear is hollow, embedded in the temporal bone, the densest bone of the body. The hollow channels of the inner ear are filled with liquid, and contain a sensory epithelium that is studded with hair cells. The microscopic "hairs" of these cells are structural protein filaments that project out into the fluid. The hair cells are mechanoreceptors that release a chemical neurotransmitter when stimulated. Sound waves moving through fluid push the filaments; if the filaments bend over enough it causes the hair cells to fire. In this way sound waves are transformed into nerve impulses.

The human ear can generally hear sounds with frequencies between 20 Hz and 20 kHz (the audio range). Although the sensation of hearing requires an intact and functioning auditory portion of the central nervous system as well as a working ear, human deafness (extreme insensitivity to sound) most commonly occurs because of abnormalities of the inner ear, rather than the nerves or tracts of the central auditory system.” (Greinwald&Hartnick, 2002)

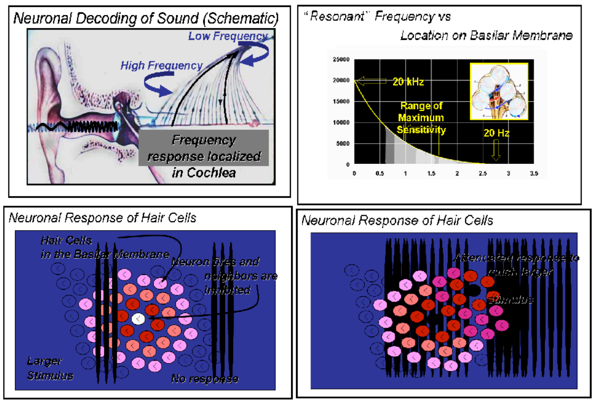

Figure 2.8 Neuronal Response of Hair Cells
(Anatomy of Ear and Eye, 2008)

2.3.2 Qualities and Specifications about Sound

Sound have different related concepts which makes up the basis of this research area. The most needed and basic information about the related concepts are given below:

> “Sound is defined as pressure differences which human ear can perceive in situations such as water, solid, air. If sound wave is in pure sinus form, sound wave has only a frequency. Multiple waves are called as complex sound waves.” (Uyan, 2008).

Sounds have different specifications:

i)Simple Sounds:

> **“**Sound consists of pressure waves carried by vibrating air molecules. Parts of the wave where air pressure is increased are called compressions and parts where pressure is decreased are called rarefactions. A sound wave's frequency corresponds to the number of alternations between compression and rarefaction in a 1-second period, i.e. the number of cycles per second (measured in Hz). High-frequency

waves are perceived as high pitch and low-frequency waves as low pitch.” (Mather, 2008)

The graph in Figure 2.9 plots the sinusoidal sound waveform of a pure tone.

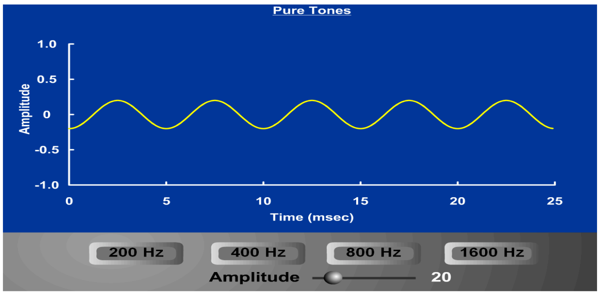

Figure 2.9 Simple sounds wave
(Mather, 2008)

ii)Complex Sounds:

“Any complex sound can be treated as a collection of simple sine waves added together. The frequencies, amplitudes and phase of the individual waves determine the overall form of the complex wave and the sound it makes. The lowest frequency in a complex wave is called the fundamental frequency. Harmonic frequency are higher frequency components that are numbered according to their distance from the fundamental frequency, e.g. the fifth harmonic has a frequency that is five times greater than the fundamental frequency. Many natural sounds do not contain harmonics, but have a continuous "spectrum" of components in which all frequencies are represented.” (Mather, 2008)

The graph in Figure 2.10 plots the sinusoidal sound waveform of a complex tone.

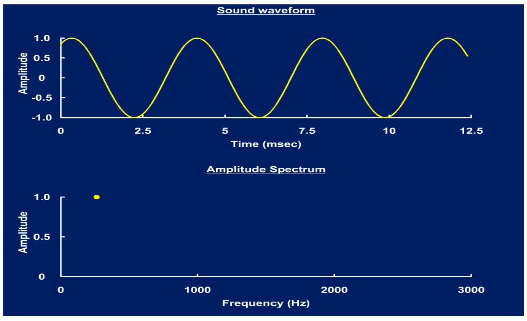

Figure 2.10 Complex sounds wave

(Mather, 2008)

iii)Standing Waves:

“When a traveling wave reflects back on itself, it creates traveling waves in both directions.The wave and its reflection interferes according to the superposition principle. With exactly the right frequency, the wave will appear to stand still. This is called a standing wave. A node occurs where the two traveling waves have the same magnitude of displacement, but the displacements are in opposite directions. Net displacement is zero at that point. The distance between two nodes is a $\lambda/2$(Half Decibel). An antinode occurs where the standing wave vibrates at maximum amplitude.” (Wisconsin, 2005).

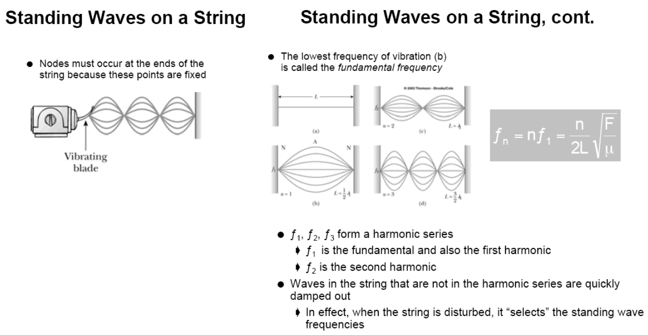

Figure 2.11 Standing waves

(Wisconsin, 2005)

iv)Sound Velocity:A sound cannot achieve a distance in space, it has to exist in an enviroment to travel.The speed of sound in some enviroments are shown in Table 2.1:

Table 2.1 Sound Velocity (Tannenbaum, B.Stillman, 1973)

Various Area (0°C)	(m/s)
Air	332
Water	1454
Wood	3828
Ferrous	5103

Stone	5971

v)Decibels: Decibels are defined as a logarithmic unit of sound intensity. Mathematically a decibel (dB) is defined as: λ

The decibel is useful for a wide variety of measurements in science and engineering (e.g., acoustics and electronics) and other disciplines. It confers a number of advantages, such as the ability to conveniently represent very large or small numbers, a logarithmic scaling that roughly corresponds to the human perception of, for example, sound and light, and the ability to carry out multiplication of ratios by simple addition and subtraction.

vi)Intensity vs. Loudness:

> **"**Sound intensity is a measurable quantity relating to acoustic energy whereas loudness is a subjective quality affected by the human auditory response to sound. The Sonic Research Studio at Simon Fraser University defines sound intensity as; the sound energy transmitted per unit time through a unit area, thereby being a measure of the magnitude of a sound. This magnitude of sound is measured in Sound Pressure Levels (SPL)." (Bengert and Upward, 2003)

vii)Pitch: When referring to sound and music the term pitch has two common uses. Pitch is considered an attribute of the human auditory system as well as a synonym for the term frequency. Although these two definitions may seem incongruous, it is possible to make a connection between the two definitions.

> "The ANSI definition of psychoacoustical terminology considers pitch the auditory attribute of sound according to which sounds can be ordered on a scale from low to high. This definition envelops both usages. The human auditory system makes distinctions between frequencies (pitches) and can order them on a scale from low to high." (Bengert and Upward, 2003)

> "Pitch is related mainly, although not completely, to the frequency of the sound. Pitch is not a physical property of the sound. Frequency is the stimulus and pitch is

the response. It is a psychological reaction that allows humans to place the sound on a scale.” (Wisconsin, 2005)

viii)Timbre:

“Timbre is defined in many ways. The ANSI defines timbre as, that attribute of auditory sensation in terms of which a listener can judge that two sounds, similarly presented and having the same loudness and pitch, are different. Musically, timbre is defined as the quality of a musical note which distinguishes different types of musical instrument. And finally, timbre is defined as everything that is not loudness, pitch, or spatial perception. Those three definitions cover the concept of timbre very well. When two sounds of the same loudness and pitch are played we can distinguish between them. Everyone will agree that a flute and a saxophone playing the same song don’t sound the same. This is due to the difference in timbre of each instrument. Timbre can be based on physical characteristics of instruments, such as airflow, embrasure, and many others.” (Bengert and Upward, 2003)

In music, the characteristic sound of any instrument is referred to as the quality of sound, or the timbre of the sound. The quality depends on the mixture of harmonics in the sound. Figure 2.12 shows the wave of timbres for various musical instruments:

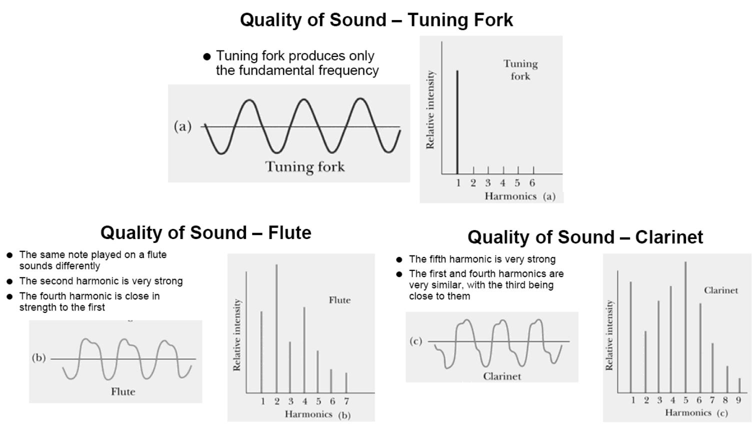

Figure 2.12 Timbre of the Musical Instruments
(Wisconsin, 2005)

ix)Resonance:

> "In physics, resonance is the tendency of a system to oscillate at maximum amplitude at certain frequencies, known as the system's resonance frequencies (or resonant frequencies) system. When the frequency of the driving force equals the natural frequency of the system, the system is said to be in resonance." (Wisconsin, 2005)

x)Doppler Effect:

The Doppler effect, named after Christian Doppler, is the change in frequency and wavelength of a wave as perceived by an observer moving relative to the source of the waves. For waves that propagate in a wave medium, such as sound waves, the velocity of the observer and of the source are relative to the medium in which the waves are transmitted.

xi)Auditory Localisation:

> "The direction of a sound source in the horizontal plane is specified by an azimuth angle relative to straight ahead. The minimum audible angle (MAA) is the smallest change in the azimuth angle of a sound source that can be detected. According to duplex theory, low-frequency sounds are localised using inter-aural time differences (ITDs) and high-frequency sounds are localised using inter-aural level differences (ILDs). Wightman and Kistler (1992) demonstrated that for complex natural sounds such as speech, ITDs in lower frequencies dominate the computation of the azimuth angle. A cone of confusion , describes a cone of space extending from a listener's head in which different locations (e.g. directly in front or behind the head) result in the same ITDs. Listeners can compensate for this by moving their heads to reduce the cones of confusion and/or by using cues in the vertical plane." (Mather, 2008)

xii)Masking: The masking of tones happens in both the frequency and the time domain. Time domain masking (forward and backward masking) happens when a tone is masked by a previous sound or by a sound that occurs shortly after the tone. Forward masking is more powerful than backward masking, meaning the time duration, t, between the masking and masked tone can be larger for forward masking than backward as shown in Figure 2.13.

