IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.53

Volume 4, Issue 4, June 2024

A Comprehensive Review of Psychoacoustic Model for Audio Compression in Signal Processing

Prof . Rahul M. Mulanjkar[1] **and Mrs. Bhagyashree Balkrushna Bhoir**[2]

Professor, Department of Electronics & Telecommunication Engineering[1]

Student, Department of Electronics & Telecommunication Engineering[2]

JCEI's Jaihind College of Engineering, Narayangaon, India

Abstract*: A Psychoacoustic Model is a framework used in an audio signal processing to understand how humans perceive sound. This models are fundamental to modern audio compression techniques. They exploit the limitations and characteristics of human hearing to reduce data rates without significantly compromising perceived audio quality. This report explores the principles of psychoacoustic models, their applications in various audio codecs and their impact on audio compression efficiency.Psychoacoustic models of human auditory perception have found an important application in the realm of perceptual audio coding, where exploiting the limitations of perception and removal of irrelevance is key to achieving a significant reduction in bitrate while preserving subjective audio quality. To this end, psychoacoustic models do not need to be perfect to satisfy their purpose, and in fact the commonly employed models only represent a small subset of the known properties and abilities of the human auditory system. This paper provides a tutorial introduction of the most commonly used psychoacoustic models for low bitrate perceptual audio coding.*

Keywords: Psychoacoustic Model, Signal Processing, Audio Compression, Masking

I. INTRODUCTION

The field of psychoacoustics, which studies the psychological and physiological responses associated with sound perception, has significantly influenced modern audio signal processing. Audio compression aims to reduce the amount of data required to represent audio signals while maintaining quality. Traditional methods focus purely on mathematical signal representations. In contrast, psychoacoustic models leverages the limitations and characteristics of the human auditory system to efficiently encode audio signals, often for the purpose of compression. The psychoacoustic model allows for efficient and perceptually transparent audio processing, making it a cornerstone of modern audio technology.

II. MECHANISM

Psychoacoustics describes characteristics of auditory perception and relates them to the physics of the sound field. This chapter discusses psychoacoustical foundations that play a role in spatial audio. For psychoacoustic sound field synthesis these are critical bands and masking, source localization and principles of auditory scene analysis[1]. It can explain how a sharp clap of the hands might seem painfully loud in a quiet library but is hardly noticeable after a car backfires on a busy, urban street. This provides great benefit to the overall compression ratio, and psychoacoustic analysis routinely leads to compressed music files that are one-tenth to one-twelfth the size of high-quality masters, but with discernibly less proportional quality loss. Such compression is a feature of nearly all modern lossy audio compression formats. Some of these formats include Dolby Digital (AC-3), MP3, Opus, Ogg Vorbis, AAC, WMA, MPEG-1 Layer II (used for digital audio broadcasting in several countries), and ATRAC, the compression used in Mini-Disc and some Walkman models[6]. Psychoacoustics is based heavily on human anatomy, it involves some basic principles as follows:

A. Human Auditory Perception

The perception of sound is influenced by psychoacoustic phenomena, including masking, which affects how sounds are heard in the presence of other sounds. This complex interplay between the physical properties of sound and the

DOI: 10.48175/568

perceptual processes of the auditory system enables humans to enjoy and understand the rich world of sounds[2]. Human auditory perception is the process by which the ears and brain work together to interpret sound waves.The journey of sound begins when sound waves enter the ear and travel through the auditory canal, causing the eardrum to vibrate[9]. These vibrations are transmitted through the ossicles (tiny bones) in the middle ear to the cochlea in the inner ear. The cochlea, a spiral-shaped organ filled with fluid, converts these vibrations into electrical signals through the movement of tiny hair cells. These electrical signals are then transmitted to the brain via the auditory nerve[11] as shown in fig1 below:

