Critical Band Encoding Technology
Audio Encoding System
From Arbitron

Technical Overview

Document: 1050-1054
Revision: E
Date: February 2008

Table of Contents

1. INTELLECTUAL PROPERTY NOTICE 3

2. ABSTRACT 3

3. INTRODUCTION 4

4. CBET TECHNOLOGY FUNDAMENTALS 5

4.1. Conceptual Encoding Approach 5
4.2. Human hearing 6
4.3. Psychoacoustic masking 7

5. DEVELOPMENT OVERVIEW OF THE ARBITRON CRITICAL BAND ENCODING TECHNOLOGY 9

5.1. Background of CBET 9
5.2. Evolution of CBET Development 9
5.3. PPM Multi-layer Message Format 13
  5.3.1. History: 13
  5.3.2. Layer Definitions: 14
  5.3.3. Message Definitions: 14
  5.3.4. PPM Detection Features: 16
  5.3.5. Additional Note Worthy Items: 16

6. CBET PERFORMANCE 17

6.1. Transparency 17
6.2. Robustness 18
6.3. A/D and D/A Conversions 19
6.4. Compression and Data Reduction Techniques 19
6.5 MPEG 21
6.6 Internet 22
6.7 Adaptive Transform Coding (ATRAC) 22
6.9 Dolby AC2 and AC3 systems 23
6.10 Resilience 23
6.13 Frequency Notches and Hopping 24
6.14. Tamper Resistance 24

7. IMPLEMENTATION 25

8. TECHNICAL CONTACTS 26

9. ARBITRON INC: THE COMPANY 27

10. ARBITRON RESEARCH AND TECHNOLOGY CENTER 27

1. INTELLECTUAL PROPERTY NOTICE

Before reading this document, please be aware that:

- Arbitron has received patents covering all of the key areas of its approach to audio encoding.
- The information contained in this document describes a proprietary system developed by Arbitron and is provided for your information. However, please be aware that any ideas you develop on audio encoding or watermarking systems may be influenced by your reading of this document and you should be aware of Arbitron's patents and proprietary rights.
- All the information contained within is copyrighted by Arbitron and may not be reproduced, copied or distributed to any third party without written consent of Arbitron.

2. ABSTRACT

This document outlines the Critical Band Encoding Technology (CBET) from Arbitron Incorporated. The document describes the background, operation and application of this technology in the domain of audio encoding for the purpose of audience measurement. Audio encoding is a mechanism for automatically identifying the source of a particular piece of material by embedding an inaudible code within the content. This code contains information about the content that can be decoded by a machine, but is not detectable by human hearing. Arbitron employs a number of patented technologies that ensure the code is unique, robust and resilient to all forms of processing or tampering while ensuring accurate decoding and inaudibility.

The CBET system has been developed with a significant investment from one of the world's most understated and successful research organizations. CBET is now in deployment and serving the broadcast industry as an audience measurement tool. In addition, it can protect and monitor intellectual property and assets in the rapidly emerging digital broadcast and distribution channels.

3. INTRODUCTION

The Arbitron Critical Band Encoding Technology (CBET) system is the enabling technology for embedding identification information within the content in the audio spectrum. This applies to music and broadcast signals alike, and does not affect the audio quality as perceived by the human ear. The CBET system is able to recover that information after transmission through common technologies and in the presence of normal acoustic noise. The CBET system provides a resilient, transparent, and indelible identity code that can be embedded into an audio program and captured either directly or acoustically. CBET is the cornerstone technology of Arbitron's Portable People Meter (PPM) system. Additional details on the PPM can be found on our website, www.arbitron.com.

There are several key requirements of audio encoding in order for it to be effective. It must be:

- **Transparent to user with no deterioration in original program material**
  The listener must not be able to hear the embedded signal in any way. Verification testing has been performed using sophisticated double-blind tests in a variety of environments to verify the transparency of the coding. Validation has continued using Arbitron on-site facilities, as well as external studies with the aid of industry audio experts.

- **Easy and cost effective to decode**
  Any embedded signal must be able to be decoded by electronics that may be supported by battery power and compact enough to be contained in a truly portable package. Cost of the decoder must be appropriate for large-scale deployment in survey panels. The portable decoder is a highly integrated design making use of the very latest in digital signal processing technologies.

- **Robust to all parts of the signal chain**
  The embedded signal survives any of the processes undertaken in the broadcast or distribution signal chain.

- **Easy to integrate into transmission and delivery environment**
  Equipment designed for the broadcast industry, fully automated and self-monitoring.

- **Able to be decoded both directly and acoustically**
  CBET is designed to support an acoustic decoder but the embedded code may also be decoded directly from the audio signal. In support of the audience measurement system, broadcast monitoring may be employed to verify the encoding in the audio signal.

4. CBET TECHNOLOGY FUNDAMENTALS

Encoding of audio relies on the ability of machines to distinguish information that is effectively transparent to human hearing. Over the years, many attempts have been made to achieve a transparent readable embedded signal that is resilient to all forms of treatment in the normal broadcast or delivery cycle as well as deliberate tampering. Only in the last few years has the processing technology been able to deliver these types of solutions.

Within the audio domain the approaches that have been attempted include: inclusion and exclusion of narrow frequency bands, modulation or patterning of the noise floor, spreading small amounts of specific frequencies amongst the program, and “hiding” signals by psychoacoustic techniques. Each of these approaches has some limitations, including detection by listeners, lack of resilience to tampering, lack of robustness to survive the signal chain, and more recently the inability to survive compression systems (e.g., Dolby AC3, MPEG, etc.). The approach Arbitron has taken is to combine multiple frequency coding and psychoacoustic masking in a unique fashion, and set as a key objective the ability to capture the embedded signal in a hostile acoustic environment.