> "The time duration for forward masking can be as large as 20-30 ms, while t for backward masking can not be much larger than 10 ms As mentioned earlier, masking in the frequency domain also occurs. Pure tones that are close together in pitch mask each other better than those widely separated.Pure tones also mask higher frequencies better than lower frequencies. The higher frequencies will mask more easily than the lower frequencies.Finally, a tone of greater intensity masks a broader ranger of tones than a tone of less intensity." (Bengert and Upward, 2003)

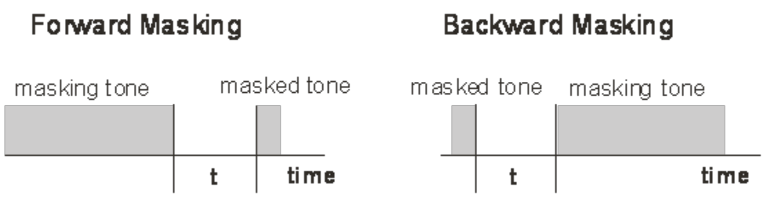

Figure 2.13 Time Domain Masking
(Bengert and Upward, 2003)

xiii)Binaural Masking: Binaural masking is simply the internal masking done by the human auditory system due to the fact that we have two ears, one on either side of the head. The processing and mapping of sound in the brain is still relatively unknown.

Binaural Masking Effects are shown in Figure 2.14:

“a is 440 Hz reference tone of one second. b plays the same reference tone added to some noise in the left ear, and nothing in the right ear. c is the sine wave plus noise in the left ear and noise in the right ear. d is the sine plus noise in the left ear and the same (in-phase) sine plus noise in the right ear.” (Bengert and Upward,2003)

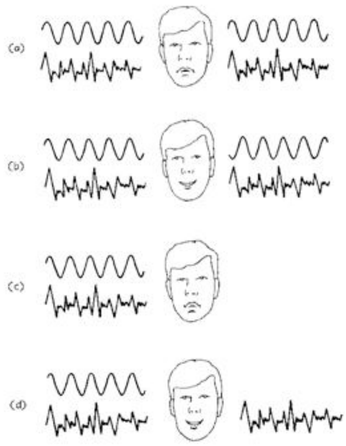

Figure 2.14 Binaural Masking Effects
(Bengert and Upward, 2003)

xiv)Grouping and Segregation:

The human auditory system groups and separates sounds in a very similar fashion to the human visual system. Due to this similarity, visual aids can help explain how the brain groups and segregates sounds. Apparent motion is a phenomenon that occurs both visually and acoustically. Animation is an example of visual apparent motion.Animation is just a series of drawings viewed in rapid succession. The brain groups these images together and interprets them as a single moving image.

“Apparent motion is possible in the auditory domain as well. The brain will group notes as a single melody if the notes are alternated in time at a slow rate.(A) The human auditory system thinks the sounds belong together, and groups them in a single group of four notes. As the time delay between notes is decreased the brain begins to group the sounds differently.(B) We no longer hear a melody, we only hear the rhythmic beats. Our auditory system is now hearing four groups of one note each. (C).”(Bengert and Upward, 2003)

Figure 2.15 illustrates these visually:

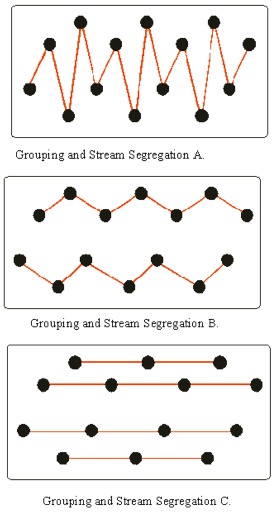

Figure 2.15 Grouping and Segregation
(Bengert and Upward, 2003)

2.3.3 Auditory Scene Analysis

In a natural acoustic environment, for example sitting in a bar talking with some friends, the sound waves reaching each ear can include components from friends' voices, music playing in the bar, and passing traffic. The auditory system must separate the incoming complex signal into each of these components to build separate perceptual representations of the sources generating each sound. Bregman (1990) called the perceptual process that decomposes complex acoustic signals "auditory scene analysis".

> “The neurophysiology of the ear suggests that sensory information arriving at the basilar membrane can be described in terms of a spectrogram.That is, the initial information available to the auditory system is that visible in a spectrogram.” (Bregman, 1990)

The corresponding spectrogram of this simple sound (no background noise, masking or interfering sounds) looks like in Figure 2.16:

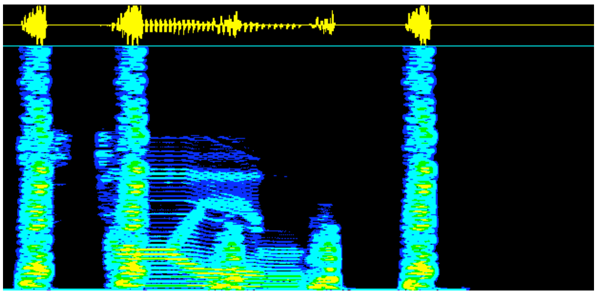

Figure 2.16 Simple Sound Spectrogram
(Bregman, 1990)

i)Fourier theory: Fourier analysis is a mathematical procedure that allows any complex sound to be broken down into its sine wave components. Fourier analysis produces a magnitude spectrum, or Fourier spectrum, containing information about the power in the original signal at each frequency and a phase spectrum, containing information about the phases of the sine waves that make up the complex signal. Together, the magnitude and phase spectrum provide a complete representation of the original signal, which can be recombined using fourier synthesis.

> "Fourier analysis assumes that the sound signal is unchanged over time; however, real signals do not remain unchanged. For example, human speech contains many frequency components that vary in amplitude over time. A simple magnitude spectrum cannot show these variations so, instead, the frequency content of the signal is analysed over a series of small time intervals and displayed graphically in a spectrogram. The spectrogram clearly shows how the frequency content of the sound signal changes over time. After fourier synthesis, the spectrum of a recombined signal can be compared to the original signal to investigate how a transmitting device or medium has modified the sound signal. For example, when a sound passes over the human head it acts as a frequency filter, as it allows low-frequency components to pass but removes the higher frequency components from the signal. The transfer function of a filter describes the amount of attenuation applied by the filter at each sine wave frequency in the signal; for example, values

> close to 1 represent little attenuation, whereas values close to 0 indicate that little of the sound is transmitted through the filter. Filtering techniques are useful for describing the properties of filters (e.g. microphone) and predicting their effects on sound signals." (Mather, 2008)

ii)Digital Audio Watermarking: Digital watermarking is a technology which allows a secret message to be hidden in a computer file, without the detection of the user. The watermark is not apparent to the user, and does not affect in any way, the use of the original file. Watermark information is predominantly used to identify the creator of a digital file, a picture, a song, or text.

iii)Watermark Insertion: The process of inserting a digital watermark into an audio file can be divided into four main processes as seen in Figure 2.17.

> "An original audio file in wave format is fed into the system, where it is subsequently framed, analyzed, and processed, to attach the inaudible watermark to the output signal." (Bengert and Upward, 2003)

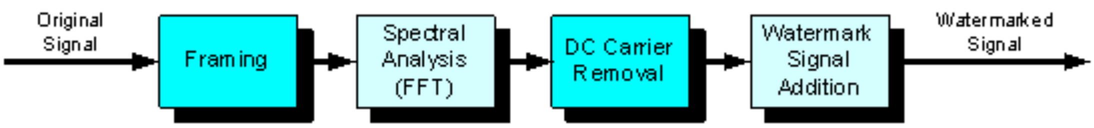

Figure 2.17 Watermark Insertion Process
(Bengert and Upward, 2003)

iv)Watermark Extraction: The process of extracting the digital watermark from the audio file is similar to the technique for inserting the watermark. The computer processing requirements for extraction are slightly lower. A marked audio file in wave format is fed into the system, where it is subsequently framed, analysed, and processed, to remove the embedded data which exists as a digital watermark.

Figure 2.18 shows the extraction process:

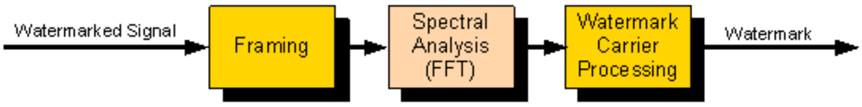

Figure 2.18 Watermark Extraction Process
(Bengert and Upward, 2003)

v)Phase Encoding:

> "This watermarking technique exploits the human auditory systems lack of sensitivity to absolute phase changes by encoding the watermark data in an artificial phase signal. Phase encoding works by breaking the audio signal into frames, and performing spectral analysis on each frame. Once the spectrum has been computed, the magnitude and phase of consecutive frames are compared, and an artificial phase signal is created to transmit data. The artificial phase is modulated in with the phase from each frame, and the new phase frames are combined to form the watermarked signal." (Bengert and Upward, 2003)

The modified phase frames can also be smoothed to limit the amount of distortion present in the marked signal (Figure 2.19), but in minimizing distortion, the data rate of the watermark is constrained respectively.

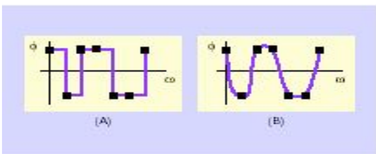

Figure 2.19 Phase Smoothing
(Bengert and Upward, 2003)

> "The phase encoding watermarking technique offers higher data rates over the previous methods, averaging from 8 to 32 bits per second. This technique is increasingly effective in the presence of noise." (Bengert and Upward, 2003)

vi)Spread Spectrum Watermarking: This watermarking technique relies on direct sequence spread spectrum (DSSS) to spread the watermarked signal over the entire audible frequency spectrum such that it approximates white noise, at a power level as to be inaudible.

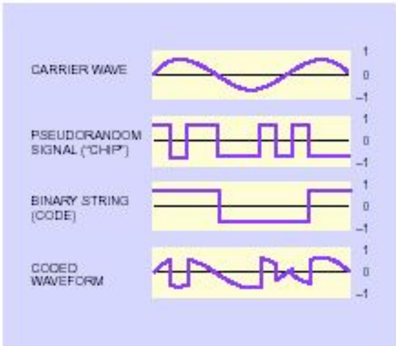

Figure 2.20 Spread Spectrum
(Bengert and Upward, 2003)

vii)Echo Watermarking:

> **"**The echo data hiding technique relies on distorting an audio signal in a way which is perceptually dismissed by the human auditory system as environmental distortion. The original audio signal is copied into two segments (kernels), one which leads the original signal in time, and one which lags. Each kernel represents either a zero or a one bit for watermark data transmission. The bit stream of watermark data is used to mix the two kernels together. The signals are mixed with gradually sloped transitions to reduce distortion." (Bengert and Upward, 2003)

2.3.4 Perception of Sound

i)Loudness Perception: Loudness is the perceptual attribute of a sound that corresponds most closely to its intensity.

> "Two techniques have been used in experimental studies to measure the loudness of sounds: loudness matching and loudness scaling.In loudness matching, the participant matches the intensity of a sound (comparison stimulus) until it sounds as

loud as a standard stimulus that has a fixed intensity. This technique can be used to investigate the frequency dependence of loudness by manipulating the frequency difference between a comparison and a standard stimulus. If this procedure is repeated for a range of comparison frequencies, an equal-loudness contour is produced, which plots comparison SPL as a function of frequency." (Mather, 2008)

ii)Frequency Response Curves: The ear is not equally sensitive to all frequencies, particularly in the low and high frequency ranges. The response to frequencies over the entire audio range has been charted, originally by Fletcher and Munson in 1933, with later revisions by other authors, as a set of curves showing the sound pressure levels of pure tones that are perceived as being equally loud.

