



دانشگاه صنعتی شریف
دانشکده مهندسی کامپیوتر

بررسی الگوریتم‌های فشرده‌سازی و کاربردهای آنها

نگارندگان:
محمد مهدی عرفانیان

ارائه مطالب علمی و فنی

دکتر همت‌یار

۱۴ دی ۱۳۹۸

چکیده

در مستندی که پیش روی خواننده عزیز قرار دارد تلاش شده تا مختصراً الگوریتم‌های فشرده‌سازی مختلف و کاربردهای آن‌ها در زمینه‌های مختلف مهندسی کامپیوتر در راستای انجام پروژه درس ارائه مطالب علمی و فنی بررسی شود. این درس در پاییز ۹۸ در دانشکده مهندسی کامپیوتر دانشگاه صنعتی شریف توسط دکتر همتیار ارائه شده است.

برای سهولت کار استاد محترم درس برای تحصیل اطمینان از درستی مستندسازی و همچنین استفاده دانش‌جویان علاقه‌مند، سیر پیشرفت مستند به همراه کدهای $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ در *Github* قرار گرفته است، لازم به ذکر است که این مستند به صورت متن‌باز ارائه شده و استفاده از آن بدون ذکر منبع برای همگان آزاد است. در انتها از استاد محترم درس، دستیار آموزشی ایشان و خوانندگان محترم تشکر می‌کنم.

با آرزوی خوش‌وقتی برای تمامی خوانندگان این مستند
نگارنده

کلمات کلیدی: الگوریتم‌های فشرده‌سازی، فشرده‌سازی تصویر، فشرده‌سازی صوت

فهرست مطالب

۱	مقدمه	۱
۲	۱.۱ تعریف	۱.۱
۲	۲.۱ انواع الگوریتم‌های فشرده‌سازی	۲.۱
۲	۱.۲.۱ الگوریتم‌های Lossless	۱.۲.۱
۳	۲.۲.۱ الگوریتم‌های Lossy	۲.۲.۱
۵	۲ بررسی نحوه فشرده‌سازی در فرمت JPEG	۲
۶	۱.۲ DCT	۱.۲
۶	۱.۱.۲ تعریف	۱.۱.۲
۶	۲.۲ تبدیل دنباله‌ای از پیکسل‌ها به تابع	۲.۲
۶	۳.۲ نحوه کار JPEG	۳.۲
۶	۴.۲ یک مثال از فشرده‌سازی JPEG	۴.۲
۸	۱.۴.۲ Quantization Table	۱.۴.۲
۱۰	۵.۲ فشرده‌سازی Huffman	۵.۲
۱۰	۱.۵.۲ تعریف	۱.۵.۲
۱۲	۳ بررسی نحوه فشرده‌سازی در فرمت MP۳	۳
۱۳	۱.۳ مقدمه	۱.۳
۱۳	۲.۳ نرخ بیتی	۲.۳
۱۳	۳.۳ نحوه کار mp۳	۳.۳
۱۳	۴.۳ بررسی دستگاه شنوایی انسان	۴.۳
۱۴	۱.۴.۳ ضعف شنوایی بزرگسالان	۱.۴.۳
۱۴	۲.۴.۳ درک کمتر جزئیات صداهای کم	۲.۴.۳
۱۴	۳.۴.۳ آستانه شنوایی انسان	۳.۴.۳
۱۵	۴.۴.۳ اثر پوشش‌دهی صداهای بلند	۴.۴.۳
۱۵	۵.۳ فشرده‌سازی mp۳ در عمل	۵.۳

فهرست تصاویر

۳	۱.۱	شکل نمونه برای تست میزان کمپرس در فرمت PNG
۷	۱.۲	توصیف فرکانسی توابع $\cos(x)$, $\cos(2x)$
۷	۲.۲	توصیف فرکانسی تابع $\cos(x) + \cos(2x)$
۹	۳.۲	ماتریس فرکانس‌های استاندارد
۱۱	۴.۲	مراحل الگوریتم Huffman
۱۴	۱.۳	آستانه شنوایی انسان برای فرکانس‌های مختلف
۱۵	۲.۳	اثر پوشش‌دهی
۱۶	۳.۳	ساختار فایل MP۳

فهرست جداول

۳	۱.۱	حجم فایل نمونه در فرمت PNG
۴	۲.۱	حجم فایل نمونه در فرمت JPG
۱۰	۱.۲	نوعی نظیرسازی حروف با طول ثابت برای هر حرف
۱۰	۲.۲	نوعی نظیرسازی حروف با طول متغیر برای هر حرف

فصل ۱

مقدمه

توضیحی اولیه مبنی بر تعریف کلی فشرده سازی، انواع الگوریتمها و کاربردها

۱.۱ تعریف

الگوریتم‌های فشرده‌سازی، الگوریتم‌هایی هستند که می‌توان با استفاده از آن‌ها داده‌ها را طوری رمزنگاری کرد که در تعداد کمتری بیت نسبت به آرایش اولیه قابل ارائه باشند. برای مثال می‌دانیم که برای ذخیره هر بیت اسکی هشت بیت فضا لازم است، می‌توان با استفاده از نگاشتی متشکل از حروف استفاده‌شده در یک متن تعداد بیت‌های مورد نیاز برای نشان دادن هر حرف استفاده‌شده در متن را کاهش داد. با استفاده از این تکنیک در حقیقت متن را در قالب جدیدی فشرده کرده‌ایم.

۲.۱ انواع الگوریتم‌های فشرده‌سازی

در یک دسته‌بندی الگوریتم‌های فشرده‌سازی را به دو نوع زیر افراز می‌کنند.

- Lossless یا بدون هدررفت داده
- Lossy یا همراه هدررفت داده

۱.۲.۱ الگوریتم‌های Lossless

در این سری الگوریتم‌ها داده ورودی بدون هیچ‌گونه هدررفتی از داده خروجی قابل بازیابی است، الگوریتم‌های این دسته با استفاده از افزودن آمارهای تلاش می‌کنند تا نحوه نمایش داده را در نگاشتی به نحوه نمایش دیگری که به فضای کمتری نیاز دارد تبدیل کنند. این الگوریتم‌ها در مواقعی که ثابت ماندن داده در طی فشرده‌سازی الزامی است استفاده می‌شوند، همچنین معمولاً برای بازیابی اطلاعات فشرده‌شده نیاز به داده‌هایی خارجی است که با کمک آن عمل بازیابی انجام می‌گیرد، از این رو می‌توان از این نوع الگوریتم‌ها در رمزنگاری نیز استفاده کرد. یکی از کاربردهای اصلی این الگوریتم‌ها در فشرده‌سازی متون است که اشتباه شدن حتی یک حرف می‌تواند باعث بدخوانی و بدفهمی متن اصلی گردد. الگوریتم‌های مشهور کمپرس Lossless به شرح زیر اند.