4.1. CONCEPTUAL ENCODING APPROACH

To effectively embed a signal code in audio material requires that the audio be modified in a predictable manner such that a receiving device can be expected to accurately extract the embedded signals. If the modification of audio material is so small as to be imperceptible by the most discriminating listener, it would make the job of extracting the code difficult and also make the signal a candidate for loss in a compression algorithm or through a reproduction signal chain.

Placing the signal outside the frequency range of the human ear would render the signal inaudible, but it may not be able to be reproduced by common audio equipment and is likely to become lost in most, if not all, broadcast methods. The signal must be significant in amplitude relative to the audio material in order to survive compression algorithms and must fall within the bandwidth of normal broadcast audio in order to be reproduced at the receiver. In other words, it must be relatively loud and centered on the frequencies to which the ear is most sensitive. These requirements work against the goal of complete inaudibility.

An effective technical approach is to use the masking characteristics of the human ear to render the embedded signal inaudible but electronically detectable. Such a system would analyze the incoming audio source and dynamically modulate the embedded signal based on masking rules consistent with psychoacoustic masking techniques and the spectral content of the original material.

4.2. HUMAN HEARING

Much has been written regarding definitions of audibility. The human ear is said to have a frequency range of 20 Hz to 20 kHz. It is also accepted that the sensitivity of the ear is such that sounds with less energy than 0 dB Sound Pressure Level (SPL is from an empirically derived scale) are below the audible threshold (see Figure 1). This is expressed as “loudness” and reflects the variation in sensitivity of the human ear in response to different frequencies, as expressed by the common Fletcher Munson and similar loudness curves. We then have two dimensions, defined as frequency range and loudness.

A third dimension is also well understood, the “resolution of the human ear”, and is described in literature as the minimum frequency separation between two sinusoidal tones sounded simultaneously, which can be individually perceived.

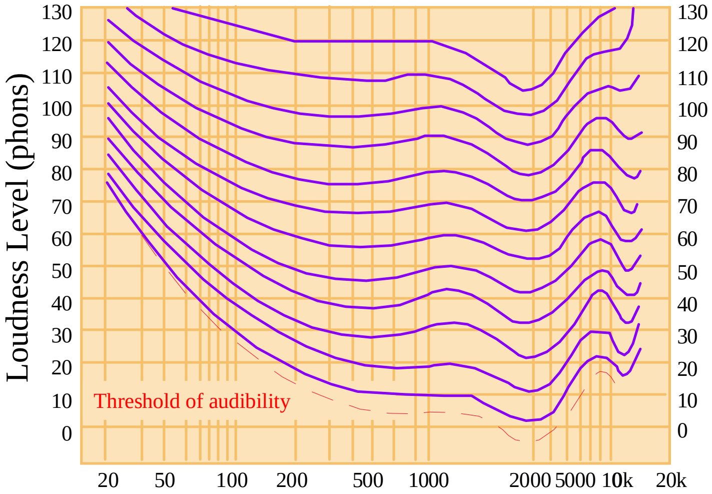

Figure 1

4.3. PSYCHOACOUSTIC MASKING

The impact and operation of masking can be understood in terms of the ear's mechanisms to discriminate tones (frequencies) from each other and those that gauge the magnitude of the sound energy (loudness). This ability to discriminate pitch and intensity is contained within the inner ear, the cochlea (see Figure 2). The cochlea is a tightly wound organ resembling a snail that when unwound reveals a semi-rigid material that is covered with small fibers and extends the length of the cochlea. This is the basilar membrane which provides the physical reason for frequency response (the length and relative location of the basilar fibers) and the loudness (deflection of the basilar membrane).

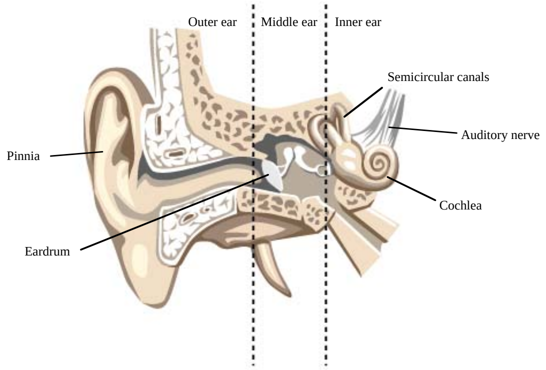

Figure 2

The membrane itself responds to loudness or intensity. The fibers vary in length, from very short at the base of the membrane to very long at the apex. The shortest fiber is sensitive to higher frequencies, and the longer fibers sensitive to the lower frequencies.

As the basilar membrane is a rigid material that will flex when excited by sound energy, the wave shape will be the same for sound energy that contains only one frequency and sound energy that contains that same frequency along with a series of other frequencies that are reduced in amplitude relative to that single frequency. In other words, the ear hears the same sound whether or not the surrounding (tertiary) frequencies have energy. If the amplitude of the tertiary frequencies is controlled to be lower than the masking curve, then the tertiary frequencies will not be heard by a human listener.

The presence of sound energy will cause the basilar membrane to be excited in such a way as to cause a wave to travel the length of the membrane, much the same as a guitar string is excited by mechanical energy. Very short waves correspond to the frequencies up to 20 kHz, and very long waves correspond to the lower frequencies, as low as 20 Hz. Large amounts of sound energy will cause a corresponding large deflection.