"The curves are plotted for each 10 dB rise in level with the reference tone being at 1 kHz. Also called loudness level contours and the Fletcher-Munson curves. The curves are lowest in the range from 1 to 5 kHz, with a dip at 4 kHz, indicating that the ear is most sensitive to frequencies in this range. The intensity level of higher or lower tones must be raised substantially in order to create the same impression of loudness. The phon scale was devised to express this subjective impression of loudness, since the decibel scale alone refers to actual sound pressure or sound intensity levels. The lowest curve repesents the threshold of hearing, the highest the threshold of pain. Curves based on the studies of Fletcher and Munson showing the response of the human hearing mechanism as a function of frequency and loudness levels.When the sound is loud (top curve, threshold of pain) all frequencies can be heard equally well." (Truax, 1999)

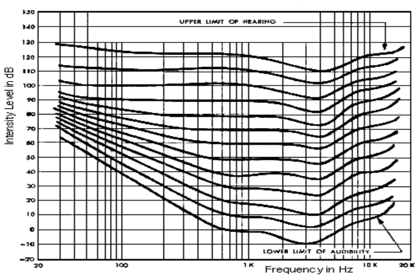

Figure 2.21 Frequency Response Curves

(Traux, 1999)

iii)Pitch Perception:

“Pitch is the perceptual attribute of a sound that corresponds to its frequency. Pitch is related to tone frequency in pure tones and to the fundamental in complex tones. Pitch perception allows us to order sounds on a musical scale. Helmholtz believed that the auditory system constructs a separate representation for each frequency component of a complex sound (Ohm's law).” (Mather, 2008)

iv)Frequency Selectivity:

“Psychophysical masking experiments measure a subject's ability to detect a pure tone signal in the presence of noise. Band-pass noise contains only sound frequencies that fall within a certain band above and below its centre frequency. Masking is effective only when the noise is close in frequency to the signal, indicating that hearing is served by overlapping band-pass filters covering a range of frequencies. These filters are remarkably similar to the tuning curves of individual auditory nerve fibres identified physiologically. In psychophysical experiments, masking of a signal that occurs only when the centre frequency of the noise falls within a certain band of frequencies is called critical-band masking.” (Mather, 2008)

v)Frequency Discrimination: Studies of frequency discrimination measure a listener's ability to detect small changes in frequency.

“Listeners hear two tones with slightly different frequencies and report whether the first or the second tone had a higher pitch. Frequency discrimination is remarkably good at low frequencies, but deteriorates at high frequencies.” (Mather, 2008)

vi)Theories of Pitch Perception:

> "A sound stimulus produces a characteristic pattern of activity across neurons in the peripheral auditory system tuned to different frequencies. The peripheral auditory system provides three ways to encode frequency: The frequency-to-place conversion of the basilar membrane. The timing of responses in auditory nerve fibres; according to Goldstein's (1973) pattern theory, the auditory system resolves the sinusoidal components of a complex sound, and encodes its pitch as the fundamental frequency that best fits the harmonic series of components." (Mather, 2008)

vii)Speech Perception: The vibration of air being forced through the vocal cords by the diaphragm produces speech sounds.

> "The vibration creates a harmonic sound whose fundamental frequency depends on the rate of vibration. When the harmonic sound reaches the vocal tract, the vocal tract acts as a resonator. The vocal tract resonates more at certain frequencies, called formant frequencies, which produce vowel sounds. Formants often have a smooth change in their frequencies during the first 50 ms of a speech sound; for example, /ba/ /da/ and /ga/ differ in terms of their formant transitions. It has been argued that speech perception involves a specific mode of auditory processing, called speech mode, involving:
>
> 1. A process that is qualitatively different from the processes dealing with non-speech sounds, because it supports categorical perception.
>
> 2. Specialised neural structures, including Wernicke's area in the cortex." (Mather, 2008)

2.4. Image and Visual Perception

2.4.1 Eye and Its Anatomy

The structure of the mammalian eye can be divided into three main layers or tunics whose names reflect their basic functions: the fibrous tunic, the vascular tunic, and the nervous tunic. The nervous tunic, also known as the tunica nervosa oculi, is the inner sensory which includes the retina. The Retina is composed of two layers:

a)Outer nuclear layer: Rods and cones are the two types of photoreceptor in the outer nuclear layer. The outer component of rods and cones consists of a stack of light-sensitive disks, and the inner segment contains the cell nucleus and synaptic terminals.

.

b)Inner nuclear layer: Bipolar cells transmit graded potentials "vertically" from photoreceptors to ganglion cells. ON bipolars are activated by an increase in the photon catch of receptors, and they depolarise.

> “The human eye is a roughly spherical, light-tight chamber, the inside of which is lined with a sheet of photoreceptors. After entering the cornea, light passes through the pupil and the lens before striking the retina (photoreceptor sheet). The interior of the eye is filled with vitreous humour, which maintains the shape of the eye and holds the retina in the correct position, and aqueous humour, which nourishes the lens and keeps the eye inflated.” (Mather, 2006)

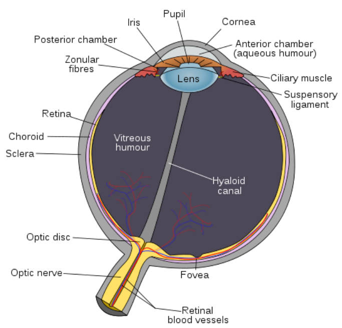

Figure 2.22 Schematic Diagram of the Human Eye

(Rhcastilhos, 2007)

2.4.2 Properties of Light

Light normally travels in straight lines or rays from a light source. Light rays are emitted from a point on a source in all directions. Opaque occluding objects in the path of the rays create shadows. Light rays are refracted (change direction) as they pass from one transmitting medium, e.g. air, into another, e.g. water.

> "Refraction occurs when a light ray changes mediums. Light traveling from air and going into water would be an example. The speed of the light ray changes upon changing mediums. In almost every case the direction of the light ray changes also." (Zobel, 2006)

Figure 2.23 demonstrates the general behavior of a light ray as it travels from air into some transparent medium such as water or glass.

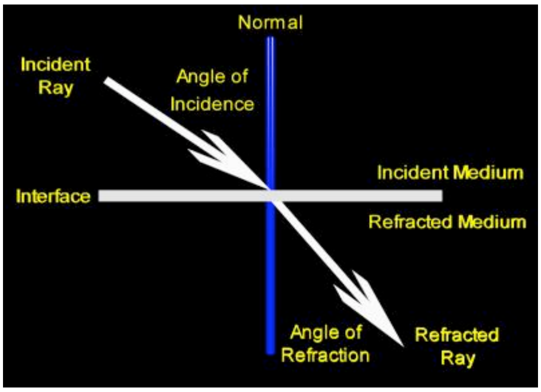

Figure 2.23 Light Refraction
(Zobel , 2006)

The descriptions for the terms in this diagram are represented by Zobel (2006) as:

- The ray of light which travels through the incident, or first, medium and strikes the boundary, or interface, is called the incident ray.
- The ray of light which travels into the refracted, or second, medium and leaves the interface is called the reflected ray.

- A line perpendicular to the surface is imagined at the point of refraction.This line is called a normal. In this context the word normal means perpendicular. In the above diagram the normal is colored blue.
- The angle between the incident ray and the normal is called the angle of incidence, or the incident angle.
- The angle between the refracted ray and the normal is called the angle of refraction, or the refracted angle.

"When light strikes the interface between two substance, e.g. between air and glass, it may be absorbed, reflected or transmitted. During absorption, light quanta are taken up by the substance and converted into thermal energy. Absorption is important for vision because, in order for light to be seen, photoreceptors must absorb light energy and convert it into electrical signals. Reflected light provides information about the properties of surfaces in the environment, because reflections behave in a very lawful manner; light rays are reflected so that the angle of incidence equals the angle of reflection. As light waves are transmitted through dense media, interference causes the waves to slow down. If the light rays enter the medium obliquely, their direction of travel changes (refraction). The refractive index of a medium refers to the degree to which it slows down light rays, and consequently alters their direction." (Mather, 2008)

2.4.3 The Dimensions and Specifications of Colour

The light that is emitted or reflected from a surface into the eye is not coloured. The experience of colour is a neural construct derived from the wavelength composition of the incoming light.

i)Traditional Colour Descriptions:

Colour has three attributes:

1. "Hue: the experience of blue, yellow, red, etc., that the colour creates. Hue is closely related to the wavelength of the light, but a simple one-to-one correspondence does not exist.

2. Saturation: a measure of the colour's purity in terms of how much neutral colour (white) it contains. A highly saturated red contains little white, and becomes pink when desaturated.

3. Brightness: the visibility of the colour, most closely related to the perceived intensity of the light that creates it." (Mather, 2006)

"The hue, saturation, and brightness of a colour have traditionally been depicted in a three-dimensional perceptual colour space, with hue varying around the circumference of a horizontal circle, saturation corresponding to the distance from the centre of the circle, and brightness being represented on the vertical axis. The arrangement of hues on the circle's circumference has been influenced by introspective observations made by Hering.

1. Hering considered red, green, blue, and yellow to be elementary colour sensations; unique in that he could describe all other experiences of colour using these four, and they themselves could not be described by any other colours.

2. He proposed that red and green were opponent colours in that the sensation of red and the sensation of green never appeared to co-exist in the same colour. Likewise, blue and yellow were mutually exclusive." (Mather, 2006)

ii)Contemporary Colour Space:

> "The universally agreed chromaticity diagram is based on extensive additive matching data, and supersedes subjective descriptions of colour space. The wavelength of pure spectral colours is plotted on the perimeter of the space; saturation decreases towards a central white region .There is no need to represent brightness, since S, M, and L cones retain their response ratio regardless of the absolute brightness of a colour. A straight line drawn between any two perimeter wavelengths will pass through the range of colours created by an additive mixture of the wavelengths at various relative intensities. Complementary wavelengths have mixture lines that pass through the central region, and produce white when mixed at appropriate relative intensities. Most colours can be created by mixing many different pairs of primary wavelengths, with the exception of the saturated purples, whose only primaries are the shortest and longest of the visible spectrum (440 and 700 nm)." (Mather, 2008)

A color model is an abstract mathematical model describing the way colors can be represented as tuples of numbers, typically as three or four values or color components (e.g. RGB and CMYK are color models). However, a color model with no associated mapping function to an absolute color space is a more or less arbitrary color system with no connection to any globally-understood system of colour interpretation.

> "The sensitivy of conic sensors located in human retina shows different varieties for different wave ranges.Some of them may sense blue/green or some of them may sense orange the best.After this process they send signals to the brain.It's not to be forgotten that %8-10 percent of men and less than %1 percent of women can't sense colours in the correct way.This becomes a matter especially for the blue/green parts of the colour spectrum.The most of the colour-blind patients have this problem by genetic heritage and some of them by another disease or accident.This problem may effect when using a colour coded screen terminal." (Aydın, 1995)

iii)Colour Theories and Dual-Process Theory: Hurvich and Jameson resolved the apparent discrepancy between theories of trichromacy and the opponent pairing of colours, by proposing that an opponent stage of processing followed trichromatic analysis at the photoreceptor level.

> "Electrophysiological evidence has now confirmed dual-process theory. Ganglion and LGN cells show red-green, blue-yellow, and light-dark opponent responses, based on signals from the triad of retinal photoreceptors (S, M, and L cones). The information is organised as follows..." (Mather, 2008)
>
> - The difference between M and L cone responses is carried by midget ganglion cells in the red-green opponent channel.
> - The difference between the responses of the S cones, and the sum of the L and M cones, is carried by bistratified ganglion cells in the blue-yellow opponent channel.
> - By summing the responses from M and L cones, wavelength information is discarded by the parasol and midget ganglion cells of the light-dark channel.

Table 2.2 Color Channel (Mather, 2008)

Opponent channel	**Cone input**	**Signal carried by**
Red-green chromatic channel	Opponent: L-M	Midget ganglion cells
Blue-yellow chromatic channel	Opponent: S-(L + M)	Bistratified ganglion cells
Light-dark achromatic channel	Non-opponent: L + M	Parasol and midget cells

iv)Colour Interactions:

- **Simultaneous Colour Contrast:**

> **"**The cone excitation ratio produced by a source will only predict its colour in spatial contexts that are chromatically neutral.Simultaneous colour contrast is one form of colour interaction.In chromatic contexts, the colour of a test patch is influenced by its

surround. Usually the hue of the test is shifted in a direction that is complementary to the surround." (Mather, 2008)

- **Colour Adaptation:**

"Colour adaptation is another form of colour interaction. In chromatic contexts, the colour of a test patch will be influenced by an adapting colour that precedes it. If the test patch is preceded by an adapting colour with a similar hue, the test will appear less saturated than a physically identical test shown in a neutral temporal context. Neutral test colours will assume the hue that is complementary to the adapting colour. Assuming the adapting colour reduces the sensitivity of the cone classes it excites, subsequently presented test colours will produce cone excitation ratios biased in favour of previously less active cones." (Mather, 2008)

- **Colour Constancy:**

The Figure 2.24 shows a bowl of fruit photographed in three lighting conditions artificial light (left), hazy daylight (middle), and clear blue sky (right).