- Run-Length Encoding (RLE)
- Lempel-Ziv (LZ)
- Huffman Encoding
- Burrows Wheeler Transform

البته لازم به ذکر است که در عمل از مجموعه‌ای از الگوریتم‌های فوق برای رسیدن به درصد مطلوب فشرده‌سازی استفاده می‌شود.

نمونه‌های الگوریتم‌های Lossless

در عمل از الگوریتم‌های Lossless در مواقعی که پایداری داده‌های ذخیره‌شده حیاتی است یا این که فایل در آینده به تعداد زیادی بار فشرده و گسترده می‌شود و از دست دادن قسمتی از داده در هربار فشرده‌سازی منجر به اختلاف فاحش نهایی خواهد شد استفاده می‌شود.

• PNG

در طراحی فرمت PNG برای فشرده‌سازی تصاویر از الگوریتم Lempel-Ziv-Welch (LZW) که الگوریتمی Lossless است استفاده شده. در شکل ۱.۱ و جدول ۱.۱ یک نمونه عکس در حالت فشرده‌نشده و فشرده‌شده با فرمت PNG و مقدار حجم آن در حالت‌های مختلف آورده شده است.

```

۸ ۱ ۸ ۵ ۱ ۰ ۰ ۱ ۱ ۸
۸ ۴ ۴ ۳ ۹ ۱ ۷ ۳ ۰ ۵
۸ ۹ ۷ ۰ ۵ ۲ ۸ ۹ ۲ ۰
۴ ۱ ۶ ۰ ۱ ۰ ۲ ۸ ۶ ۳
۱ ۴ ۷ ۲ ۱ ۵ ۴ ۷ ۵ ۶
۷ ۱ ۲ ۳ ۶ ۷ ۵ ۵ ۹ ۰
۴ ۸ ۲ ۱ ۲ ۲ ۱ ۶ ۲ ۶
۶ ۰ ۵ ۹ ۵ ۴ ۹ ۹ ۶ ۱
۲ ۳ ۹ ۹ ۵ ۴ ۵ ۴ ۰ ۹
۴ ۴ ۴ ۸ ۶ ۶ ۰ ۹ ۹ ۴

```

شکل ۱.۱: شکل نمونه برای تست میزان کمپرس در فرمت PNG

جدول ۱.۱: حجم فایل نمونه در فرمت PNG

Format	Size
BMP	۷.۷ مگابایت
PNG	۹۸ کیلوبایت

• Free Lossless Audion Codec (FLAC)

الگوریتمی که با استفاده از اطلاعات ذاتی داده‌های صوتی به فشرده‌سازی آنها می‌پردازد، نرخ فشرده‌سازی این الگوریتم با توجه به سطح فشرده‌سازی سازی آن معمولاً بین ۴۰ تا ۶۰ درصد می‌باشد اما در حالت بیشینه ممکن است تا ۸۰ درصد هم برسد.

۲.۲.۱ الگوریتم‌های Lossy

در این الگوریتم‌ها پس از هر بار فشرده‌سازی مقداری از داده‌ها از دست می‌روند، معیار ارزیابی این الگوریتم‌ها مقدار فشرده‌سازی با توجه به میزان هدررفت داده می‌باشد، به علت هدررفت مقداری از داده این الگوریتم‌ها معمولاً در مواردی که هدررفت اندک داده توسط انسان یا ماشین قابل تشخیص نباشد استفاده می‌شوند، مثلاً تکنیک‌های ذخیره‌سازی تصاویر و ویدئوها در کامپیوترها مبتنی بر الگوریتم‌های Lossy است زیرا چشم انسان قادر به تشخیص عوض شدن تعدادی پیکسل در صفحه پس از بازیابی فایل فشرده‌شده نیست.

معمولاً در فشرده‌سازی Lossy از Transform Coding استفاده می‌شود که داده‌ها را از فضای حقیقی به فضایی دیگر (معمولاً فرکانس) می‌برد و در آنجا از قسمت‌هایی از داده که تاثیرگذاری و حس‌پذیری کمتری نسبت به دیگران دارند صرف نظر می‌شود، سپس وارون تبدیل اجرا شده و داده‌های کوچک‌شده جدید با الگوریتم‌های Lossless فشرده می‌شوند. یکی از مشهورترین Transform Coding ها الگوریتم Discrete Cosine Transform (DCT) است که در فصول آتی این مستند بیشتر با آن آشنا خواهیم شد.

نمونه‌های الگوریتم‌های Lossy

الگوریتم‌های Lossy با همه‌گیر شدن اینترنت و اشتراک‌گذاری بیشتر مدیا در فضای اینترنت بسیار همه‌گیر شدند، از این رو اکثر این الگوریتم‌ها برای فشرده‌سازی فایل‌های صوتی-تصویری یا به اصطلاح Media

استفاده می‌شوند.

- JPEG
- MP3
- MP4
- H.26x

به عنوان نمونه برای الگوریتم Lossy حالت فشرده‌شده عکس ۱.۱ با فرمت JPEG با کیفیت‌های مختلف در جدول ۲.۱ آورده شده است.^۱

جدول ۲.۱: حجم فایل نمونه در فرمت JPG

Format	Size	Quality(%)
BMP	۷.۷ مگابایت	۱۰۰
PNG	۹۸ کیلوبایت	۱۰۰
JPG	۸.۹۶ کیلوبایت	۹۰
JPG	۹.۶۲ کیلوبایت	۵۰
JPG	۴۹ کیلوبایت	۲۰

^۱ در صورتی که تمایل به دیدن اصل فایل‌ها دارید می‌توانید به *Github* مستند مراجعه کنید.