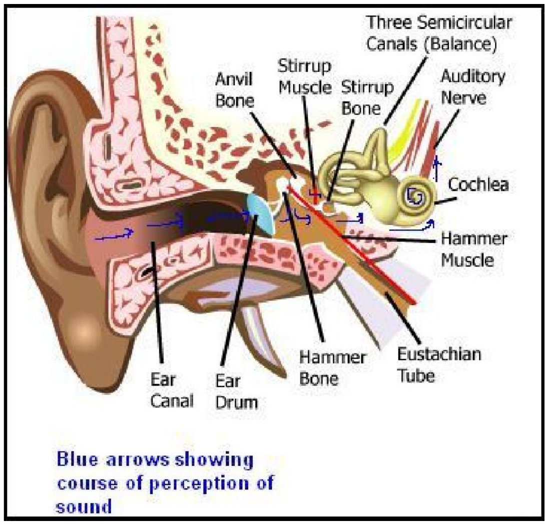


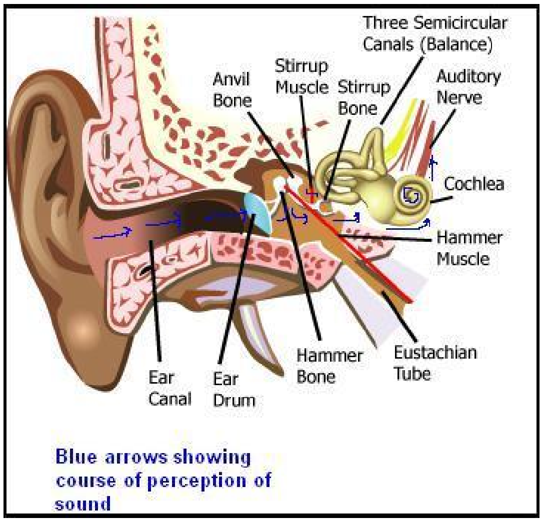
Fig. 1. Perception of Sound

The brain processes these signals in the auditory cortex, where it interprets aspects such as pitch, loudness and spatial location of the sound. Human hearing is sensitive to a wide range of frequencies, typically from 20Hz to 20KHz, with varying sensitivity across this range.

B. Masking Effect

A. Masking

Masking is a process to make weaker signal inaudible in the presence of louder signal. Therefore the louder signal is called as Masker and the weaker signal is called as Maskee. Also we can say that, the masking of soft sound by louder ones. Making is basically a part of our everyday experience[10]



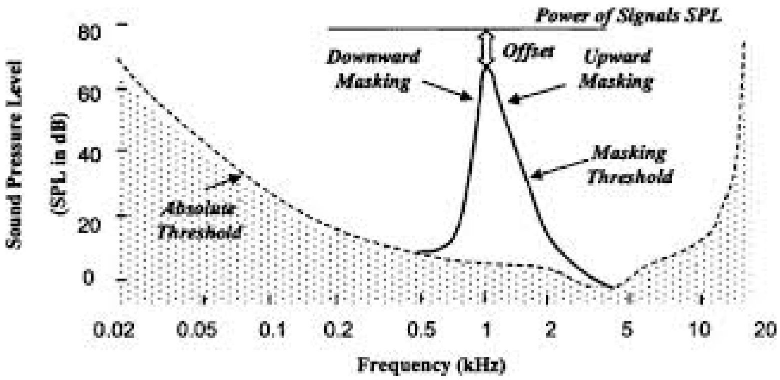


Fig. 2. Audio Masking Graph

From above fig.2 shows, the audio masking graph. Suppose a listener can hear a given acoustical signal under silent conditions. When a signal is playing while another sound is being played (a masker), the signal has to be stronger for the listener to hear it. The masker does not need to have the frequency components of the original signal for masking to happen. A masked signal can be heard even though it is weaker than the masker[12]. Masking happens when a signal and a masker are played together—for instance, when one person whispers while another person shouts—and the listener doesn't hear the weaker signal as it has been masked by the louder masker[8]. There are two types of masking happens in a psychoacoustic model for audio compression as follows:

2. Simultaneous Masking or Frequency masking

When both masker and maskee occurs at same time Or close in frequency and it is no longer possible to hear the normally audible maskee, then this phenomenon is called as simultaneous masking or frequency masking[7].For Example,Music and background noise, speech in noisy environments, audio compression, hearing tests, sound design in movies etc.