"Notice the marked variation in colour balance caused by the spectral properties of the illuminant. We are not normally aware of this variation because colour constancy mechanisms discount illumination effects." (Mather, 2008)

Figure 2.24 Colour Constancy
(Mather, 2008)

- **Colour Deficiency:**

Table 2.3 Colour Deficiency (Mather, 2008)

Colour deficiency	Type	Number of cone classes	Abnormality	Consequences	Neutral point	Prevalence (%)
Anomalous trichromacy	Protanomaly	3	Peak of L cones closer to M than normal	Never require more than 3 primaries to match any colour experience (but in different proportions to normal observers). More sensitive to greens than normal observers. Colour discrimination poorer than normal observers.	None	1.73
	Deuteranomaly	3	Peak of M cones closer to L than normal	Never require more than 3 primaries to match any colour experience (but in different proportions to normal observers). More sensitive to reds than normal observers. Colour discrimination poorer than normal observers.	None	4.78
Dichromacy	Protanopia	2	L cones absent	Never require more than 2 primaries to match any colour experience. No distinction between reds and greens (< 520 nm).	492 nm	0.81
	Deuteranopia	2	M cones absent	Never require more than 2 primaries to match any colour experience. No distinction between reds and greens (< 520 nm).	498 nm	0.48
	Tritanopia	2	S cones absent	Never require more than 2 primaries to match any colour experience. No distinction between blues and yellows (450 nm-480 nm).	575 nm	0.45
Monochromacy	Rod monochromacy	0	All cones absent	Poor visual acuity. Never require more than 1 primary of correct intensity to match any colour	All wavelengths neutral	0.3

Colour deficiency	Type	Number of cone classes	Abnormality	Consequences	Neutral point	Prevalence (%)
				stimulus. No distinction between any colours.		
	Cone monochromacy	1	Some receptors that resemble cones	Near normal visual acuity. Never require more than 1 primary of correct intensity to match any colour stimulus. No distinction between any colours.	All wavelengths neutral	0.0001

- **Colour Vision:**

"The long-held belief that human colour vision is trichromatic was based on the additive combination of coloured lights, and has since been confirmed by the identification of three cone classes, with peak sensitivities in the blue, green, and red regions of the spectrum. Colour is also organised opponently, on two principal chromatic axes (red-green and blue-yellow), in a way that seemingly contradicts trichromacy.Current physiological evidence demonstrates that dual-process theory, in which retinal trichromacy is followed by opponent analysis at the LGN, accurately represents human colour vision. Many colour phenomena can be explained by cone excitation ratios. In neutral chromatic contexts, a mixture's colour can be predicted by the cone excitation ratio of its constituents, and metameric colours appear identical, despite different spectral compositions, because their cone excitation ratios are identical.In chromatic contexts, contrast effects highlight the need to compare a colour's cone excitation ratio to that of its surround. Sensitivity changes in individual cone classes underlie adaptation. Both effects help to achieve colour constancy in widely different conditions of illumination. Since the congenital colour deficiencies of monochromats, dichromats, and anomalous trichromats result from cone photoreceptor abnormalities, their colour experiences can be predicted by atypical cone ratios." (Mather, 2008)

- **Accommodation:**

"The purpose of the optics of the mammalian eye is to bring a clear image of the visual world onto the retina. Because of limited depth of field of the mammalian eye, while an object at one distance from the eye might project a clear image an object either closer to or further from the eye will not. To make images clear for objects at different distances from the eye, its optical power needs to be changed. This is accomplished mainly by changing the curvature of the lens. For distant objects, the lens needs to be made flatter, for near objects the lens needs to be made thicker and more rounded." (Mather, 2008)

- **Vergence movement:**

"When a creature with binocular vision looks at an object, the eyes must rotate around a vertical axis so that the projection of the image is in the centre of the retina in both eyes. To look at an object closer by, the eyes rotate 'towards each other' (convergence), while for an object farther away they rotate away from each other' (divergence).Exaggerated convergence is called cross eyed viewing (focusing on the nose for example).When looking into the distance, or when 'staring into nothingness', the eyes neither converge nor diverge.Vergence movements are closely connected to accommodation of the eye. Under normal conditions, changing the focus of the eyes to look at an object at a different distance will automatically cause vergence and accommodation." (Mather, 2008)

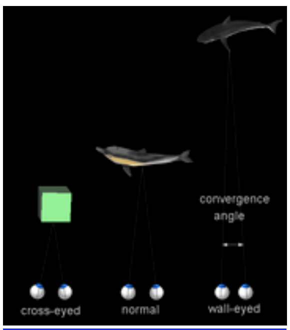

Figure 2.25 The Two Eyes Converge
(Hsu, 2005)

- **Visual Acuity:**

“Visual acuity is often measured in cycles per degree (CPD), which measures an angular resolution, or how much an eye can differentiate one object from another in terms of visual angles.Resolution in CPD can be measured by bar charts of different numbers of white–black stripe cycles. For example, if each pattern is 1.75 cm wide and is placed at 1 m distance from the eye, it will subtend an angle of 1 degree, so the number of white–black bar pairs on the pattern will be a measure of the cycles per degree of that pattern. The highest such number that the eye can resolve as stripes, or distinguish from a gray block, is then the measurement of visual acuity of the eye. For a human eye with excellent acuity, the maximum theoretical resolution would be 50 CPD (1.2 minute of arc per line pair, or a 0.35 mm line pair, at 1 m). However, the eye can only resolve a contrast of 5%. Taking this into account, the eye can resolve a maximum resolution of 37 CPD, or 1.6 minute of arc per line pair (0.47 mm line pair, at 1 m).” (Mather, 2008)

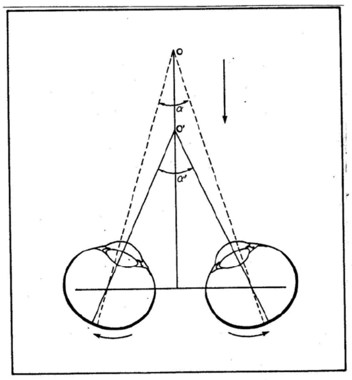

Figure 2.26 Visual Acuity
(Onaygil, 2005)

- **Equivalent Resolution:** The experience of wide sharp human vision is in fact based on turning the eyes towards the current point of interest in the field of view, the brain thus perceiving an observation of a wide sharp field of view.The narrow beam of sharp vision is easy to test by putting a fingertip on a newspaper and trying to read the text while staring at the fingertip,it is very difficult to read text that's just a few centimeters away from the fingertip.

- **Eye Movement:**

  "The visual system in the brain is too slow to process that information if the images are slipping across the retina at more than a few degrees per second. Thus, for humans to be able to see while moving, the brain must compensate for the motion of the head by turning the eyes. Another complication for vision in frontal-eyed animals is the development of a small area of the retina with a very high visual acuity. This area is called the fovea, and covers about 2 degrees of visual angle in people. To get a clear view of the world, the brain must turn the eyes so that the image of the object of regard falls on the fovea. Eye movements are thus very important for visual perception, and any failure to make them correctly can lead to serious visual disabilities. Having two eyes is an added complication, because the brain must point both of them accurately enough that the object of regard falls on corresponding points of the two retinas; otherwise, double vision would occur. The movements of different body parts are controlled by striated muscles acting around joints. The movements of the eye are no exception, but they have special advantages not shared by skeletal muscles and joints, and so are considerably different." (Mather, 2008)

- **Spatial Contrast Sensitivity:** In Figure 2.27, spatial frequency increases from left to right, while contrast increases from top to bottom.

  "The shape of the visible, lower part of the image gives an indication of our relative sensitivity to different spatial frequencies; medium frequencies are visible at the lowest contrast (higher up the image). Spatial sensitivity for stationary gratings is band-pass. This image is known as a Pelli-Robson Chart." (Mather, 2008)

Figure 2.27 Spatial Contrast Sensitivity
(Mather, 2008)

Natural images contain a diverse range of:

- Coarse-scale information: low spatial frequency sine wave gratings representing the general shape of objects.
- Fine-scale information: high spatial frequency sine wave gratings representing surface texture and sharp edges.

"Low-pass filters blur images by removing component gratings with high spatial frequencies; high-pass filters remove component gratings with low spatial frequencies. Neurons in the visual system act as band-pass spatial filters, carrying information about different narrow ranges of spatial frequencies in an image." (Mather, 2008)

- **Stereoscopic Depth and Motion Processing:**

"A difficult task in the analysis of stereoscopic depth involves correctly matching the features that have been extracted from the image in one eye with the corresponding features extracted in the other eye. Spatial frequency-tuned filters could simplify this stereo-correspondence problem by restricting the search for matches to certain bands of spatial frequency. Motion processing uses filters tuned to relatively low spatial frequencies. Their large receptive fields sample greater segments of the motion sequence, and improve estimates of velocity and direction." (Mather, 2008)

2.4.4 Perception of Vision

i)Shape and Object Perception:

Gestalt laws describe the tendency for elements of images to be grouped together on the basis of properties such as proximity, similarity, common fate, and good continuation. Shape segmentation and contour integration can be achieved through image-based computations and through symbolic computations. There may also be a role for parsing processes that link together disparate shapes that belong to the same, partially occluded object surface.

"Object representations are important for several reasons: recognition, discrimination, and interaction. Object processing is complex because images of objects reflect both intrinsic and extrinsic factors. View-independent theories suggest that the visual system removes all extrinsic factors and creates a structural description of objects in terms of their component parts. According to view-dependent theories, objects are stored as a few discrete prototypical views. Recognition occurs by comparing a novel 2-D view with intermediate views between prototypes. Currently it is not clear which theory best accounts for human object perception, and indeed there may be two or more systems working in parallel." (Mather, 2008)

ii)Depth Perception:

"To successfully interact with a 3-D world, the visual system must reconstruct the dimension of depth from numerous visual cues in the 2-D retinal image, and from non-visual cues regarding the state of the intra- and extra-ocular muscles. Multiple monocular cues, derived from a single eye, vary as a function of distance and can be used to estimate depth. Most offer highly correlated, quantitative distance estimates, relative to fixation (relative depth) or to the observer (absolute depth). The binocular horizontal disparities that mediate stereopsis require information from both eyes, and are considered the most important cue to depth. The random-dot stereogram demonstrates their power as an independent cue, extracted by early disparity-selective cells capable of fine distance discriminations relative to fixation. Research on cue combination has been prompted by the multiplicity of depth cues available. The suggestion that cues are combined according to a context-dependent weighted average is consistent with current empirical research, and with the knowledge that all cues have limitations constraining their utility." (Mather, 2008)

iv)Visual Motion Perception:

"Physiological evidence tells us that comparator-type circuitry exists in the rabbit retina and primate cortex to encode motion. In humans there is compelling psychophysical evidence, from the motion after-effect and direction-specific threshold elevation, that analogous neural motion detectors exist. For a meaningful percept of complex motion, the visual system must combine the many signals it receives from the individually ambiguous, localised motion detectors.

Separate theories of integration have been proposed for moving planar surfaces, 3-D objects, and articulated biological forms. The original two-process motion theory (short-range/long-range) accounted for empirical differences between random dot kinematograms (involving elementary motion detectors) and classical motion displays (thought to involve higher-level processing). The validity of this distinction has since been criticised, and the theory reappraised. The recent, widely accepted multiple-process theory acknowledges the role of higher-level processes, and sub-divides neural motion detectors into those that respond to luminance-defined motion (first-order detectors), and those that respond to texture-defined motion (second-order detectors)." (Mather, 2008)

3. AUDIO VISUAL SYNCHRONIZATION

3.1. Audio to Video Synchronization

Synchronization or synchronisation is timekeeping which requires the coordination of events to operate a system in unison. The familiar conductor of an orchestra serves to keep the orchestra in time. Systems operating with all their parts in synchrony are said to be synchronous or in sync.