فصل ۲

بررسی نحوه فشرده‌سازی در فرمت JPEG

توضیح نحوه کار فشرده‌سازی در فرمت JPEG، مقدمه‌ای بر DCT و توضیح کلی Huffman encoding

۱.۲ DCT

۱.۱.۲ تعریف

تبدیل کسینوسی گسسته (به انگلیسی: Discrete cosine transform)، دنباله‌ای محدود از اعداد (داده‌ها) را به صورت مجموع توابع کسینوسی با فرکانس‌های متفاوت نمایش می‌دهد. تبدیل کسینوسی گسسته، شباهت بسیاری به تبدیل فوریه گسسته (DFT) دارد، با این تفاوت که حاصل تبدیل فقط مقادیر حقیقی دارد (بر خلاف تبدیل فوریه که منجر به مقادیر مختلط می‌شود).
به صورت علمی‌تر می‌توان نوشت که DCT تابعی معکوس‌پذیر و خطی از R^N به R^N است. فرمول کلی DCT برای فضای یک‌بعدی به شکل زیر است.

$$X_k = \sum_{n=0}^{N-1} x_n \cos[\pi/n(n+1/2)k], k = 0, \dots, N-1 \quad (1.2)$$

۲.۲ تبدیل دنباله‌ای از پیکسل‌ها به تابع

هر تصویر در حقیقت به صورت چهار ماتریس دوبعدی ذخیره می‌شود که هر ماتریس مقدار رنگی پیکسل را برای آن کانال رنگی (RGB) و α نشان می‌دهد، برای کمپرس کردن یک عکس ماتریس‌های کانال‌های رنگی مختلف را جداگانه فشرده می‌کنیم، کانال رنگی R را در نظر بگیریم، در این حالت یک ماتریس $n \times m$ داریم که هر خانه آن عددی بین ۰ تا ۲۵۵ را نشان می‌دهد، برای سادگی یک سطر از این ماتریس را در نظر بگیریم، این سطر معادل با یک آرایه از اعداد می‌باشد، در صورتی که تمامی اعداد را از دامنه ۰ تا ۲۵۵ به دامنه ۱۲۸ - ۱۲۷ ببریم می‌توانیم این دنباله را با مجموعه‌ای از توابع کسینوسی بازتولید کنیم، این ضرایب همان ضرایبی است که با استفاده از DCT می‌خواهیم به دست بیاوریم.
شکل ۱.۲ نمونه‌ای از توصیف عکس با توابع کسینوسی را نشان می‌دهد.

حال اگر دو تابع نشان‌داده شده در شکل ۱.۲ را با هم جمع کرده و میانگین بگیریم، به نمایش فرکانس دیگری می‌رسیم که در شکل ۲.۲ نشان داده شده است. در صورتی که به شکل‌های متفاوت و با ضرایب متفاوت توابع مختلف کسینوسی را با هم جمع کنیم می‌توانیم هر فرکانسی را بسازیم. این کاری است که در عمل الگوریتم DCT برای ما انجام می‌دهد.

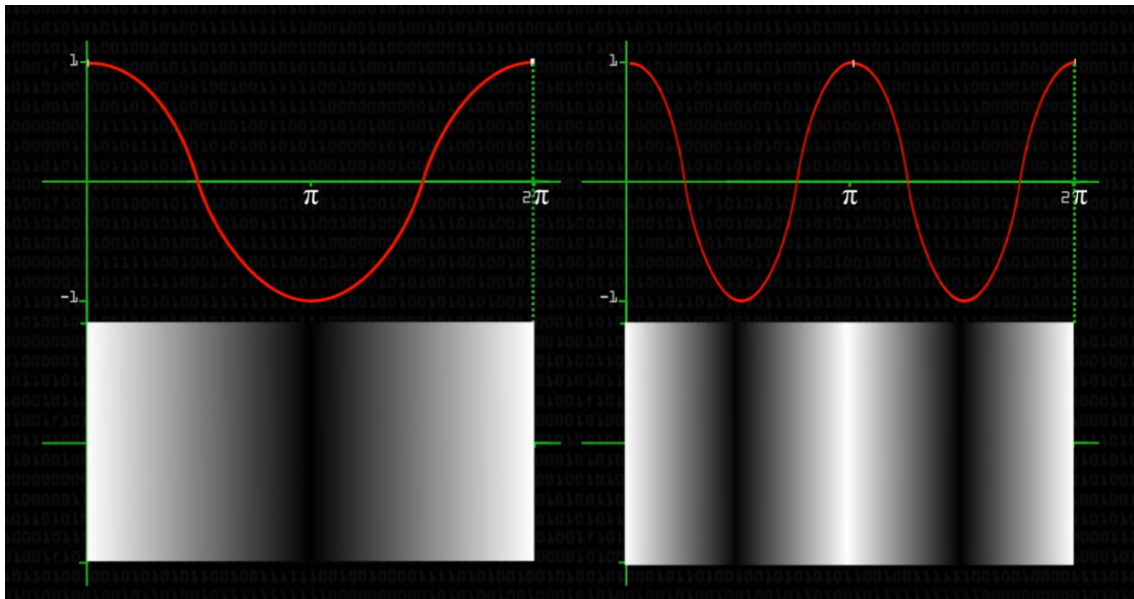
۳.۲ نحوه کار JPEG

برای فشرده‌سازی یک تصویر در JPEG مراحل زیر انجام می‌شود.

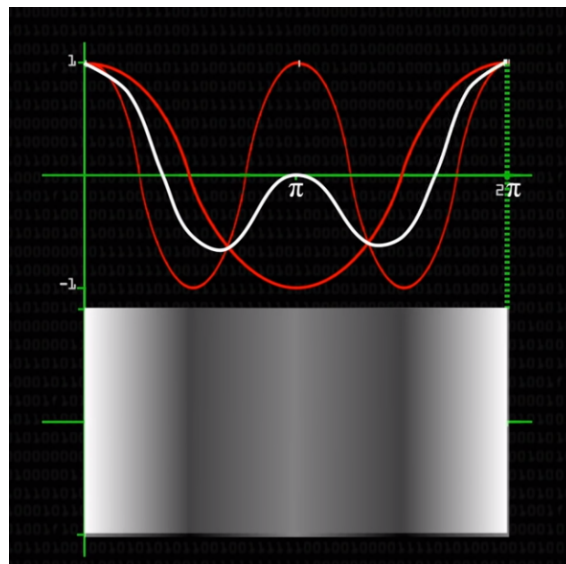
- تبدیل تصویر به بلوک‌های کوچک
- تبدیل هر بلوک به مجموعه‌ای از توابع کسینوسی استاندارد با DCT
- تنظیم ضرایب متناسب برای بلوک‌های DCT
- کمپرس کردن ضرایب با استفاده از Huffman

۴.۲ یک مثال از فشرده‌سازی JPEG

برای درک بهتر مراحل گفته شده تلاش می‌کنیم تا ماتریس زیر را (که در حقیقت می‌تواند یک کانال رنگی از قسمتی از یک تصویر باشد) را با الگوریتم‌های گفته‌شده فشرده‌سازی کنیم.