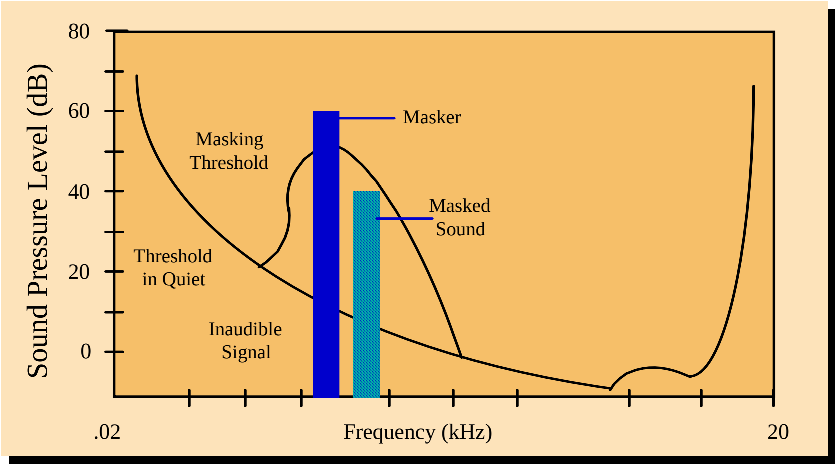

Figure 3

The basilar membrane is not capable of responding to two frequencies very close together when one is dominant (Figure 3). A technology such as the one described here is capable of determining the relative levels of sound regardless of the level of the dominant frequency (Figure 4). This technology works because there is a difference between the workings of the human ear and those of a suitably designed machine.

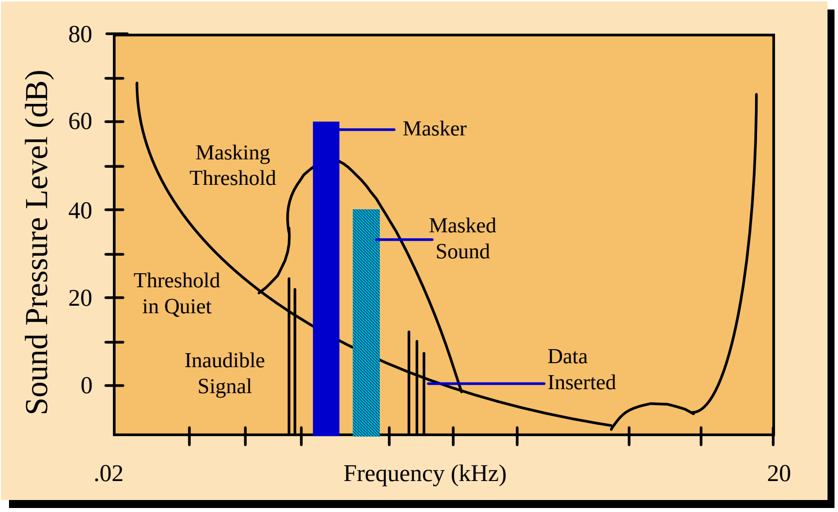

Figure 4

5. DEVELOPMENT OVERVIEW OF THE ARBITRON CRITICAL BAND ENCODING TECHNOLOGY

5.1. BACKGROUND OF CBET

Arbitron developed CBET in conjunction with Martin Marietta through proof-of-concept development and testing. Currently, CBET has multiple significant patents awarded and several others pending in the USA, Europe and worldwide. The CBET technical approach is to use the masking characteristics of the human ear to render the embedded signal inaudible but electronically detectable. The CBET system analyzes the incoming audio source and dynamically modulates the embedded signal based on rules consistent with psychoacoustic masking techniques and the spectral content of the original material. This approach provides the levels of transparency and robustness required for both acoustic and direct-coupled applications.

5.2. EVOLUTION OF CBET DEVELOPMENT

The CBET technology was originally envisioned with the purpose of better understanding when and what people listened to, along with the pursuit of improving audience measurement. For the audience measurement application, it was critical to develop a technology that could retrieve codes in the presence of environmental noise. This objective was defined by Arbitron in the mid-1980s.

The project took shape in 1992. A wide range of approaches were evaluated; low frequency and high frequency in-band encoding (20 Hz to 20 kHz), filter notching, and spread spectrum methodologies.

Simple “out-of-hearing-range” solutions that involve placing the encoding at higher-than-perceptible frequencies were rejected early on, due to the relatively narrow bandwidth of common broadcast techniques. Low frequency methods placing data in the 20- to 50-Hz range were also rejected due to the inadequate acoustic reproduction capabilities of most consumer equipment.

Filter notching was considered, and, in fact, a proof-of-concept implementation was developed (see Figure 5). This technique involves the removal of energy at specific frequencies to create a code pattern. This approach was ultimately rejected because code retrieval was not reliable in an acoustic implementation and the code was susceptible to destruction or removal by digital compression technologies.