> "The history of synchronization goes back to the 17th century when the Dutch scientist Christiaan Huygens reported on his observation of synchronization of two pendulum clocks which he had invented. Another important observation of synchrony of organ pipes was described by Lord Reyleigh in his Theory of Sound. Being, probably, the oldest scientifically studied nonlinear effect, synchronization was understood only in 1920's when E. V. Appleton and B. Van der Pol systematically, theoretically and experimentally studied synchronization of triode electronic generators." (Rosenblum&Pikovsky, 2003)
>
> "In our physical world we are used to the fact that light travels much faster than sound. An extreme example is lighting and thunder. We are used to the fact that we first see the lightning and a few moments later we hear the thunder. When a basketball hits the court in a large sports venue, for the people on the first rows would see the ball making contact with the floor and almost simultaneously hear the sound of the ball hitting the floor. The further back you get, the more the sound lags, but it still seems natural. If this timing was reversed (sound comes first), you would hear the sound of the ball hitting the court before you saw making the ball contact with the floor. This is very unnatural and would seem incorrect even if you were in the first few rows where there was just a small delay between audio and video.
>
> The conclusion is that human perception is much more forgiving for sound lagging behind sight as this what we are used to seeing in everyday occurrences." (NXP, 2007)

There are different ways where the AV-sync can get mis-synchronized:

> "1- Internal AV-sync error: Different processing delays between image and sound in video camera and microphone. The AV-sync interdelay is normally fixed.
>
> 2- External AV-sync error: If a microphone is placed remotely of the sound source, the image is out of sync, because the speed of sound is much lower than the speed of light. If the sound source is 340 meters from the microphone, then the sound arrives approximately 1 second later than the light.The AV-sync interdelay varies with distance.
>
> 3- During mixing of video-clips and sound-clips you normally need to delay either sound or video so they are synchronized. The AV-sync interdelay is static, but can vary with the individual clip
>
> 4- Video editing effects." (Aggarwal&Jindal, 2008)

Examples of transmission (broadcasting), reception and play-back, that can get the AV-sync mis-synchronized:

> "1- A video camera with built-in microphones or line-in ought to delay sound and video paths by the same amount of milliseconds. A video camera should have some sort of explicit AV-sync timing put into the video and audio streams. Solid state video cameras can delay the video signal by one or more frames.
>
> 2- If an audio video stream during transmission gets bit-errors because of electrical glitches (wired) or wireless interruptions - and the AV-stream either miss explicit AV-sync timing or the processing is not respecting the explicit AV-sync timing. The AV-sync interdelay normally increases with time.

3- There is extensive use of audio and video signal processing circuitry with significant delays in television systems. Particular video signal processing circuitry which is widely used and contributes significant video delays include frame synchronizers, digital video effects processors, video noise reduction, format converters and pre-preprocessing.

4- The video monitor processing circuit delay the video stream. Pixellated displays require video format conversion and deinterlace processing which can add one or more frames of video delay.

5- A video monitor with built-in speakers or line-out ought to delay sound and video paths by the same amount of milliseconds. Some video monitors contain internal user adjustable audio delays to aid in correction of errors." (Aggarwal&Jindal, 2008)

Synchronization is an important concept in the following fields: Computer science, Telecommunication, Physics, Cryptography, Multimedia, Photography, Music (rhythm) Stage arts, Live visuals

"Whenever audio and video streams are stored and processed, there is necessarily a separation between them. When the audio and video are presented for viewing, care must be taken to keep the audio and video in exact synchronization (AV-sync)." (ATSC, 2003)

3.2 Effect of No Explicit AV-Sync Timing

When a digital or analog audio video stream do not have some sort of explicit AV-sync timing these effects will break AV-sync as follows:

- "In film movies these timing errors are most commonly caused by worn films skipping over the movie projector sprockets because the film has torn sprocket holes.
- Errors can also be caused by the projectionist misthreading the film in the projector, although this is rare with competent projectionists.
- AV-sync is commonly corrected and maintained with an audio synchronizer. Television industry standards organizations have established acceptable amounts of audio and video timing errors and suggested practices related to maintaining acceptable timing.
- AV-sync errors are becoming a significant problem in the digital television industry because of the use of large amounts of video signal processing in television production, television broadcasting and pixelated television displays such as LCD, DLP and plasma displays.
- In the television field, audio video sync problems are commonly caused when significant amounts of video processing is performed on the video part of the television program.
- Typical sources of significant video delays in the television field include video synchronizers and video compression encoders and decoders. Particularly troublesome encoders and decoders are used in MPEG compression systems utilized for broadcasting digital television and storing television programs on consumer and professional recording and playback devices.
- A source of significant video delay is found in pixelated television displays (LCD, Plasma display, DLP) which utilize complex video signal processing to convert the resolution of the incoming video signal to the native resolution of the pixelated display.
- In broadcast television, it is not unusual for lip flap or lip-sync error to vary by over 100 ms (several video frames) from time to time." (ATSC, 2003)

3.3 The Development of Avant-Garde Sound-on-Film Techniques

Avant-garde sound-on-film techniques developed around 1930 and usedwell into the 1950s. These techniques constitute one of the least-known yet most striking parallels to the working methods and ideas of electro-acoustic music. But, before discussing these techniques and their relationship to electro-acoustic music, some historical and technological background is necessary.

> "Almost from its inception in 1895, the silent film was provided with ac- companying music and sound effects by local theater musicians. These live accompaniments, though occasionally first rate, were plagued by inappropriate music, poor performance quality, and the difficulty of synchronization with the image on the screen. Eventually, these problems helped to precipitate the development of a synchronizable studio recording of both music and sound effects that could be distributed with the film. The initial breakthrough in this research, a synchronizable phonograph mechanism, received world ac- claim with the 1927 premikre of Warner Brothers' The Jazz Singer. The tre- mendous though short-lived popularity of the synchronizable record overshad-owed the far more portentous advent, in 1929, of the ability to record sound and image side by side on film. This sound-on-film method proved far more preferable and is still in use today.
>
> Sound is recorded on film as black-and-white patterns. It is made audible again through a process depicted in the highly simplified abstraction. Inside a movie projector, the sound track is passed between a light source and a photoelectric cell. Both are individually and completely enclosed except for one slit in each container; these slits face each other through the sound track. As a result, the only light reaching the photoelectric cell must pass through a thin width of film.As the film moves through the projector, the black-and-white sound patterns recorded on it cause the amount of light striking the photoelectric cell to fluctuate rapidly. The photoelectric cell generates electricity when illuminated, while the amount of the resulting current varies with the intensity of the light. Thus the sounds recorded on film are converted to electrical impulses which can be made audible by means of an amplifier and loudspeaker. The sound-on-film patterns can take several basic forms,

the left-hand pair are known as variable area sound tracks while the third represents the variable density type, labels that indicate the means of varying the light transmission. Film music innovators soon discovered a number of unexpected and intriguing secondary benefits to sound-on-film technology. For instance, it allowed one to freeze and visually preserve a sound for study. These sound pictures could also be manipulated and altered, generating entirely new sounds and sound arrangements. Parts of a sound might be excised while individual sounds or entire pieces of music could be reordered, reversed, and superimposed by cutting, rearranging, and splicing the film." (James, 1986)

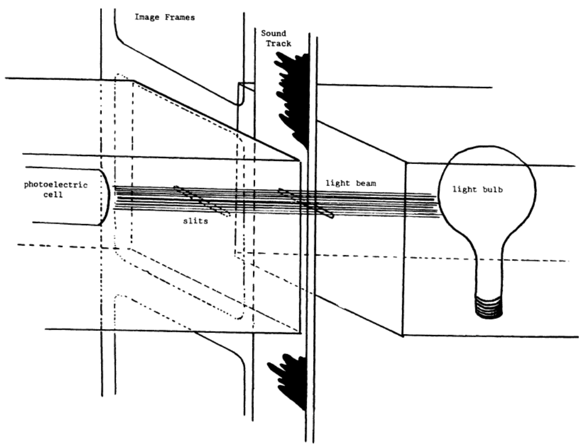

Figure 3.1 Diagram of the Sound Track Reader
(James, 1986)

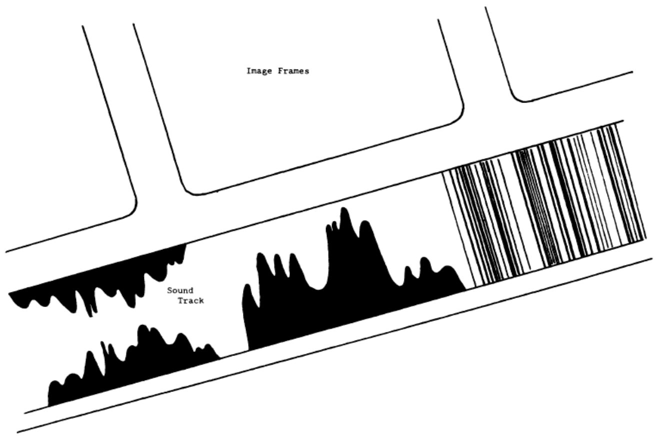

Figure 3.2 Sound Track Patterns
(James, 1986)

Sounds could also be altered by adding marks to the sound track by hand. Finally, the recording process itself could be circumvented completely, either by painting sounds directly onto the sound-track film or by taking pictures of various designs with an ordinary animation camera and then reducing and combining them like an animation sequence.

> "Sound animation, the second type of experimental work with optical sound recording, is quite different from the montage techniques described thus far. Sound montage involves the manipulation of sounds already recorded on film. Animated or drawn sound. on the other hand, constitutes the actual synthesis of sound by application of visual animation techniques to the sound track. Like its visual counterpart, animated sound encompasses two basic methods: sound drawn with paint or ink directly onto blankfilm, and the more common approach of using a standard animation camera to photograph pictures of sound which are then combined in a series to produce a soundtrack.
>
> An interest in animated sound surfaced among abstract film and animation pioneers in several European countries around 1930. The first thorough investigation of its potential began in Russia at the Scientific Experimental Film Institute in Leningrad. There, music theorist and mathematician Arseni M. Avraamov took an empirical approach, experimenting with sound tracks created by photographing drawings of repeated geometric shapes. He found that any repeated pattern would produce a distinctive sound. Pitch was controlled by the frequency with which the shape appeared on the film and volume by the length of exposure, for example the grayer the print, the quieter the sound. It was possible to achieve limited polyphony by means of double exposure, very rapid alternation between tones, or dividing the sound track into multiple strips.
>
> On a more subtle level, sound-on-film techniques offered composers freedom from the limitations of the human performer, musical notation, and conventional musical instruments, allowing them to work directly with sound and to control every minute detail of the music and its realization. The range of sounds included in the category of musical resources was broadened, new range was brought to almost every

musical parameter, and fundamentally new ways of working with individual sounds were explored. All of these possibilities would become central, attractive features of the electro-acoustic music field and are nowhere more fully prefigured than in the avant-garde film sources, experiments, and conjectures of the 1930s. Finally, the range of radically new resources made available by sound on-film techniques necessitated an experimental attitude toward composition similar to that seen in the early electro-acoustic music studios." (James, 1986)

3.4 Synchronization Implementation and Standards

The Advanced Television Systems Committee, Inc.(ATSC), is an international, non-profit organization developing voluntary standards for digital television. The ATSC member organizations represent the broadcast, broadcast equipment, motion picture, consumer electronics, computer, cable, satellite, and semiconductor industries.

The end-to-end DTV audio-video production, distribution, and broadcast system is a complex array of digital processing, compression, decompression, and storage devices. Each component in the system imposes a latency on the audio and/or video signals flowing through it. System design goals often call for the relative audio-video latency through each component to be in the sub-millisecond range. Operationally, unequal delays can be imposed on the audio and video signals respectively, and these delays compromise audio-video synchronization.

"One of the overarching goals of the DTV broadcasting system is to deliver audio and video in proper synchronization to the viewer. Because each digital audio and video component in the chain from production to reception imposes some degree of latency on the signals passing through it, and the delays imposed on the audio and video signals are typically unequal, each component harbors the potential to cause an audio-video synchronization error at its output. The overall audio-video synchronization error is the algebraic sum of the individual synchronization errors encountered in the chain. While a given synchronization error may cause either a positive or negative differential shift in audiovideo timing, the video signal is typically subjected to greater delay than the audio signal, and the tendency is therefore toward video lagging behind audio. Thus, there is a requirement to monitor audio-video synchronization at various points within in the system and to make corrections, where required, in order to deliver to the viewer audio-video synchronization within the required tolerance." (ATSC, 2003)

In addition, there are points within the end-to-end chain that require AV-sync to be maintained, such as switching points, monitoring points, and transmission/encoding points.