شکل ۱.۲: توصیف فرکانسی توابع $\cos(x)$, $\cos(2x)$



شکل ۲.۲: توصیف فرکانسی تابع $\cos(x) + \cos(2x)$

$$M = \begin{bmatrix} 52 & 55 & 61 & 66 & 70 & 61 & 64 & 73 \\ 63 & 59 & 55 & 90 & 109 & 85 & 69 & 72 \\ 62 & 59 & 68 & 113 & 144 & 104 & 66 & 73 \\ 63 & 58 & 71 & 122 & 154 & 106 & 70 & 60 \\ 67 & 61 & 68 & 104 & 126 & 88 & 68 & 70 \\ 79 & 65 & 60 & 70 & 77 & 68 & 58 & 75 \\ 85 & 71 & 64 & 59 & 55 & 61 & 65 & 83 \\ 87 & 79 & 69 & 68 & 65 & 76 & 78 & 94 \end{bmatrix}$$

هر عنصر این ماتریس عددی بین ۰ تا ۲۵۵ دارد اما از آنجایی که تابع کسینوسی مقادیر بین ۱- تا ۱ می گیرد نیازمندیم تا با کم کردن هر عنصر این ماتریس از ۱۲۸ هر عنصر این ماتریس را به عددی بین ۱۲۷- تا ۱۲۸ تبدیل کنیم. ماتریس تبدیل شده به شکل زیر است.

$$M_{shifted} = \begin{bmatrix} -76 & -73 & -67 & -62 & -58 & -67 & -64 & -55 \\ -65 & -69 & -73 & -38 & -19 & -43 & -59 & -56 \\ -66 & -69 & -60 & -15 & 16 & -24 & -62 & -55 \\ -65 & -70 & -57 & -6 & 26 & -22 & -58 & -59 \\ -61 & -67 & -60 & -24 & -2 & -40 & -60 & -58 \\ -49 & -63 & -68 & -58 & -51 & -60 & -70 & -53 \\ -43 & -57 & -64 & -69 & -73 & -67 & -63 & -45 \\ -41 & -49 & -59 & -60 & -63 & -52 & -50 & -34 \end{bmatrix}$$

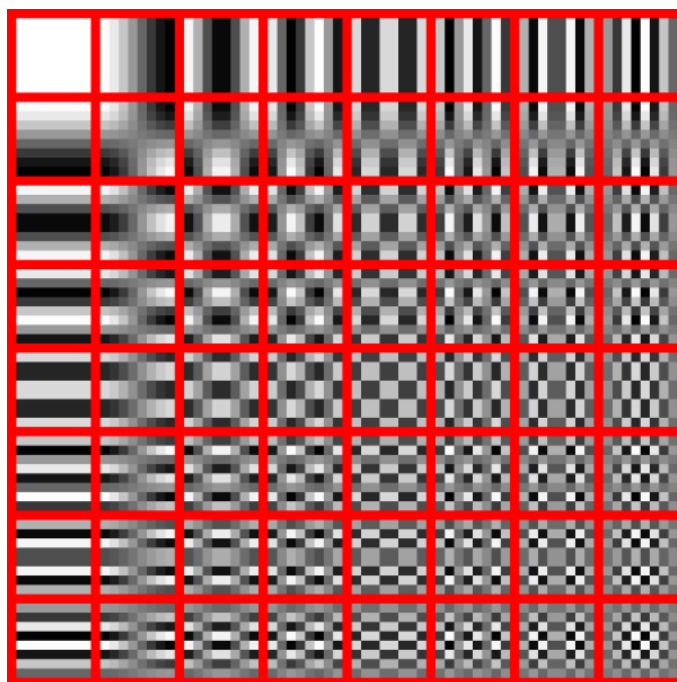
در گام بعدی مقادیر این ماتریس را با استفاده از الگوریتم DCT به فضای فرکانس می بریم هر مقدار از این ماتریس جدید برابر با ضریب فرکانس معادل در ماتریس استاندارد است.

$$F = \begin{bmatrix} -415.38 & -30.19 & -61.20 & 27.24 & 56.12 & -20.10 & -2.39 & 0.46 \\ 4.47 & -21.28 & -60.76 & 10.25 & 13.15 & -7.09 & -8.54 & 4.88 \\ -46.83 & 7.37 & 77.13 & -24.56 & -28.91 & 9.93 & 5.42 & -5.65 \\ -48.53 & 12.07 & 34.10 & -14.76 & -10.24 & 6.30 & 1.83 & 1.95 \\ 12.12 & -6.55 & -13.20 & -3.95 & -1.87 & 1.75 & -2.79 & 3.14 \\ -7.73 & 2.91 & 2.38 & -5.94 & -2.38 & 0.94 & 4.30 & 1.85 \\ -1.03 & 0.18 & 0.42 & -2.42 & -0.88 & -3.02 & 4.12 & -0.66 \\ -0.17 & 0.14 & -1.07 & -4.19 & -1.17 & -0.10 & 0.50 & 1.68 \end{bmatrix}$$

هر کدام از عناصر ماتریس بالا مقدار ضریب تاثیر فرکانس نشان داده شده در شکل ۳.۲ می باشند. تا به این لحظه هیچ مقداری از داده را برای فشرده سازی از دست نداده ایم اما در این گام می خواهیم قسمت هایی از تصویر که برای چشم انسان قابل تشخیص نیستند را حذف کنیم تا مقدار داده کمتری را ذخیره کنیم. باید توجه داشت که چشم انسان معمولاً قادر به تشخیص و تمیز تصاویر با فرکانس های بالا در تصاویر نیست و همان طور که در ماتریس هم مشاهده می شود ضریب این فرکانس ها نسبت به فرکانس های پایین بسیار کم است، می توانید به مقدار بسیار بزرگ ۴۱۵ برای فرکانس بسیار پایین در راستای x و y در ماتریس توجه کنید تا این نکته روشن شود. حال باید ضرایب فعلی ماتریس را با تقریبی گرد کنیم و تاثیر فرکانس های بالا را کمتر از تاثیر فرکانس های پایین قرار دهیم، از این رو از جدولی به نام Quantization Table استفاده می کنیم.