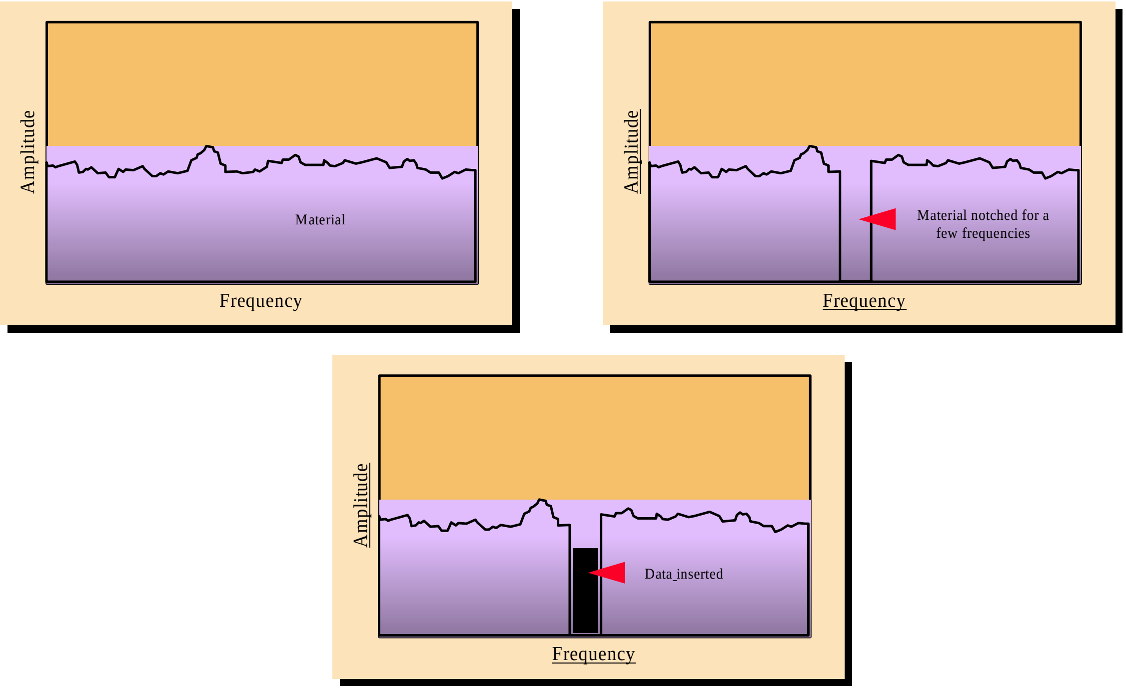

Figure 5

In parallel with the filter notching proof-of-concept, spread spectrum techniques were explored to a proof-of-concept stage (see Figure 6). For this method, a low-level signal is inserted into the noise floor. There are a wide variety of techniques for accomplishing this, and Arbitron pursued several with input from Martin Marietta in 1992 and 1993. A spread spectrum approach was pursued through first working prototypes in 1993. However, the spread spectrum approach by itself proved vulnerable to defeat by compression technologies. In order to maintain inaudibility, the signal must be at or below the noise floor, which works against a high code retrieval rate and is likely to be removed in some types of compression methods. Additionally, the definition of the noise floor continues to move downward as better equipment comes to the marketplace. On the other hand, a higher-level signal that could ensure a high retrieval rate and survival through compression algorithms works against the goal of complete inaudibility, which was a primary objective.

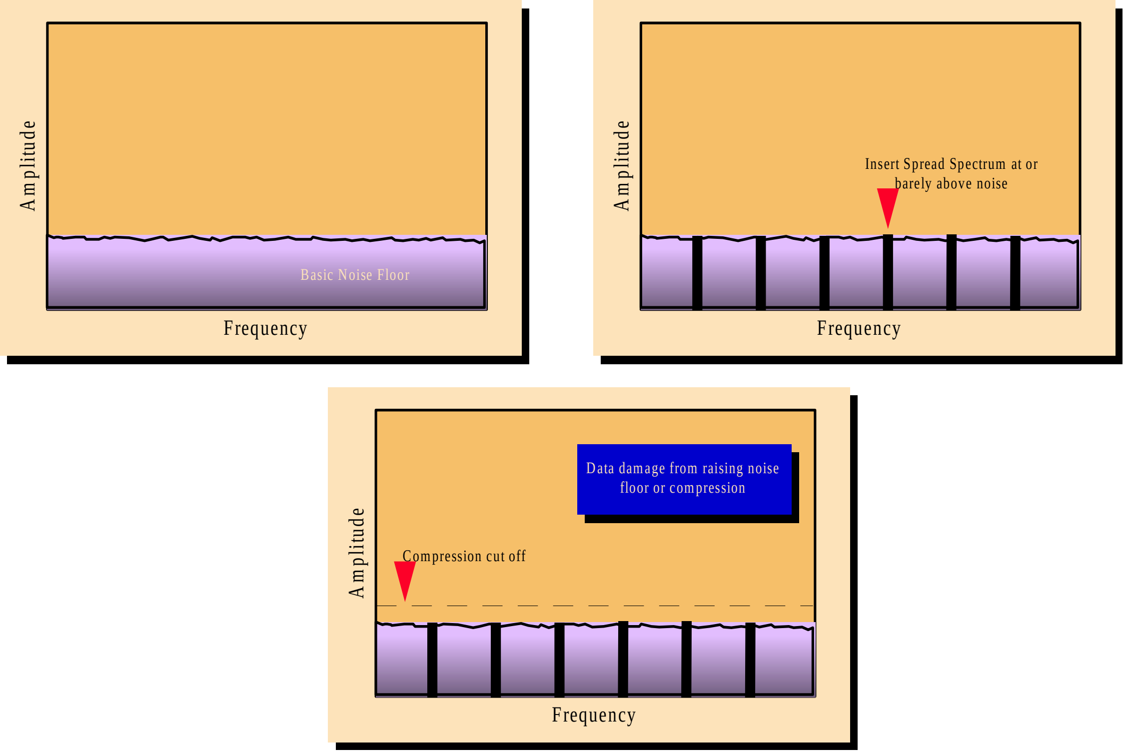

Figure 6

As a result, further innovation and development during 1993 resulted in the adoption and patenting of the current CBET approach that modulates a form of multiple frequency coding using masking rules consistent with the psychoacoustic principles established for human hearing.

Simple data transfer paradigms resembling bit serial binary systems were also tried and rejected early in the development process. Applying normal data transmission rules to the embedded data stream resulted in an unacceptable data transmission rate. CBET data transfer is based on a symbol repertoire that supports multiple layers of intelligence and, over a given time period, provides higher information transfer than do common binary systems. By mid-1994 portable acoustic decoder prototypes incorporating the current CBET approach were produced with the dimensions of a TV remote control.

By mid-1995 the algorithms had been further refined to result in a robust and resilient system. The acoustic decoder had been further developed into an operational unit. During this period, Arbitron undertook more than 20 field tests to establish the validity of the approach and ensure the transparency, robustness and resilience of the CBET technology.