"For the purposes of this discussion, the end-to-end DTV system may be divided into four segments: acquisition and production/post production (contribution system), release facility and distribution system, local broadcast station, and home receiver.

At the production-post production stage, AV-sync errors can occur in the capture stage, in film-to-video transfer, and in editing. The product may be delivered on video tape or by various electronic means, but whatever the delivery medium, that steps must be taken to ensure that audio-video synchronization in the delivered product falls within the required tolerance .The release facility segment contains a number of devices through which the DTV audio and video signals are passed, which variously impose compression and de-compression, processing, and storage and their attendant differential delays on the signals. The process of distributing the signals to affiliate stations typically requires compression and decompression steps. It is incumbent on the release facility to correct the differential audio-video delays that the signals experience within the plant so that the initial timing relationship is restored to a tight tolerance before the signals reach the distribution encoder. The synchronization to a tight tolerance should be maintained in any encode/decode process that is involved in delivering the signals to the affiliate station, so that the tight AV-sync can be monitored at switching and other points.

The affiliate station segment contains a number of devices that are similar to those encountered in the release facility segment and that generate the same types of differential audio-video delays, including switching and monitoring points.AV-sync should be restored to a tight tolerance before the signals are input to the broadcast station's ATSC audio and video encoding devices to assure that the presentation time stamps placed on the audio and video access units by the encoder faithfully represent correct synchronization.

Under all operational situations, at the inputs to the DTV encoding devices, the sound program should be tightly synchronized to the video program. The sound program should never lead the video program by more than 15 milliseconds, and should never lag the video program by more than 45 milliseconds.

MPEG-2 models the end-to-end delay from an encoder's signal input to a decoder's signal output as constant. This end-to-end delay is the sum of the delays from encoding, encoder buffering, multiplexing, transmission, demultiplexing, decoder buffering, decoding, and presentation. Presentation time stamps are required in the MPEG bit stream at intervals not exceeding 700 milliseconds. The MPEG System Target Decoder model allows a maximum decoder buffer delay of one second. Audio and video presentation units that represent sound and pictures that are to be presented simultaneously may be separated in time within the transport stream by as much as one second. In order to produce synchronized output, the receiver must recover the encoder's System Time Clock (STC) and use the Presentation Time Stamps (PTS) to present the audio-video content to the viewer with a tolerance of +/-15 milliseconds of the time indicated by PTS.

Although real aural and visual presentation devices typically have finite and different inherent delays, and may have additional delays imposed by post-processing or output functions, the System Target Decoder models these delays as zero. Such delays must be corrected before the audio and video signals are presented to the viewer." (ATSC, 2003)

3.5 Defining the Problem of AV-sync and Implementing Solutions

The enemy of precise network time synchronization is non-determinism. Latency estimates are confounded by random events that lead to asymmetric round-trip message delivery delays; this contributes directly to synchronization error. To better understand the source of these errors, it is useful to decompose the source of a message's latency. Kopetz and Schwabl (1989) characterize its distinct components as:

- Send Time
- Access Time
- Receive Time

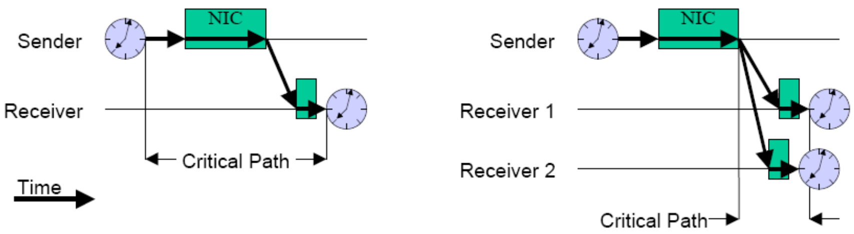

Figure 3.3 Critical Path Analysis
(Elson, 2003)

> "The fundamental property of Reference-Broadcast Synchronization is that a broadcast message is only used to synchronize a set of receivers with one another, in contrast with traditional protocols that synchronize the sender of a message with its receiver. Doing so removes the Send Time and Access Time from the critical path. This is a signicant advantage for synchronization on a LAN, where the Send Time and Access time are typically the biggest contributors to the nondeterminism in the latency. Mills attributes most of the phase error seen when synchronizing an NTP client workstation to a GPS receiver on the same LAN ((500 sec)2000 sec in his 1994 study) to these factors; Ethernet jitter and collisions." (Elson, 2003)

With the introduction of advanced digital delivery systems for audio and video, there is an increased awareness of the timing relationship between audio and video. Owing to advanced data compression technologies such as Dolby Digital (AC-3) for audio and MPEG-2 for video, sound is clearer and pictures are sharper. Technologies such as Digital Television (DTV), DVD, Direct Broadcast Satellite (DBS), and Digital Cable use these compression techniques to deliver extremely high-quality programming to consumers. However, it is a misalignment of these same systems that is the root cause of most AV-sync problems.

3.5.1 AV-Sync Measures

The human body can tolarate the fact that light travels faster than sound, but when it is reversed the react to the conclusion is commonly sensed as an error. The measure timing between the sight and sound should be calculated and fixed before broadcasting so the potential lagging is minimized for the viewer.

> "Most film editors are able to detect AV-sync errors as short as +/- ½ film frame. As film is projected at 24fps (frame per second) in the US and 25fps in Europe, this equates to approximately +/- 20msec. It is claimed that some editors can detect even smaller errors, but this might be more accurately attributed to their familiarity with the material being viewed. Other figures abound, such as +/- 1 video frame (+/-33-40ms). Dolby Laboratories has specified that any Dolby Digital decoder must be with the range of +5ms audio leading video to –15ms audio lagging video. This is because human perception of AV-sync is weighted more in one direction than the other.
>
> The International Telecommunications Union (ITU) released ITU-R BT.1359-1 in 1998. It was based on research that showed the reliable detection of AV-Sync errors fell between 45ms audio leading video and 125ms audio lagging behind video. That was just for detection, while the acceptability region, and therefore the recommended maximum was quite a bit wider. In summary, the recommendation states that the tolerance from the point of capture to the viewer and or listener shall

> be no more than 90ms audio leading video to 185ms audio lagging behind video. This range is probably too wide for truly acceptable performance, and tighter tolerances are generally obeyed." (Linear Acoustic, 2004)

3.5.2 Delays in the Television Broadcasting

AV-sync issues within the TV plant are not new to digital television, they are perhaps more noticeable. Although the obvious may be re-stated, there is enough of a problem that still exists in NTSC that it is worth the trouble to identify and fix them prior to beginning. Unlike some basic points to keep in mind are that in general, audio operations are very low latency.

> "Generally, no compensating video delay needs to be added as the latency is so low. Video processing, on the other hand, takes substantial amounts of time, usually no less than one video frame. Similar to audio, any time a video signal is digitized, operations upon that signal will take longer. As most video effects are unable to be performed in the analog domain, delay is inevitable." (Linear Acoustic, 2004)

3.5.3 Video Frame Synchronizing

Due to its very nature, a video frame synchronizer causes between one and two variable frames of delay. In this case, a special audio delay that is able to track the variable delay of the frame synchronizer is required.

> "Most video frame synchronizers are available with matching and tracking audio delays and should always be purchased as a set as there is currently no standard interface that represents AV-sync values. Digital video effects can add from one to many video frames. As the delay of a DVE is generally a fixed value, a fixed value audio delay can also be used. Devices with fixed delays are easier to deal with if they are always kept in line, or if they must be removed then a fixed video delay equal to that of the device is inserted in its place. This will prevent having to

dynamically adjust audio delay and create an audible disturbance." (Linear Acoustic, 2004)

3.5.4 AV-Sync in MPEG-2 System

"The MPEG system provides the proper tools to make AV-sync absolutely correct. Each audio and video frame has a Presentation Time Stamp (PTS) that allows the decoder to reconstruct the sound and pictures in sync. These PTS values are assigned by the Multiplexer in the MPEG encoder. The decoder receives the audio and video data ahead of the PTS values and can therefore use these values to properly present audio and video in sync. It is imperative that audio and video are applied to the Dolby Digital (AC-3) and MPEG-2 encoders in sync. A very common mistake is to calibrate the multiplexer to compensate for plant differences. Although this may work fine in the short term, it should be avoided in permanent installations. Larger problems will likely result, including issues with some consumer decoders." (Linear Acoustic, 2004)

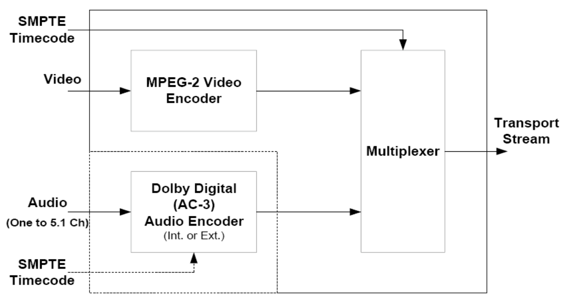

Figure 3.4 Simplified MPEG-2 Encoding System
(Linear Acoustic, 2004)

i)Aligning the MPEG-2 Encoding System

Video and audio encoding take some time to accomplish, and the multiplexer must know exactly how long. This delay depends on the manufacturer of the equipment, but the value is crucial for getting the PTS values correctly assigned.

"Many of the AV-Sync problems encountered thus far in the field can be attributed to these delays not being properly accounted for or just not set at all. In practical terms, this simply means that if your transmission system uses an external Dolby Digital (AC-3) encoder, there is a known, fixed audio encoding latency that must be entered into the MPEG-2 encoding system. There is usually a setting called MPEG-2 Encoder Audio Delay, or possibly AC-3 Delay.Once set, it need not be changed unless either the audio or video encoder latency is reset." (Linear Acoustic, 2004)

ii)Testing the MPEG-2 Encoding System & Decoder

"In its simplest terms, testing entails feeding typical AV-sync test material such as beep/flash (audio pip with simultaneous video flash) to the encoder, capturing the resulting transport stream from the multiplexer, and using analysis software to determine compliance." (Linear Acoustic, 2004)

The Figure 3.5 below was captured from Sync Check and clearly shows the timing relationship between audio and video.

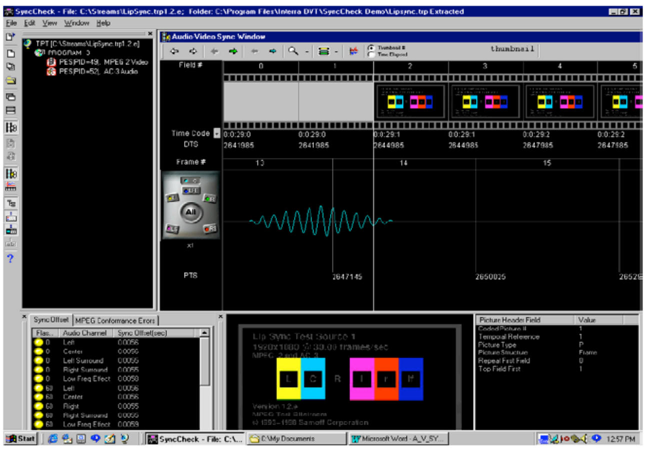

Figure 3.5 Testing the MPEG-2 Encoding System
(Linear Acoustic, 2004)

"The latency of the Dolby Digital (AC-3) encoding algorithm is fixed regardless of the number of encoded audio channels. Therefore, a two-channel signal adequate to test the AV-sync relationship of an MPEG-2 encoded video signal and a Dolby Digital (AC-3) encoded audio signal. This means that a test tape can be easily created with a video flash and an audio beep, verified with an oscilloscope, and used as the source for testing with Sync Check. Testing the MPEG-2 decoder is a very straight forward process. It requires that a reference transport stream be applied to the decoder under test. This transport stream is again a beep/flash type signal that has been encoded and verified for proper synchronization as described above. The audio and video outputs are then displayed on a dual trace oscilloscope and compared." (Linear Acoustic, 2004)

Figure 3.6 shows what a typical measurement might look like on an oscilloscope screen:

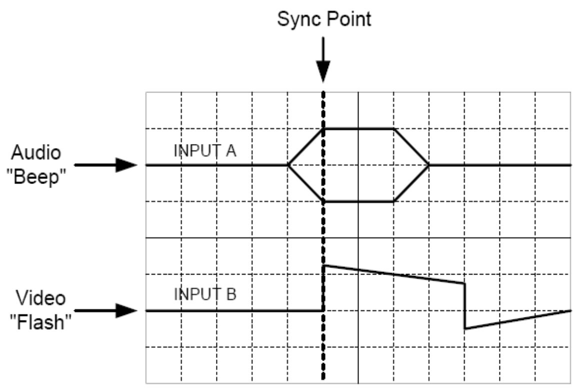

Figure 3.6 Outputs of Oscilloscope
(Linear Acoustic, 2004)

3.6 Second Generation AV-Sync

The most obvious result of audio to video mismatch is visible "lip sync" errors. This problem certainly can and does happen in today's systems, with the frequency of occurrence becoming a significant concern to advertisers and station management. The mistiming of audio and video will always cause a subconscious degradation of the program's entertainment quality as perceived by the home viewer when the audio is advanced with respect to the video the cause of this effect is believed to be the unnatural sound relationship which the television program presents.