Quantization Table ۱.۴.۲

Quantization table جدولی است که در آن مقدار تاثیر هر فرکانس برای هر رده فشرده سازی برای



شکل ۳.۲: ماتریس فرکانس‌های استاندارد

عکس‌های JPEG به صورت جهانی استانداردسازی و تعیین شده‌است. برای مثال برای نرخ کمپرس ۵۰ درصد Quantization Table به شکل زیر است.

$$Q = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

حال برای آخرین گام باید مقادیر ماتریس Q را بر مقادیر ماتریس F تقسیم کنیم و عدد به دست آمده را گرد کنیم. ماتریس نهایی برای فشرده‌سازی به شکل زیر است.

$$R = \begin{bmatrix} -26 & -3 & -6 & 2 & 2 & -1 & 0 & 0 \\ 0 & -2 & -4 & 1 & 1 & 0 & 0 & 0 \\ -3 & 1 & 5 & -1 & -1 & 0 & 0 & 0 \\ -3 & 1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

همان‌طور که در ماتریس R مشهود است تعداد بسیار زیادی از عناصر ماتریس جدید مقدار صفر دارند (که بیشتر فرکانس‌های بالا را شامل می‌شوند) در ادامه فشرده‌سازی این ماتریس به وسیله الگوریتم فشرده‌سازی بدون هدررفت داده Huffman فشرده می‌شود و به همراه مقداری داده افزوده^۱ مانند درصد فشرده‌سازی

^۱ meta data

و... ذخیره می‌شود.

۵.۲ فشرده‌سازی Huffman

برای این که ماتریس مرحله آخر را به یک رشته متنی تبدیل کنیم می‌خواهیم از فشرده‌سازی یا الگوریتم Huffman استفاده کنیم. در این قسمت مختصراً الگوریتم Huffman برای فشرده‌سازی یک رشته متنی شرح داده می‌شود.

۱.۵.۲ تعریف

برای نمایش رشته در حالت ascii برای هر حرف ۸ بیت فضا گرفته می‌شود و تمامی حروف با یک آرایه هشت‌بیتی نمایش داده می‌شوند. رشته زیر را در نظر بگیرید.

$$S = \text{bananasc}$$

در این رشته در حالت ascii به $8 * 8 = 64$ بیت فضا نیاز داریم، در صورتی که بخواهیم از طراحی با طول بیت ثابت برای هر حرف در این رشته استفاده کنیم می‌توانیم نظیرسازی زیر را در نظر بگیریم.

جدول ۱.۲: نوعی نظیرسازی حروف با طول ثابت برای هر حرف

نماد	حرف
۰۰۱	a
۰۱۰	b
۰۱۱	n
۱۰۰	s
۱۰۱	c

در این روش برای نشان دادن هر حرف به سه بیت فضا نیاز داریم و در کل به $8 * 3 = 24$ بیت فضا نیاز داریم.

اما در صورتی که به صورت دقیق‌تر به رشته نگاه کنیم متوجه می‌شویم که تعداد تکرار هر حرف یکسان نیست و می‌توانیم برای نشان دادن هر حرف از تعداد بیت متغیر استفاده کنیم به شکلی که حروفی که بیشتر تکرار شده‌اند تعداد بیت کمتری بگیرند و حروفی که کمتر تکرار شده‌اند از تعداد بیت بیشتری برای ذخیره‌سازی استفاده کنند. مثلاً نظیرسازی زیر را برای این رشته در نظر بگیرید.

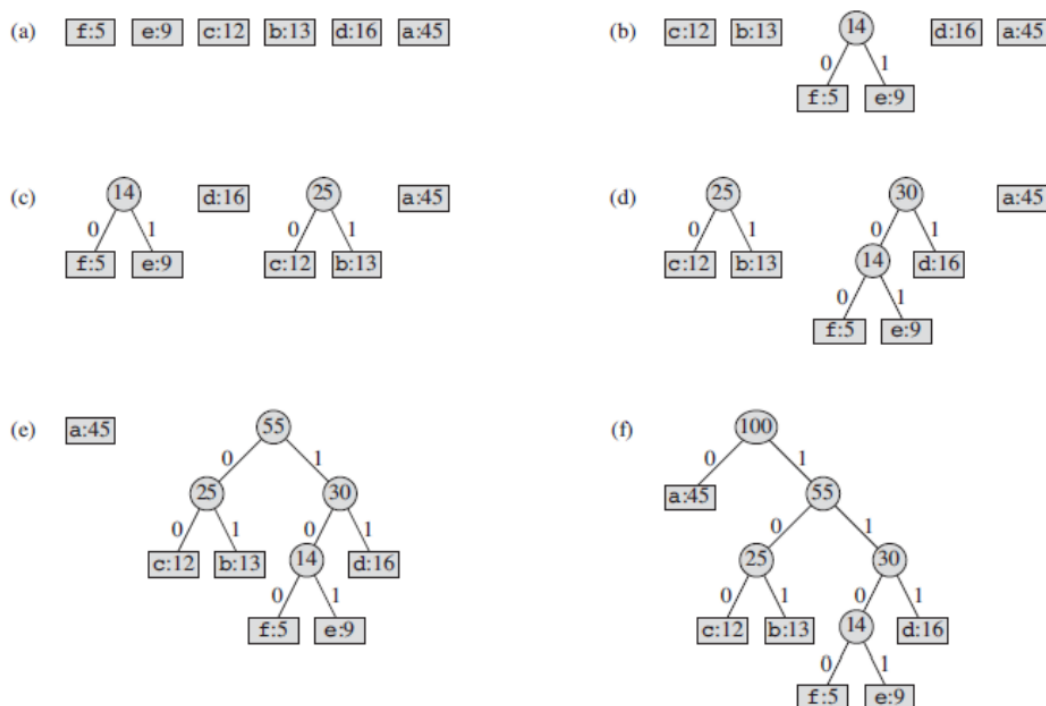
جدول ۲.۲: نوعی نظیرسازی حروف با طول متغیر برای هر حرف

نماد	حرف
۰	a
۱۱۰	b
۱۰	n
۱۱۱	s
۱۱۱۰	c

در صورتی که رشته اصلی را با این نظیرسازی فشرده کنیم به بیت‌های زیر می‌رسیم.