Non-Simultaneous Masking or Temprol masking

It is phenomenon in auditory perception where the presence of sound i.e masker affects the perception of another sound i.e maskee that occurs shortly before or after the masker. This effect occurs due to the temporal characteristics of the auditory system, which means sounds are not processed instantaneously and can influence each other over short periods of time[7].For example, Music production, speech perception, sound effects in movies and everyday life etc.

3. Threshold of Hearing

In the Psychoacoustic model, the threshold of hearing refers to the minimum sound pressure level that an average human ear can detect[10]. Typically measured in decibels(dB) relative to a reference sound pressure of 20 micro-pascals (µPa), this threshold represents the quietest sound a person can hear under ideal conditions. In the Psychoacoustics, this threshold is crucial for understanding how humans perceive sound, influencing areas such as audio compression, noise reduction and the design of hearing aids[12].

4. Critical Bands & Bark Scale

Bark Scale

The Bark scale is a psychoacoustical scale proposed by Eberhard Zwicker in 1961. It is named after Heinrich Barkhausen, who proposed the first subjective measurements of loudness. The mathematical expression for converting a frequency “f(Hz)“into Bark is given by[8]:

$\text{Bark}=13 \arctan(0.00076f) + 3.5 \arctan[(f/7500)2]$

$\text{Bark}=[(26.81f)/(1960+f)]-0.53$

$\text{Bark}=6[(\sinh-1)(f/600)]$

Critical band:

Determined by psychoacoustic experiments, their values typically vary depending on the type of response being measured. About 24 critical bands along basilar membrane.

Critical Band	Lower cut-off (Hz)	Central Frequency (Hz)	Higher cut-off (Hz)	Bandwidth (Hz)
1	0	50	100	100
2	100	150	200	100
3	200	250	300	100
4	300	350	400	100
5	400	450	510	110
6	510	570	630	120
7	630	700	770	140
8	770	840	920	150
9	920	1000	1080	160
10	1080	1170	1270	190
11	1270	1370	1480	210
12	1480	1600	1720	240
13	1720	1850	2000	280
14	2000	2150	2320	320
15	2320	2500	2700	380
16	2700	2900	3150	450
17	3150	3400	3700	550
18	3700	4000	4400	700
19	4400	4800	5300	900
20	5300	5800	6400	1100
21	6400	7000	7700	1300
22	7700	8500	9500	1800
23	9500	10500	12000	2500
24	12000	13500	15500	3500

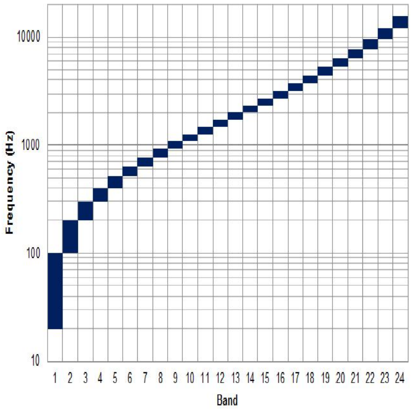


Fig. 2. Critical Band & Bark Scale According to Critical band

Fig3 shows, the basic concept of the Critical Bands and the Bark Scale according to the critical bands used in this model. Each critical band is about 1.3 mm long and embraces about 1300 neurons[8].

In the Psychoacoustics, the critical band concept is fundamental in understanding auditory masking and designing audio compression algorithms. Bark scale is essential for applications like speech processing, audio encoding and psychoacoustic modelling.