In 1997 an encoder was developed that was capable of being placed into a studio’s analog broadcast chain. It required the use of four rack units in height.

In 1998 Arbitron built its Audio Lab on site to enhance its ability to develop and test new algorithms. It was designed by a New York based studio designer and is equivalent to industry mastering studios.

1999 and 2000 saw the enhancement to the decoding algorithms to improve detection of codes in harsh environments. The encoder hardware was enhanced and reduced in size by half. With these changes came a test deployment in Manchester, England.

With the success in Manchester came the decision to deploy in Philadelphia in 2001-2002. As part of the deployment came another group of system enhancements. These included:

- A new encoder design than reduced the size to only one rack height. Thus meeting the broadcast engineer’s requirement to minimize space necessary in the rack. Also available was a Digital AES/EBU input version as well as the Analog version.

- To satisfy the need to ensure the broadcast station was being encoded an In-Studio encoding monitor was provided. It is capable of providing an alarm if encoding is lost for any reason.

- The Base Station and Household Collector were reduced in size to 1/3 the size of the versions tested in Manchester. The units were much more acceptable to placement in a panelist’s home.

- To solve the issue of radio headphone use an accessory (Headphone Cap) was introduced.

2002 was a success in Philadelphia with much valuable data collected and presented.

Internationally PPM is being accepted into a number of countries for trial and deployment. These currently include Canada, Belgium, Singapore, and others that can not be noted here.

2003 was spent in enhancement of the system as well as continued testing. These include:

- A new version of the encoder that has an SDI interface. (A version for use with HDTV is under development for 2004 deployment.)

- A Portable Recharger has been developed to allow a panelist to take their meter with them for the weekend without data loss.

- A new feature has been added to give an At-Home indication.

- On going system evaluation has included testing of response rate improvements and confirmation of detection accuracy.

In 2004 multi-layer (3) capability was added and system deployment began in Houston.

2005 saw continued product enhancement and testing in Houston.

2006 saw the development of RoHS (hazardous material certified) to meet European standards. Panel deployment in Philadelphia also occurred.

In 2007 deployment in the top U.S. markets continued. The RoHS versions of the equipment began to be introduced.

In 2008 plans are to declare currency in multiple U.S. markets. A new version of the HUB for cellular only homes will be introduced.

In looking to the future, there will be new modernized version of the PPM and new encoder models to deal with changes in the radio and television broadcast pathways.

5.3. PPM MULTI-LAYER MESSAGE FORMAT

The PPM message format consists of three layers capable of being inserted by individual encoders in a parallel format. Thus at any time one, two, or three layers can be present in the encoded data stream. They are used for Network and Local Station identification. Features of each layer are identified in the following section.

5.3.1. History:

The encoding algorithm as implemented in originally Philadelphia test defined a layer of coding to be used for station identification, and a second 'synchronized' layer used for inserting a five minute Time Stamp (inserted at a lower energy level).

The refined requirement identified was to have a more granular time stamp (one minute) and separate layers for Local and Network identification, as well as a separate layer for Content identification (prime usage for commercials). Reporting was to be done on a 15 second segment (Philadelphia was on a 30 second segment). These additional requirements were implemented without change to neither inaudibility nor detection.

The following sections define what has been implemented for PPM Multi-layering. This definition has been tested in the lab and field for inaudibility and detection and found to be acceptable. It is presently implemented in all fielded panels.

5.3.2. Layer Definitions:

There are three layers defined for PPM encoding. The PPM X is designed to detect and report each of these layers during a 15 second period.
The layers are defined for Content, Network, and Local identification. A representation of these layers is as shown below:

1-3 Kilohertz Frequency Band

Local Symbol Codes (Including one minute Time Stamp)

Network Symbol Code (Including one minute Time Stamp)

Content Symbol Code

5.3.3. Message Definitions:

- Local Symbol Codes: The message format for this layer is as follows (Note all symbol definitions use unique frequency definitions from the Network layer.):

S1	**D1**	**D2**	**D3**	**S2**	**D1’**	**D2’**	**D3’**	**T1**	**T2**	**T3**	**T4**

- The symbol time is .4 seconds, resulting in a 4.8 second message

- S1 and S2 are unique synchronization symbols. Their definition has unique frequencies within the 1-3 kilohertz band.

- The data symbols (D1-D3 and D1’-D3”) are hexadecimal, with each possible symbol defined with a unique set of frequencies. The second group of three symbols is distinguished from the first by a 1 of 15 offset. As an example, if the code is 1 2 3 with an offset of 2, D1-D3 would be 1 2 3, and D1’-D3’ would be 3 4 5.There are approximately 50,000 possible unique codes.

- The Time Stamp is inserted as T1-T4. Each is a hexadecimal character, with T4 being the lowest order symbol. The count rolls over to zero after 28 days.

- Network Symbol Codes: The message format for this layer is as follows (Note all symbol definitions use unique frequency definitions from the Local layer.):

S1	**D1**	**D2**	**D3**	**S2**	**D1'**	**D2'**	**D3'**	**T1**	**T2**	**T3**	**T4**

- The symbol time is .4 seconds, resulting in a 4.8 second message

- S1 and S2 are unique synchronization symbols. Their definition has unique frequencies within the 1-3 kilohertz band.

- The data symbols (D1-D3 and D1'-D3") are hexadecimal, with each possible symbol defined with a unique set of frequencies. The second group of three symbols is distinguished from the first by a 1 of 15 offset. As an example, if the code is 1 2 3 with an offset of 2, D1-D3 would be 1 2 3, and D1'-D3' would be 3 4 5. There are approximately 50,000 possible unique codes.

- The Time Stamp is inserted as T1-T4. Each is a hexadecimal character, with T4 being the lowest order symbol. The count rolls over to zero after 28 days.