110010001001111110

همان‌طور که در رشته جدید مشهود است تعداد بیت‌های استفاده‌شده برای نمایش رشته اصلی به ۱۸ بیت کاهش پیدا کرده، نکته اساسی در این فشرده‌سازی این است که رشته به صورت یکتا قابل بازبینی باشد، در



شکل ۴.۲: مراحل الگوریتم Huffman

صورتی که به رشته بالا دقت کنیم متوجه می‌شویم که تنها حالتی که می‌توان با توجه به حروف جدول برای بازیابی بیت‌ها متصور شد همین حالتی است که به رشته اصلی منجر می‌شود، این اتفاق به این دلیل رخ می‌دهد که شروع هیچ حرفی زیرمجموعه پیشوندی هیچ رشته دیگری نیست، برای مثال در صورتی که حرف a با 0 و حرف b با 010 و نمایش داده می‌شدند برای رشته بی‌تی 010 دو حالت an و b می‌توانستیم متصور شویم، علت این اتفاق این است که رشته حرف a زیرمجموعه حرف b است.

کلیت الگوریتم Huffman پیدا کردن بهترین کدگذاری برای هر حرف در رشته اصلی است که طول رشته نهایی کم‌ترین اندازه را داشته باشد، برای این کار از شبه کد زیر استفاده می‌شود.

```

1 //Huffman Algorithm
2 n := |C|;
3 Q := C;
4 for i := 1 to n - 1 do
5     allocate a new node z
6     z.left := x := Extract-Min(Q);
7     z.right := y := Extract-Min(Q);
8     z.freq := x.freq + y.freq;
9     Insert(Q, z);
10 end for
11 return Extract-Min(Q); {return the root of the tree}

```

شکل ۴.۲ مراحل الگوریتم Huffman را نشان می‌دهد.

فصل ۳

بررسی نحوه فشرده‌سازی در فرمت MP۳

توضیح نحوه کار فشرده‌سازی در فرمت MP3، تحلیل صدای انسان و توضیح کانال‌های فرکانسی

۱.۳ مقدمه

در اواخر قرن بیستم و با شروع فراگیری اینترنت در جوامع مختلف، اشتراک‌گذاری مدیاهای مختلف مانند صوت، فیلم و تصویر در فضای اینترنت به یکی از خواست‌های همگانی و نیازهای اصلی مردم تبدیل شد اما مشکل اصلی در این میان، حجم زیاد فایل‌های مدیا و سرعت پایین انتقال داده در فضای اینترنت بود، فرمت‌های فشرده‌سازی مختلفی برای حل این معضل پیشنهاد شدند که هر کدام با تکیه بر تحلیل‌های آماری و یا حذف داده‌هایی غیرضروری در راستای کم کردن حجم فایل‌های مدیا می‌کوشیدند. در زمینه فشرده‌سازی فایل‌های صوتی فرمت MP۳ یقیناً فراگیرترین و موفق‌ترین فرمت به شمار می‌رود. در این فصل نگاهی کلی به نحوه کار این فرمت و الگوریتم‌های با هدررفت داده به کاررفته در این فرمت خواهیم داشت. برای تقریب اذهان به بزرگی حجم فایل‌های صوتی فشرده‌ناشده لازم به توجه است که هر فایل صوتی در حالتی که هیچ مقداری از داده فشرده نشود در هرثانیه حدود ۱۷۶۰۰۰ بایت فضا می‌گیرد.

۲.۳ نرخ بیتی

برای درک بهتر مفاهیمی که در ادامه متن با آن‌ها سر و کار داریم لازم است تا ابتدا با مفهوم نرخ بیتی^۱ آشنا شویم، در مفاهیم علوم کامپیوتر، نرخ بیتی معمولاً مقدار بیتی‌ست که در برای واحد زمانی ذخیره می‌شود، مثلاً وقتی bitrate یک موسیقی 320 kb/s اعلام می‌شود به این معنی‌ست که برای هر ثانیه از این صوت ۳۲۰ کیلوبیت داده ذخیره شده است؛ همچنین نرخ بیتی می‌تواند سرعت انتقال داده در یک کانال را بیان کند، مثلاً در طراحی مودم‌های شبکه از این مفهوم برای نمایش سرعت انتقال داده استفاده می‌شود.

۳.۳ نحوه کار mp۳

تمرکز فرمت mp۳ در فشرده‌سازی بر حذف اصواتی‌ست که توسط گوش انسان شنیده نمی‌شوند، همان‌طور که از فصل دو به یاد دارید برای فشرده‌سازی تصاویر می‌توانستیم از بیت‌هایی که نمایانگر تصاویری با فرکانس بالا بودند صرف نظر کنیم زیرا توسط چشم انسان قابل تشخیص نبودند، در فشرده‌سازی صوت نیز می‌توانیم با مطالعه ساختار شنوایی انسان اصواتی که به طور معمول توسط گوش انسان شنیده نمی‌شود را از صوت حذف می‌کنیم تا حجم فایل کاهش یابد.

به شکل خلاصه می‌توان مراحل فشرده‌سازی در فرمت mp۳ را به شکل زیر خلاصه کرد.