“Viewing a television program with advanced audio is unnatural for the viewer, and is believed to cause subconscious stress. Psychological tests at Stanford University demonstrate that viewers who watch television commercials having advanced audio "evaluate people on television more negatively (e.g. less interesting, more unpleasant, less influential, more agitated, less successful)" than the same commercials which were played with the audio in sync with the video. It was also discovered that this effect takes place with relatively small audio advances where the mere existence of an audio problem was detected by very few average viewers. In addition to the negative perception of the commercials in the presence of advanced audio, there was also evidence this caused the test subject's memory of the negative aspects of the commercial to be remembered longer than normal. The worst possible scenario takes place, the viewer perceives the advanced audio commercial in a bad light, and also remembers it longer than a commercial which is properly presented. Obviously, such problems can cause a great deal of concern for television advertisers.” (Reeves&Voelker, 1993)

3.6.1 New Audio Synchronization System

“A method and system for providing synchronization of the timing of various multimedia events, including an audio event is disclosed. Clock objects are defined in the storage and associated with an internal or external source of current time. The clock objects are able to be displayed on the display, but can be hidden once their linkages are defined. One or more multimedia objects representative of audio, visual or other multimedia events, including an audio object, are defined and linked to a particular clock object or clock objects. Then, a processor synchronizes the multimedia objects, including a MIDI data object, with the associated clock object or objects. Finally, the various multimedia events are performed in synchronization with their associated clocks.

Multimedia is perhaps the fastest growing application for computer systems. Increasingly, users are employing computers to present graphic, sound and imaging information to end users.Users are increasingly demanding ergonomic interfaces for managing multimedia presentations. In the past, the system clock was often used to commence a sound playback at a certain time, or present information on a computer display at a specific time. However, tools for synchronizing the presentation of music or sound with the display of information as a multimedia presentation unfolded was not possible.

Examples of current multimedia systems that do not have the synchronization capability of the subject invention are Apple's Quicktime and Microsoft's Video for Windows.In particular, audio data must be properly synchronized with other multimedia events to create an aesthetic presentation." (Patent Storm, 1997)

3.6.2 CCD Camera Generated Vision Delays

Audio to video synchronization errors are becoming more troublesome as television technology progresses.Some charge coupled recording devices generate delay due their systems.

"The wide use of cameras having CCD sensors is aggravating this synchronization problem. All CCD sensors have an inherent visual delay mechanism. Depending on the sensor type, the visual delay may be several fields for newer camera types .In particular, the liberal use of digital frame store based image processing in newer cameras is creating previously unknown vision delays of several fields, with a four field delay not being uncommon.

"The exposure time is equivalent to aperture time. The ratio of exposure time to frame rate is the aperture ratio. It is known from sampling theory that the aperture ratio effect on frequency response, which in this case is the ability to accurately convey motion. For short exposures, the ability to convey motion to the viewer increases dramatically. The shorter exposure time gives brighter and less blurred moving edges which result in the viewer's improved ability to perceive motion. The CCD camera induced improved motion perception aggravates the corresponding increased image delay time, and makes any audio to image mismatch easier for the viewer to consciously or subconsciously detect." (Reeves&Voelker, 1993)

3.6.3 Video Processing Delays

> "Video signals are often passed through a special effects generators, color correctors, noise reducers, frame synchronizers and a variety of other editing and image processing functions. As memory costs continue to decline, these devices increase in complexity, and many incorporate frame based processing functions which add delays which are switched in and out.
>
> Unlike the past where video delays slowly drifted due to differing sync generator phases, the video delay in many of today's systems take instant jumps of one or more frames, as editors and other operators select different processing modes. This situation is especially true of many current noise reduction and color correction products where extra frames of delay are added for each additional selected function." (Reeves&Voelker, 1993)

Clearly, television facilities need to be designed with audio synchronization in mind. It is impractical to remove the offending video delays, so the only remaining solution is to ensure that the program audio receives the same delay as the associated video.

> "Part of the solution is to measure the video delay at each significant delaying device so that a corresponding audio delay can be inserted at that point. Several video synchronizer manufacturers have a digital delay output (DDO) which provide a current video delay value signal for use by a companion audio synchronizer. Additionally, video delay detectors are available for devices which do not provide DDO signals. The audio synchronizer receives the DDO signal and automatically delays the audio signal by a corresponding amount. Delay detectors for video devices without DDOs operate by storing a given input video frame and comparing all output frames to the stored frame. By counting the number of frames which pass until the previously input frame is output, the video delay is obtained. These devices are easy to add to an existing system, requiring only that input and output video be looped through their inputs." (Reeves&Voelker, 1993)

4. RESEARHCES AND EXPERIMENTS ABOUT AUDIO VISUAL SYNCHRONIZATION ON PERCEPTION

The following experiments in this topic are applied on various people groups. The purpose of these experiments are to show the importance of sound on supporting visuals and how it increases the effectiveness or vice versa and the main goal is to prove the strong effect of proper use of audio visual synchronization on perception and memory for finalizing the thesis subject.

4.1. Experiment 1: Effect of Distorted AV-Sync Media on Perception

i)Research Questions, Observers and Apparatus:

This is basically a test about perception and memory of audio-visual synchrony in which the both media are connected to each other in harmony but no other significant meaning as individuals.

1. Does only audio with no video is more effective on perception or vice versa?
2. Does audio-video asynchrony is effective on memory?
3. How does proper audio-video synchrony effect on memory and behavior?
4. How does only audio with music effect the perception and how does audio with music and proper synchronized video effect the perception?
5. How does a visual/video that is created based on a music synchrony, effect the perception without the audio?

10 naive observers participated in the experiments as 2 groups. First group of 5 were in the age range 20-30 and the second group of 5 were through 30-50. Mixed genders and all with no hearing or no vision disabilities.

Apparatus within the properly isolated reference room in a music studio, the visual stimuli appeared on a LCD display(1680X1050 pixels:widescreen) positioned at eye level. The sound stimuli were generated through a M-Audio Audiophile professional sound card and a pair of M-Audio BX5 studio reference monitor speakers, positioned to either side of the display.

ii)Experimental Method:

Observers were asked to watch to music video; Gantz Graf by Autechre, director and creator: Alex Rutterford. This is a perfect example of audio-visual synchrony, the visual is spessificaly created for the audio timeline and graphics were designed in the song's dark and technologic mood. The musical genre is IDM (Intelligent Dance Music) and it is pretty much about noises and timing effects for the first timers. The audio and video complete each other so well that is why it is chosen for this experiment about perception and memory.

Figure 4.1 Gantz Graf Music Video by Autechre – Director: Alex Rutterford (Warp Records, 2002)

iii)Procedure:

The video is shown to observes in various stages:

- Sound, No Video
- Video, No Sound
- Sound and Video in Asynchrony
- Sound and Video Synchronized

iv)Results:

Just audio without the video has showed interesting results on memory and perception since the video is generated out from sound effects and noises 8 out of 10 didn't find the audio effective or memorable. Just video showed a similar effect too because the motion graphics in the video had a total connection to the audio so 7 out of 10 didn't find the video gripping.

Observers were asked what they remember from the sound or video after 30 mins break time. 4 out of 10 could remember some details but with mistakes and 6 could not remember anything at all.

After the break asynchornized version of the video was shown with sound. Similar results like the first two stages were occurred; not attractive and not memorable. The complete video with sound and synchrony showed a great example of what the synchrony is all about in visual arts. 10 out of 10, this time found the Gantz Graf video effective and the audio meaningful.

All 10 observers could remember what the vision and sound was about after watching it in sync. This experiment proved the great union of audio and video on perception and memory if there was a connection with a meaning between them.

4.2. Experiment 2: Effect of Harmonic AV-Sync Media on Perception

i)Research Questions, Observers and Apparatus:

The same experiment was tested again with a proper meaningful music that follows the video's timeline from time to time but has a no total connection as the first experiment. Both the media has a meaning separate and a harmony in together.

1. Does only audio with no video is more effective on perception or vice versa?
2. Does audio-video asynchrony is effective on memory?

3. How does proper audio-video synchrony effect on memory and behavior?
4. How does only audio with music effect the perception and how does audio with music and proper synchronized video effect the perception?
5. How does a visual/video that is created based on a music synchrony, effect the perception without the audio?

10 naive observers participated in the experiments as 2 groups. First group of 5 were in the age range 20-30 and the second group of 5 were through 30-50. Mixed genders and all with no hearing or no vision disabilities.

Apparatus within the properly isolated reference room in a music studio, the visual stimuli appeared on a LCD display(1680X1050 pixels:widescreen) positioned at eye level. The sound stimuli were generated through a M-Audio Audiophile professional sound card and a pair of M-Audio BX5 studio reference monitor speakers, positioned to either side of the display.

ii)Experimental Method:

Observers were asked to watch to music video; 1.618 by Brian Transeau, director and creator: Scott Pagano. This is an example of audio-visual synchrony once again the visual is created for the audio timeline and graphics were designed in the song's mood but the music contains orchestral tones that gradually transition into buzzing insect sounds and acoustic guitar. The purpose of the second version of this experiment is to see if an audio and video with both have a meaning with together and alone is effecting the memory and perception or not.1.618 is the value of the Golden ratio rounded to the nearest thousandth. The video shows off examples of the golden ratio in nature and mechanisms.

Figure 4.2 1.618 Music Video by Brian Transeau – Director: Scott Pagano (Binary Acoustics, 2006)

iii)Procedure:

The video is shown to observes in various stages:

- Sound, No Video
- Video, No Sound
- Sound and Video in Asynchrony
- Sound and Video Synchronized

iv)Results:

Just sound without the video has showed a different result on memory and perception than the first experiment since the music has a meaning and a composition independent from the video, 7 out of 10 found the music memorable. Only video without the music had a similar effect too but this time 5 out of 10 found the video interesting and thought both of the elements are quite effective without each other but need each other.

Observers were asked what they remember from the sound or video after 30 mins break time. 7 out of 10 could remember what was in the video and 5 out 10 could remember what the music was like. So this shows if audio and video is not connected to each other in someways they are memorable since they have a meaning on their own. The strange thing is video was more effective on percetion, so we could say visual perception is stronger than the hearing related perception.

After the break asynchornized version of the video was shown with sound once more. The observers who found the audio and video memorable alone, this time found the asynchronized combination not that interesting. Since they are connected in someways the beats couldn't support the visuals on time so it made this version unpleasant.

The synchronized version of the audio and video showed once again the perfect result. 9 out of 10 had a memorable time watching it and could remember what was going on after. The perception of sound and video has the strongest potential when the audio and video are in proper synchrony and connected in someways and also can be experienced alone as media.

4.3. Experiment 3: Effect of Informative AV-Sync Media on Perception

i)Research Questions, Observers and Apparatus:

The same method of experiment was tested for the last this time with a 30 minutes long football game videoclip with an information giving commentary audio on it.