- Content Symbol Codes: The message format for this layer is as follows (Note all symbol definitions use a distribution of frequencies across the 1-3 kilohertz spectrum.):

S1	**D1**	**D2**	**D3**	**S2**	**D1'**	**D2'**	**D3'**

- The symbol time is .3 seconds, resulting in a 2.4 second message

- S1 and S2 are unique synchronization symbols. Their definition has unique frequencies within the 1-3 kilohertz band.

- The data symbols (D1-D3 and D1'-D3") are hexadecimal, with each possible symbol defined with a unique set of frequencies. The second group of three symbols is distinguished from the first by a 1 of 15 offset. As an example, if the code is 123 with an offset of 2, D1-D3 would be 1 2 3, and D1'-D3' would be 3 4 5. There are approximately 50,000 possible unique codes.

5.3.4. PPM Detection Features:

The PPM has been developed to detect all three layers and report any 15second segment that contains a code. Implementation of reporting is defined in an available Protocol Document.

- Key features of this implementation are:

    - For the Local and Network code layers a word length buffer (4.8 seconds) is used and data is accumulated prior to reporting. If a partial code is detected within the 15 second reporting period, it is reported and accumulation continues. If a full code is detected, it is reported and the accumulation register is cleared. If no full code is detected within five minutes of accumulation, the register is cleared and accumulation is restarted.

    - For the content layer the buffer is 15 seconds in duration. Accumulation beyond the rolling 15 seconds is not done. This method allows better detection of short encoded segments (example: commercials).

5.3.5. Additional Note Worthy Items:

- In each case the number of codes available for each layer is defined as approximately 50,000.

- Expectation is that Content layer encoding applications will require development of a software encoder and special application dependent software. This is required to ensure security of algorithms and codes, and operation on the pertinent workstation.

6. CBET PERFORMANCE

The three key areas of audio encoding that determine the viability of any potential encoding technology are:

- The inaudibility and transparency of the embedded signal;
- The robustness of the embedded signal to any signal processing in the transmission or distribution of the program material;
- The resilience of the embedded signal to any efforts to remove it from the program material.

6.1. TRANSPARENCY

Arbitron has undertaken a number of tests to ensure inaudibility of its encoding. At each stage of the algorithm development extensive on-site testing is done using our audio experts. This is followed by testing at external facilities with the assistance of industry audio consulting experts.

Arbitron has also undertaken two strictly monitored tests involving more than 100 people in each test, including industry professionals and consumers. The outcome of these tests was successful in that the participants could not differentiate between the encoded music and the unencoded music. The tests were repeated to ensure accuracy and validity of the results.

The approach of the tests was to establish whether in listening to a pair of audio samples the participants could distinguish any differences between them. The results were expressed whether they discerned no difference or whether one version sounded “better” than another. If they selected that one version was different, then they were requested to verify which source (A or B). The tests were designed as a double blind, with strict control of all possible variable parameters to ensure there was no bias or sway to the statistical analysis. A double-blind test is constructed such that neither the test participant nor the test administrator can know the “correct” answer. The participants selected CD program material from a collection covering all genres of music.

The results were analyzed to determine:

1. The number of selections correctly identified as having a difference between the versions and identifying the unencoded version as sounding better;
2. How each respondent answered overall;
3. How respondents with “good scores” (i.e., high selection of correct identifications) fared in consistently identifying one source as “better”.

The results convincingly demonstrated that people could not distinguish between encoded and unencoded program material.

As part of the implementation of test panels, encoders have been placed at radio and television stations to encode live broadcasts. The placement of these units has been preceded by individual testing by the responsible broadcast engineer. We have met their expectations. Encoders have been in continuous use in these environments since late 1997.

Specific extensive inaudibility testing has also been done by the BBC in the U.K. and multiple U.S. television network engineering staff. All testing has been successful.

6.2. ROBUSTNESS

To test the performance of the CBET system in a range of configurations found in audio replication, transmission and broadcasting, a selection of content, covering all the main musical styles, including speech, were encoded using the CBET system. The following conditions were then tested, all with positive results.

6.3. A/D AND D/A CONVERSIONS

In the tests, numerous A/D and D/A conversion techniques were studied. This included CD-R and DAT. In these tests, all the embedded codes were detected and successfully recovered without error. A/D conversion has no perceived impact on the CBET system.

6.4. COMPRESSION AND DATA REDUCTION TECHNIQUES

Over the years there has been significant growth in the development and use of data reduction systems for audio and video. Some of these systems, such as MPEG, have become standards of current and future media and transmission technologies.

All of these techniques use a number of baseline technologies, sometimes in combination. These include predictive coding, sub-band coding, and spectral coding that are often combined with Huffman coding and nonlinear quantization or transform coefficients.

These techniques give data reductions from 49:1 to 2:1, producing equivalent bit rates from 28.8 kbits/sec. to 768 kbits/sec. The majority of those used for quality audio reproduction are in the range of 128 kbits/sec. to 384 kbits/sec.

All of the techniques make use of three attributes of human hearing. They are psychoacoustics masking, nonlinear sensitivity and critical banding. A discrete simplification of the last is given in Table 1.

The CBET system makes use of those frequencies from 1 kHz to 3 kHz, and is only affected by those mid-bands 8 to 15 in Table 1. As individual components of the frequency sets used for the embedded signal are only a few Hz wide, and are applied dynamically, the Data Reduction CODEC treats the embedded signal in the same way as the program content.

All of the technologies mentioned use some form of sub-band coding, with the critical area of 1 kHz to 4 kHz being handled with additional granularity, compared with the bass and treble components. Since that is also the region with the CBET-embedded signal, it will not be significantly affected if the program material is of listenable quality.