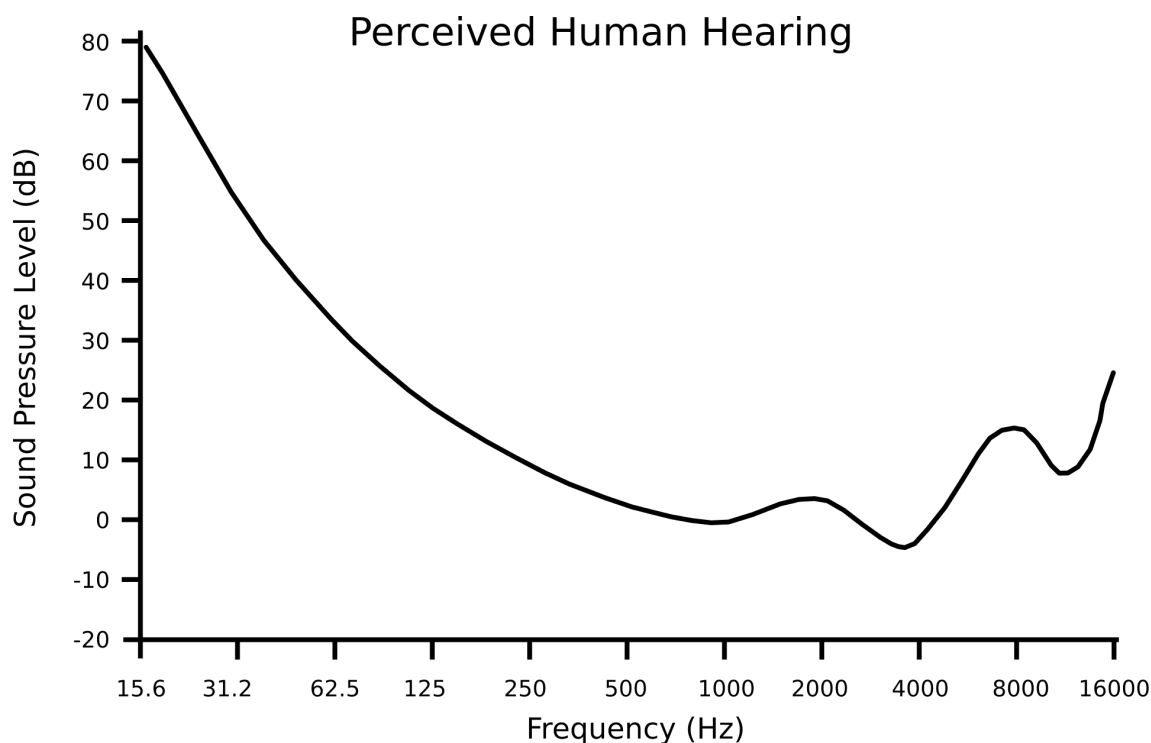
- تبدیل صوت به قسمت‌های کوچک
- حذف صوت‌های خارج از محدوده شنوایی انسان
- نمونه‌برداری از موسیقی با توجه به نرخ بیتی خواسته شده
- اضافه کردن افزونه‌ها و فشرده‌سازی نمونه

۴.۳ بررسی دستگاه شنوایی انسان

برای فشرده‌سازی صوت باید در ابتدا اصواتی که توسط گوش انسان قابل شنیدن نیستند یا گوش انسان با جزییات کمتری آن‌ها را درک می‌کند را حذف کنیم، برای این کار به تحلیل صوت‌شناسی^۲ نیاز دارد. نتایج تحلیل‌های صوت‌شناسی برای دستگاه شنوایی انسان موارد مفید زیر را اعلام می‌کند.

- ضعف شنوایی بزرگسالان

^۱ bitrate
^۲ Psychoacoustics



شکل ۱.۳: آستانه شنوایی انسان برای فرکانس‌های مختلف

- درک کمتر جزییات صداهای کم
- آستانه شنوایی انسان
- اثر پوشش‌دهی صداهای بلند

به تفصیل موارد بالا و کاربردهای آنان در فشرده‌سازی را بررسی خواهیم کرد.

۱.۴.۳ ضعف شنوایی بزرگسالان

در کودکی انسان‌ها معمولاً می‌توانند اصواتی را که بین فرکانس‌های ۱۰۰ تا ۲۰ کیلوهرتز باشند را بشنوند، اما با بزرگ شدن انسان معمولاً حد بالای شنوایی به مقدار ۱۵ کیلوهرتز می‌رسد و در صورتی که به درصد فشرده‌سازی بالایی نیاز داشته باشیم می‌توانیم از فرکانس‌های بیش از ۱۵ کیلوهرتز صرف نظر کنیم.

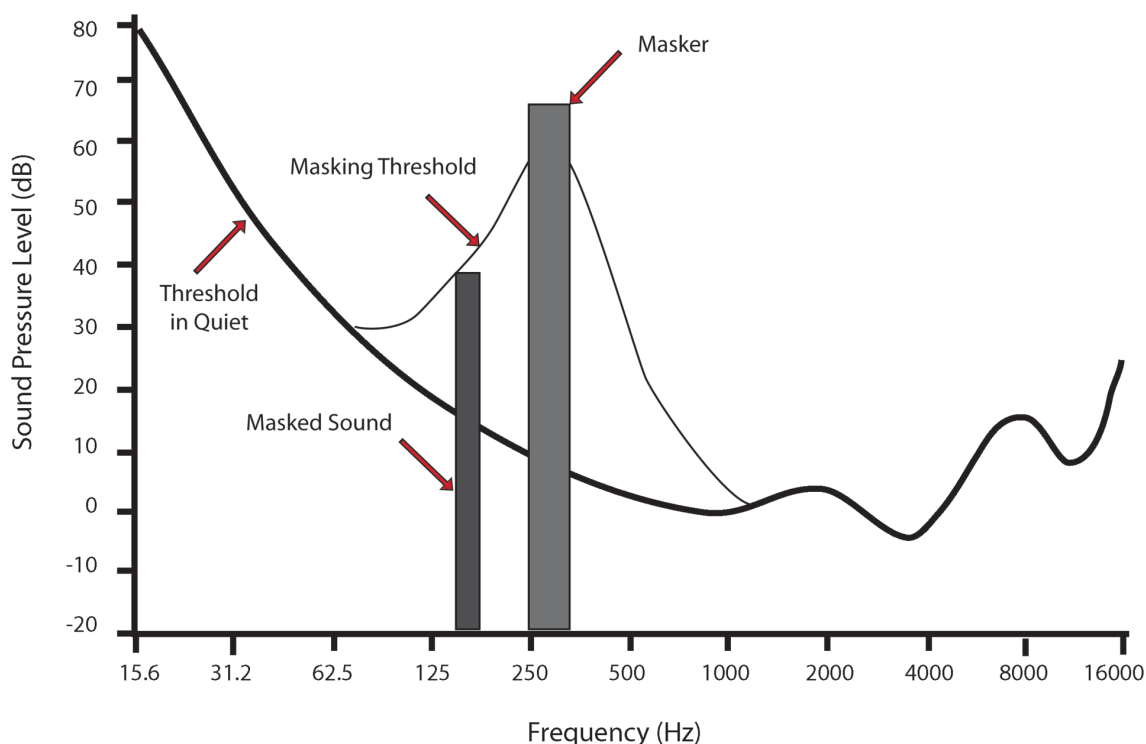
۲.۴.۳ درک کمتر جزییات صداهای کم

دستگاه شنوایی انسان معمولاً به جزییات صداهای بلند بیش از صداهای آرام توجه می‌کند، در نتیجه می‌توانیم برای صداهایی که سطح فشار صوت ۳ کمتری دارند از bitrate پایین‌تری استفاده کنیم.