5. Perceptual Entropy

Perceptual Entropy is a concept used in psychoacoustic models to quantify the amount of information in an audio signal that is perceptually significant to human listeners. It leverages the principles of human auditory perception to determine which parts of a sound are most important for maintaining perceived audio quality. By understanding which elements of the sound are less likely to be noticed if removed or altered, perceptual entropy helps in efficient audio compression

This approach allows for the reduction of data size while preserving the quality of the sound as perceived by human ears, making it fundamental in technologies such as MP3 and other audio codes.

III. A PSYCHOACOUSTIC MODEL OVERVIEW

In generally, there are two basic psychoacoustic models are used in the audio and video compression technique. Out of those, Psychoacoustic model 1 uses a 512-sample analysis window for Layer I and a 1,024-sample window for Layers II and III. Because there are only 384 samples in a Layer I frame, a 512-sample window provides adequate coverage. Here the smaller window size reduces the computational load. Layers II and III use a 1,152-sample frame size, so the 1,024-sample window does not provide complete coverage[12]. While ideally the analysis window should completely cover the samples to be coded, a 1,024-sample window is a reasonable compromise. Samples falling outside the analysis window generally will not have a major impact on the psychoacoustic evaluation.

Psychoacoustic model 2 uses a 1,024-sample window for all layers. For Layer I, the model centers a frame's 384 audio samples in the psychoacoustic window as previously discussed. For Layers II and III, the model computes two 1,024-point psychoacoustic calculations for each frame. The first calculation centers the first half of the 1,152 samples in the analysis window, and the second calculation centers the second half. The model combines the results of the two calculations by using the higher of the two signal-to-mask ratios for each subband. This in effect selects the lower of the two noise-masking thresholds for each sub band[10].

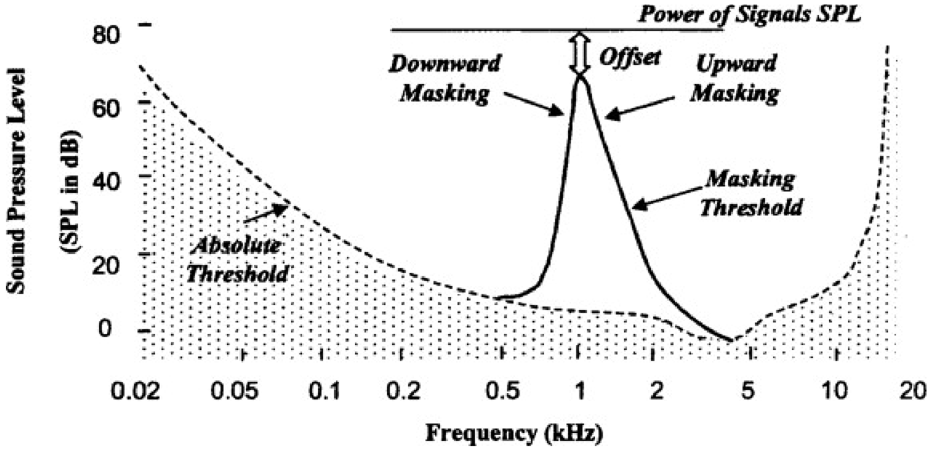


Fig. 2 An Overview of A Psychocoustic Model

IV. FUTURE DIRECTION

- **Improved Perceptual Models**: Continued research into human auditory perception can lead to more accurate and efficient models, incorporating more refined understanding of temporal and spatial masking effects.
- **Machine Learning and AI**: Utilizing machine learning and AI can help develop adaptive psychoacoustic models that can learn from a wide range of listening scenarios and individual hearing profiles.
- **Real-Time Processing Enhancements**: Advances in hardware and software optimization will enable more complex psychoacoustic models to be used in real-time applications without excessive computational demands.
- **Integration with Advanced Audio Formats**: Future models will need to integrate seamlessly with emerging audio formats and technologies , such as immersive 3D audio and high resolution audio formats.
- **Personalization**: Developing models that can adapt to individual listener's hearing profiles and preferences, potentially through user calibration processes, can enhance perceived audio quality.
- **Inter-Disciplinary Research**: Collaborations between audio engineers, neuroscientists and psychologists can foster new insights and methodologies for improving psychoacoustic models.
- **Standardization and Testing**: Establishing standardized benchmarks and testing protocols for psychoacoustic models can ensure consistent and objective evaluation of their performances across different platforms and applications.
- **Sustainabililty** : Research into more energy efficient algorithms can reduce the environmental impact of audio compression technologies, making them more sustainable for widespread use.