Critical Band (Hz)

Band	Low	High	Width
0	0	100	100
1	100	200	100
2	200	400	100
3	300	400	100
4	400	510	110
5	510	630	120
6	630	770	140
7	770	920	750
8	920	1080	160
9	1080	1270	190
10	1270	1480	210
11	1480	1720	240
12	1720	2000	280
13	2000	2320	320
14	2320	2700	380
15	2700	3150	450
16	3150	3700	550
17	3700	4400	700
18	4400	5300	900
19	5300	6400	1100
20	6400	7700	1300
21	7700	9500	1800
22	9500	12000	2500
23	12000	15500	3500
24	15500	22050	6550

Table 1: Discrete critical bands

6.5 MPEG

MPEG fundamentally uses the PASC (Perceptual Audio Sub-band Coding) method to achieve data reduction. All MPEG systems have four key components as identified below.

MPEG Layer	Layer-1	Layer-2	Layer-3 (MP3)
Time frequency mapping	Polyphase filter bank 32 sub-bands equally spaced	as 1	as 1 with DCT filters
Psychoacoustic model	512-point FFT	1024-point FFT	Polynomial FFT
Quantizer/encoder	6-bit scale	6-bit scale with three frames	Increased sophistication
Frame packer	4 bits per sub-band	4 bits low, 3 bits mid, 2 bits high	Includes bit reservoir
Optimum bit rate	384 kbits	256 kbits	Down to 64 kbits

Table 2

Although a basic principle of PASC and MPEG is the removal of redundant data, the CBET system is relatively unaffected by this process because of the narrow width of the multiple frequency bands used for the symbols.

The use of the error checking symbols, and the redundancy built into the continuous signal, provide additional resilience.

MPEG 2 compression of CBET-encoded material was undertaken at a range of bit rates covering from 64 kbits/sec. to 384 kbits/sec. The results show little degradation of detection.

Arbitron has conducted many tests with MP3 encoded audio. The tests have shown that CBET encoding and decoding works extremely well with MP3 material.

We believe these tests substantiate our view that MPEG encoding and the use of Perceptual Audio Sub-band Coding techniques have a negligible effect on the retrieval of CBET-embedded signaling

6.6 INTERNET

The Internet presents many challenges for audio and video coding. CBET audio encoded Webcasts have been successfully decoded. It is important to realize that Webcasts or audio transmissions via the Internet have many variables. These include countless codecs and various bit rates. We have moved beyond testing Internet transmissions of CBET encoded material. Our technology has worked with every broadcaster that has Webcast CBET encoded material to date.

6.7 ADAPTIVE TRANSFORM CODING (ATRAC)

The ATRAC system provides for a data reduction to approximately one-fifth of the original data rate. The ATRAC system is primarily used in the Sony Mini disc format. This uses 52 sub-bands in the frequency domain (after FFT), with finer bands in the low frequencies contrasting with the 32 equal-width sub-bands in the MPEG use of the PASC approach

ATRAC also uses a psychoacoustic transfer function in bit allocation to gain further data reduction resulting in a bit rate of 292 kbits/sec. The ATRAC system breaks the incoming signal into three Block Floating Units (BFUs). The CBET codes are present only in the BFU, which covers the 0-Hz to 5.5-kHz region, and as such are treated as part of the program.

Mini disc was tested in a range of configurations. The direct-to-Mini-disc test produced no degradation of encoding. Mini disk was then tested with multiple encode/decode cycles and again produced similar results. This was also confirmed when the Mini disc output was transferred to DAT again with no degradation.

A final test was the combination of Mini disc with AC3 Compression at 64 kbits/sec., to simulate low bit rate compression on a compressed format, such as might be used for future portable digital music systems, which produced little degradation of the encoded signal.

6.8 ADAPTIVE SUB-BAND CODING

Although adaptive sub-band encoding is used throughout data reduction techniques, the use of PASC has been evaluated in two products using this technology.

The first is DCC, where a form of PASC, similar to MPEG Layer-1 at 4:1 data reduction, is used. As is the case with MPEG, the use of PASC gives no discernible reduction in the retrieval of the embedded signal over multiple encode/decode cycles.

The second is the PAC (Perceptual Audio Coding) system developed by AT&T. The CBET system was used in single pass mode at a range of bit rates with no discernible reduction in retrieving the embedded signal.

6.9 DOLBY AC2 AND AC3 SYSTEMS

The Dolby system is a family of transform coders that are based on TDAC. The Dolby system makes use of the logarithmic spectral envelope of the signal to drive a perceptual model that dynamically allocates the bits. The signal is sampled in overlapping windows to produce sine and cosine transforms, which are then selectively combined in sub-bands which approximate the critical bands.

The CBET system is masked within the program material and, as such, any transform coefficient established for the program will include the embedded signal. The reconstruction of the program through inverse transforms will not affect the embedded signal.

The Dolby system was tested with both AC3 and AC2 CODEC's at a range of bit rates. The AC2 CODEC was tested at both 256 kbits/sec. and 384 kbits/sec. Resulting in very little degradation of signal.

The AC3 CODEC was tested at 128 kbits/sec. and 256 kbits/sec., producing a retrieval index of 98+. The bit rate was then lowered to 64 kbits/sec., with very low losses.
The effect of the Dolby CODEC on the CBET system performance appears to be negligible.

6.10 RESILIENCE

One of the key issues of any encoding system is the ability of the embedded signal to survive in a hostile environment. This includes deliberate attacks to remove the encoding signal from the program material as well as operation in noisy acoustic environments. Within the development of CBET, a key focus has been to be resilient in these circumstances. Although the CBET system has a dependence on the amount of program material energy in the 1-kHz to 3-kHz band, development to date has focused on the acoustic retrieval from material that is light in this band, in an ambient noise field.

CBET uses the program material as the carrier waveform but does not use program power level information to decide whether to place a message packet. The message packets are placed in a continuous stream, giving a data rate that is independent of the program material power level or type.