۳.۴.۳ آستانه شنوایی انسان

گوش انسان برای هر فرکانس صوتی آستانه شنوایی دارد که اگر سطح فشار صدا از آن کمتر باشد آن را نمی‌شنود، شکل ۱.۳ این آستانه را برای فرکانس‌های مختلف نشان می‌دهد؛ در صورتی که سطح صوتی فرکانسی در هر قسمت از مقدار آستانه آن کمتر باشد آن صدا شنیده نمی‌شود و می‌توان آن را حذف کرد.

Sound Pressure Level (db)^۳



شکل ۲.۳: اثر پوشش‌دهی

۴.۴.۳ اثر پوشش‌دهی صداهاى بلند

هنگامی که در یک محدوده فرکانسی یک صدای بلند رخ دهد گوش انسان نمی‌تواند صداهاى آرام با فرکانس‌هاى نزدیک به آن را تشخیص دهد و بشنود حتی اگر سطح صوتی آن از آستانه شنوایی انسان بیشتر باشد، به این اثر در گوش انسان اثر پوشش‌دهی^۴ گفته می‌شود. مثالی از اثر پوشش‌دهی در شکل ۲.۳ نشان داده شده است.

۵.۳ فشرده‌سازی mp۳ در عمل

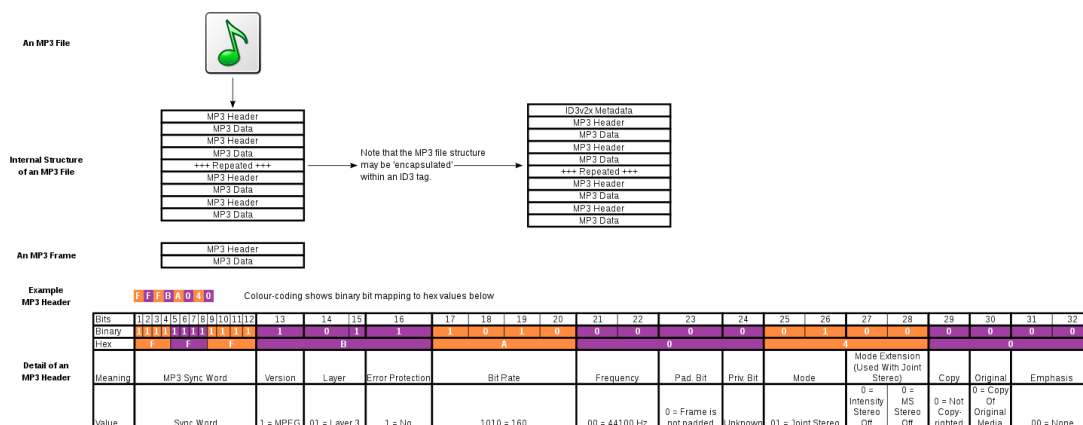
پس از بررسی انتزاعی اتفاقاتی که برای فشرده‌سازی در فرمت mp۳ می‌افتد نیازمند آنیم تا با یک مثال جزئیات فنی پیاده‌سازی این فرمت را نیز بررسی کنیم.

هنگامی که یک صوت برای فشرده شدن انتخاب می‌شود ابتدا به تکه‌های صوتی کوچک‌تر تقسیم می‌شود و با استفاده از نسخه اندکی تغییر یافته الگوریتم DCT که در فصل دوم معرفی شد به فضای فرکانس برده می‌شود، پس از آن اصواتی که فرکانس‌هایی پایین‌تر از آستانه شنوایی انسان دارند حذف می‌شوند و همین‌طور برای هر کانال صوتی عامل پوشش‌دهنده^۵ و عوامل پوشش‌پذیر^۶ شناسایی می‌شوند و بیت‌های مربوط به عوامل پوشش‌پذیر حذف می‌شوند، همچنین برای صداهاى بلند مقداری از جزئیات اصوات آرام‌تر را حذف می‌کنیم زیرا توسط ذهن انسان تشخیص داده نمی‌شوند.

پس از این تغییرات دوباره به فضای زمان برمی‌گردیم تا نمونه‌گیری را با توجه به bitrate خواسته شده انجام دهیم؛ دقت کنید که تا این‌جا کار هیچ مقداری از بیت‌هایی که توسط دستگاه شنیداری انسان قابل شنیدن باشند از دست نرفته‌اند.

در گام سوم با توجه به bitrate از هر قسمت کوچک ساخته شده نمونه برداری می‌کنیم و سپس به گام

Masking Effect^۴
Masker^۵
Maskee^۶



شکل ۳.۳: ساختار فایل MP۳

آخر می‌رسد، در این گام اطلاعاتی که برای بازیابی هر قسمت مورد نیاز است به همراه تعدادی بیت که برای تشخیص خطا قرار می‌گیرند را در Header هر قسمت قرار می‌دهیم و سپس داده اصلی را قرار می‌دهیم. مختصراً می‌توان گفت که هر بلوک داده موارد زیر را در خود دارد.

• سریتیر

- کد همگام‌سازی
- نسخه
- لایه صوتی
- بیت تشخیصی خطا
- نرخ بیتی
- فرکانس
- بیت padding
- بیت خصوصی
- حالت صوت
- داده کپی‌شده
- داده اصلی
- بیت تاکید

• داده صوتی

موارد بالا به صورت جدولی در شکل ۳.۳ آورده شده‌اند.