1. Does only audio with no video is more effective on perception or vice versa?
2. Does audio-video asynchrony is effective on memory?
3. How does proper audio-video synchrony effect on memory and behavior?

10 naive observers participated in the experiments as 2 groups. First group of 5 were in the age range 20-30 and the second group of 5 were through 30-50. Mixed genders and all with no hearing or no vision disabilities.

Apparatus within the properly isolated reference room in a music studio, the visual stimuli appeared on a LCD display(1680X1050 pixels:widescreen) positioned at eye level. The sound stimuli were generated through a M-Audio Audiophile professional sound card and a pair of M-Audio BX5 studio reference monitor speakers, positioned to either side of the display.

ii)Experimental Method:

Observers were asked to watch to a 30 minutes long football clip with an audio commentery. Both sides score a goal at a random time. This is a example of audio-visual synchrony test with an information giving audio with the video. The purpose is to detect the effect of sound and video together and alone, in synchrony and asynchrony on memory.

iii)Procedure:

The video is shown to observes in various stages:

- Sound, No Video
- Video, No Sound
- Sound and Video in Asynchrony
- Sound and Video Synchronized

iv)Results:

Only video with no sound was more effective than only the audio. 7 of 10 could remember the scoring times with smalltime mistakes on sides. Sound showed various results depending on the observer. Some could remember the scoring sides order correct, some could remember the scoring times correct but generally the results were not quite sharp as the first just the video version.

Asynchornized version of the video with sound created stress and made the video not watchable. Since the information giving commentery was linked to the game second by second, 5 of 10 observers asked to skip this step and commented this version could not effect the memory in any good means.

The synchronized version of the audio and video together has shown the perfect result.
9 out of 10 could remember what the game was about, what the scoring times were and what the correct sides scored in order.

4.4. General Results About the Experiments

As a result of the three experiments it is clear that audio supported video with a proper synchrony has the strongest effect on memory and the perception as well, than all the other versions that was tested.

Asynchrony of the audio supported video lead to some problems like not remembering the details and the timings. Probably the worst results were observed in this situation. It is clear that asynchrony of the audio-visual media is effecting the perception in negative ways. It's better to have only one media type in an informative giving, educational or art related source than having them both in an asynchronized state.

Only video with no audio has a larger potential effect on perception than audio only but this may show different variences since effect of sound perception is stronger in some cases. The conclusion is, visual perception is stronger and more memorable than the audio perception as speaking of media formats and audio adds speed and develops better understanding to the video.

The results enlightened the fact that even a sound that has no specific meaning itself; like a noise, can gain an interesting potential to effect human interest and perception when it is in proper synchronization with the visual changes in the video. The human perception reacts to this harmony in a positive way and the combination of the media becomes more memorable and enjoyable than the solo experience.

The informative audio-visual media should be synchronized with no errors in all cases since this is the most important detail that should be considered if the source is the main information and the visuals are supportive elements. Because the asynchrony of this type of media has a subversive effect on the information itself.

The results of the three experiments are summarized below:

Table 4.1 Summary of Results

	Experiment 1	Experiment 2	Experiment 3
Only Sound	Not effective	Effective	Depends
Only Video	Not effective	Depends	Effective
AV-Synchronized	Effective	Effective	Effective
AV-Asychronized	Not effective	Not effective	Not effective

5. CONCLUSION

In this research we tried to give a wide perspective to explain audio-visual synchronization. By explaining the most concepts, in relation with communication, and basic principles of audio-visual features of the topic is investigated.

The aim of this thesis is to show the importance of audio-visual synchronization usage in visual communication as well as the other related media mediums by revising the general topics like hearing and seeing, explaning how human perception works, defining the usage of synchronization and supporting the value of this subject by the experiments given.

Synchronization, which means transferring multiple multimedia data varieties at the same time, states that sound and visual are suitable to each other. By considering human perception, technical preparations which will be made in this area are needed to be based on these principals to increase the effect of communication.

The audio and video are generally mixed with proper softwares in digital mediums, the correction of lags and clippings due the recording is the first process that should be done by the editor than the fixed media is embedded to the same timeline and workspace manually. The restored and synchronized media is filtered and re-checked for the last time then sent through distribution amplifiers.

Because the multimedia data is in high dimensions such as video, sound and animation especially in internet and mobile channels, upgrading efforts for these areas will decrease synchronization problems. For example, the usage of compressed digital audio for the first was a revolutionary progress in the online audio-visual field by providing a more fluent broadcast.

The researches and the given experiments in this thesis show the audio and video has a larger scale of effect both in perception and impression on the audience when used together in perfect synchronization, especially the videos/animations which are mainly created for sync based audio-visual timelines such as experimental music, interactive computer games and all related nu-media.

Videos that follow the audial timeline on track have a lot more potential than keeping the audience interested. The asynchrony of the audio-visual combination is one of the worst things that might effect the quality of the media and human perception when it occurs. The informative component as an audio totally loses the aim and control on the viewer when it is in such asynchronized condition with the video. As a result the most important negative indicator in synchronization topic should be assessed as the occurring of delay.

For videos which is broadcasted mainly in all media, while visual is moving, sounds should be suitable to visual, should complete the visual, also there should not be any delay which creates disharmony between them. These kinds of problems affect negatively the watched program's rational and perceived quality. This will cause the broadcast or related media to lose audience and interest.

REFERENCES

Aggarwal, S. and Jindal, A., 2008, "Comprehensive Overview of Various Lip Synchronization Techniques", Biometrics and Security Technologies, International Symposium on Volume, Issue, pp 23-24

Allegro, S., Büchler, M. and Launer, S., 2008, "Automatic Sound Classification Inspired by Auditory Scene Analysis". Available on site: http://www.ee.columbia.edu/ln/crac/papers/allegro.pdf

Allen, G., 1998, "Communicating". Available on site: http://telecollege.dcccd.edu/mgmt1374/book_contents/3organizing/commun/communic.htm

Anatomy of Ear and Eye, Available on site: 163.20.152.20/~nature/mwf12.ppt

ATSC(Advanced Television Systems Committee.), 2003, "ATSC Implementation Subcommittee Finding: Relative Timing of Sound and Vision for Broadcast Operations". Available on site: http://www.atsc.org/standards/is_191.pdf

Aydın, E.D., Büro Otomasyonunda Ergonomi, Beta Basım Yayım, 2.Baskı, İstanbul, 1995.

Bengert, J. and Upward, A., 2003, "Perceptual Audio Demonstrations". Available on site: http://www.ece.uvic.ca/~ece499/2003a/group09/p/demos.htm

Clark, D., 2005, "Communication & Leadership". Available on site: http://www.nwlink.com/~donclark/leader/leadcom.html

Elson , J.E., 2003, "Time Synchronization in Wireless Sensor Networks". Available on site: http://lecs.cs.ucla.edu/~jelson/dissertation-final.pdf

Foulger, D., 2004, "Models of the Communication Process". Available on site: http://foulger.info/davis/research/unifiedModelOfCommunication.htm

Hsu, F., 2005, "Stereogram Tutorial Eye Convergence". Available on site: http://www.answers.com/topic/eye

James, R.S., 2008, "Avant-Garde Sound-on-Film Techniques and Their Relationship to Electro-Acoustic Music". Available on site: http://links.jstor.org/sici?sici=0027-4631%281986%2972%3A1%3C74%3AASTATR%3E2.0.CO%3B2-A

Linear Acoustic Inc., 2004, "Audio and Video Synchronization: Defining the Problem and Implementing Solutions". Available on site: http://www.linearacoustic.com/pdf/A_V_SYNC_R2.pdf

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Auditory Localisation". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch05-tp-03

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Classification ". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch01-tp-01

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Colour Deficiency". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch12-tp-05

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Colour Interactions". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch12-tp-04

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Colour Space". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch12-tp-02

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Colour Theories". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch12-tp-03

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Colour Vision". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch12-tp-01

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Complex Sounds". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch04-tp-01

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Foundations of Perception". Available on site: http://www.psypress.com/mather/resources/chapter.asp?chapter=10

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Foundations of Perception". Available on site: http://www.psypress.com/mather/resources/chapter.asp?chapter=10

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Fundamental Functions". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch08-tp-01

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Individual Differences in Perception". Available on site: http://www.psypress.com/mather/resources/chapter.asp?chapter=13

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Loudness Perception". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch05-tp-02

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Object Representation". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch09-tp-02

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Perceptual Principles". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch01-tp-02

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Pitch Perception". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch05-tp-02

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Representation at Multiple Spatial Scales". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch08-tp-02

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Shape and Object Perception". Available on site: http://www.psypress.com/mather/resources/chapter.asp?chapter=09

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Shape Representation". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch09-tp-01

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Simple Sounds". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch04-tp-01

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Some Important Properties of Light". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch06-tp-02

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Speech Perception". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch05-tp-04

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "The Eye". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch06-tp-03

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "The Retina". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch07-tp-01

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "The Senses". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch01-tp-01

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Theoretical Principles". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch01-tp-02

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Uses of Spatial Filters". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch08-tp-03

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "Visual Motion Perception". Available on site: http://www.psypress.com/mather/resources/chapter.asp?chapter=11

Mather, G., 2008, Foundations of Sensation and Perception, Second Edition, Chapter "What is Light?". Available on site: http://www.psypress.com/mather/resources/topic.asp?topic=ch06-tp-01

NXP, 2007, "Audio Video Synchronization Application Note", Available on site: http://www.tcshelp.com/pnx1517_app/MPTK_AVsync.pdf

Onaygil, S., 2005, "Görme Olayı". Available on site: http://atlas.cc.itu.edu.tr/~onaygil/disayd_Gorme_Olayi.pdf

Patent Storm Issued., 1997, "Audio Synchronization System". Available on site: http://www.patentstorm.us/patents/5655144-fulltext.html

Reeves, B. and Voelker D., 1993, "Effects of Audio-Video Asynchrony on Viewer's Memory, Evaluation of Content and Detection Ability", Available on site: http://www.pixelinstruments.tv/pdf/Articles/Effects%20of%20Audio-Video%20Asynchrony.PDF

Rhcastilhos, 2007, "Shematic Diagram of the Human Eye". Available on site: http://www.nationmaster.com/encyclopedia/Image:Schematic-diagram-of-the-human-eye-en.svg

Rosenblum, M. And Pikovsky, A., 2003, "Contemporary Physics", Taylor & Francis Ltd, volume 44, number 5, pp 401 – 416

Rosenblum, M. And Pikovsky, A., 2003, "Contemporary Physics", Taylor & Francis Ltd, volume 44, number 5, pp 401 – 416

Tannenbaum, B. and Stillman M., Understanding Sound, Mc. Grawhill, Newyork, 1973.

Truax, B., 1999, "Equal Loudness Contours" . Available on site: http://www.sfu.ca/sonic-studio/handbook/Equal_Loudness_Contours.html

Uyan , A., 2008, “Işık Tasarımı”.,
Available on site: http://www.tiyatronline.com/mutisik105.htm

Uyan , A., 2008, “Ses ve Efekt”.,
Available on site: http://www.tiyatronline.com/mutisik106.htm

Wisconsin, U., 2005, “Physics”. Available on site:
http://tycho.physics.wisc.edu/courses/phys103/spring08/Lectures/Lect27Print.pdf

Zobel, E.A., 2006, **“**Ray Optics, Light Refraction”. Available on site:
http://id.mind.net/~zona/mstm/physics/light/rayOptics/refraction/refraction1.html

CURRICULUM VITAE OF THE AUTHOR

Mete Yafet

Personal Information :

Birth Date:	20.06.1983
Birth Place:	İstanbul
Marital Status:	Single

Education :

High School	1998-2001	İstek Semiha Şakir High School
Undergraduate	2001-2005	Yeditepe University, Fine Arts Faculty Graphic Design
Graduate	2005-2006	Yeditepe University, Social Sciences Visual Communication Design

Working Experience:

2007-2008 Bukalemun Design – Visual Communication&Interface Designer

2005-2006 Grafitürk Design – Graphic Designer