The amplitude of each of the frequencies making up a symbol is modulated by the program material through the use of the masking rules. The CBET system has a near perfect record of detection, giving an extremely low error rate.

6.11. NOISE

The embedded signal power levels allowed by the psychoacoustic masking rules provide robustness to levels of noise that render the program material unlistenable.

A function of the design of the CBET system is the ability to overcome background noise in acoustically coupled applications. Testing of the acoustic coupled application in noisy environments has repeatedly shown acceptable data recovery at background noise levels greater than the program material power level.

6.12. FREQUENCY RESPONSE MANIPULATION

The Arbitron system places the embedded signal in the mid-frequencies of 1kHz to 3kHz. Given a typical input source dynamic range, such as that from a CD source, 15-dB changes in the mid-control setting will not significantly disturb retrieval of the embedded signal.

6.13 FREQUENCY NOTCHES AND HOPPING

The close binding of the embedded signal with the program material, and the large number of individual frequency components involved with the symbol set for a layer, makes CBET relatively impervious to any attempt to notch out the frequencies used in embedding the signal.

Tests involving a set of frequency notches inserted into the program material and then iterated to different frequencies to achieve frequency hopping resulted in complete and accurate recovery and decode. The choice and spread of the CBET multiple frequencies would make it virtually impossible to impact on the symbols without seriously corrupting the music content.

6.14. TAMPER RESISTANCE

The close degree of integration of the embedded signal and the source material makes it extremely difficult to separate them. This makes the Arbitron system relatively impervious to any attempt to tamper with or corrupt the embedded signal. Any such attempt will effectively make the content unplayable.

Even with the use of professional audio processing tools, specific knowledge of the proprietary Arbitron algorithms would be required to identify the embedded signal in order to attempt to remove it. Arbitron believes that removal of the embedded signal, without destruction of the content, is virtually impossible.

7. IMPLEMENTATION

The current PPM equipment consists of an encoder, a portable decoder (the PPM or Portable meter) for the respondent to keep, a base unit/recharger for each meter, and a household data collector for each home.

Figure 9: Current generation of PPM equipment

The Encoder, also known as the Arbitron SGE, is 1RU (1 standard rack unit) or 1.75" high. Typically the encoder is installed at either a broadcaster's studio or transmitter facility. It fits into a standard 19-inch studio equipment rack and provides continuous, real-time stereo encoding of program material as it is broadcast from the station. This equipment is self-monitoring to ensure continuous operation and non-interference with the station's processes. Analog, Digital AES/EBU, SDI, and a HDTV SDI.

The In-Studio Monitor (not shown, but similar in size to the encoder) is deployed at an encoded station to monitor the CBET encoding in the on-air broadcast. If the proper code at the expected level is not detected an alarm is given.

The PPM is the size of a small cellular telephone. It consists of an especially sensitive audio transducer, digital signal processing (DSP) circuitry to analyze input for code, compliance-monitoring motion detector, extensive memory accommodating at least one day of event codes, and a rechargeable battery. The battery can operate one day without recharging.

The base unit extracts data from the PPM, recharges the battery in the PPM, passes data to the hub and offers immediate respondent feedback. Immediate feedback is given to the respondents about their compliance via messages over a LCD messaging screen in the base unit.

Data from each base unit is transmitted to the household data collector (hub) through the household wiring. The hub receives data from all of the base units in the household and passes this data to the central computer system over the household telephone line. (A Wireless version of the unit will be available in 2008.) Both the base unit and the hub have been designed for ease of installation and ease of use by the participants. All parts of the system meet RoHS certification requirements.

Also available but not shown is a Headphone Cap to be used when the panelist is using a headphone device, and a Portable Recharger to allow the panelist to take their meter with them for a weekend away, without loss of data.

8. TECHNICAL CONTACTS

Technical inquiries should be directed to:

United States
Arbitron Inc.
David Forr
Director, U.S. Encoding Operations
9705 Patuxent Woods Drive
Columbia, MD 21046
Voice: (410)312-8447
Fax: (410)312-8604
E-mail: david.forr@arbitron.com

International
Arbitron Inc.
Richard Fiasconaro
Director, International Customer Engineering
9705 Patuxent Woods Drive
Columbia, Maryland, 21046
United States
+1 (410) 312-8665 Phone
+1 (410) 312-8604 Fax
+1 (443) 858-5556 Mobile
E-mail: richard.fiasconaro@arbitron.com

9. ARBITRON INC: THE COMPANY

Arbitron Inc. is an international media and marketing research firm serving radio and TV broadcasters, cable companies, advertisers and advertising agencies, magazines, newspapers and the online industry in the United States and Europe.

Founded in 1949, Arbitron measures radio audiences in local markets across the United States, surveys the retail, media, and product patterns of local market consumers and provides application software used for accessing and analyzing media audience and marketing information data.

The company is organized into three operational groups: U.S. Media Services, Portable People Meter Development and Webcast Services and has approximately 700 full-time employees worldwide.

Scarborough Research is a joint venture of Arbitron Inc. and VNU Marketing Information. Scarborough Research conducts syndicated surveys of consumers in 75 local markets, providing information to newspapers, television stations, radio stations, cable systems, Internet companies, advertisers and agencies. Scarborough tracks more than 1,200 products, services and retail shopping categories into each local database reporting consumer retail shopping behavior, product consumption, media usage, lifestyle and demographic characteristics.

10. ARBITRON RESEARCH AND TECHNOLOGY CENTER

The Arbitron Research and Technology Center, which was established in 1949, supports all of the company's business units. This facility is credited with many of the media research innovations that have revolutionized the industry during the past five decades, including personal diaries, electronic meters, survey sampling techniques, and new definitions and descriptions of the radio and new definitions and descriptions of the radio and television audience.