By addressing these challenges and pursuing these future directions, the field of psychoacoustic modelling for audio compression can continue to evolve, providing compression techniques

V. CONCLUSION

Psychoacoustic models revolutionize audio compression by leveraging the intricacies of human hearing. Despite their complexity, they are indispensable for efficient and high-quality audio compression. Ongoing research and technological advancements promise further enhancements in this field. The psychoacoustic model refers to computational frameworks that aim to mimic human auditory perception for various applications, technologies like MP3.

REFERENCES

[1]. Fastl, Hugo; Zwicker, Eberhard (2006). "Psychoacoustics: Facts and Models." Springer. pp. 21–22. ISBN 978-3-540-23159-2.

[2]. Ziemer, Tim (2020). Psychoacoustic Music Sound Field Synthesis. Current Research in Systematic Musicology. Vol. 7. Cham: Springer. doi:10.1007/978-3-030-23033-3. ISBN 978-3-030-23032-6. ISSN 2196-6974. S2CID 201136171.

[3]. Moore,B.C.J. (2012).”An Introduction to the Psychology of Hearing”(6th ed.). Emerald Group publishing.

[4]. Johnston, J.D. (1988).”Transform coding of audio signals using perceptual noise criteria.” IEEE Journal selected areas in communications, 6(2), 314-323.

[5]. Herre, J., & Johnston,J.D.(1996).”Exploiting mask patterns in perceptual audio coding.” Proceedings of the IEEE international conference on Acoustics,speech and signal processing(ICASSP),12-15.

[6]. Zwicker,E.,& Fastl,H.(1999). “Psychoacoustics: Facts and Models.”(second edition), Springer.

[7]. Fletche,H.(1940).”Auditory Patterns”Review of Modern Physics,12(1),47-65,deals with concept of auditory masking and the critical band theory which are foundational concepts in psychoacoustics.

[8]. Moore,B.C.J., &Glasberg,B.R.(1983).”Suggested formulae for calculating auditory-filter bandwidths and excitations patterns.” The Journal Of the Acoustical Society Of America,74(3),750-753.

[9]. James O. Pickles,”An Introduction to the Physiology of Hearing” deals with the anatomy and physiology of the auditory system.

[10]. Stanley A. Gelfand.”Hearing: An Introduction to Psychological And Physiological Acoustics”.

[11]. Christopher J. Plack,”The Sense Of Hearing” gives an accessible introduction to auditory perception,frequency analysis,loudness,pitch perception and spatial hearing.

[12]. Gearge Mather,”Foundations of Sensation and Perception”deals with basic mechanisms of auditory perception and complex interactions.

[13]. E. Larsen and R.M. Aarts (2004), Audio Bandwidth extension. Application of Psychoacoustics, Signal Processing and Loudspeaker Design., J. Wiley.

[14]. Ziemer, Tim (2020). Psychoacoustic Music Sound Field Synthesis. Current Research in Systematic Musicology. Vol. 7. Cham: Springer. doi:10.1007/978-3-030-23033-3. ISBN 978-3-030-23032-6. ISSN 2196-6974. S2CID 201136171.

[15]. https://www.researchgate.net/figure/Critical-bands-for-the-Bark-scale_tbl1_363521752

[16]. https://www.researchgate.net/figure/The-BARK-scale-is-a-psycoacoustic-model-of-the-human-perception-of-loudness-in-relation_fig3_355664561