فشرده‌سازی وفقی سیگنال صحبت باند وسیع و صوت با استفاده از تبدیل موجک

سید طه مرتضوی و محمد حسن ساوجی

چکیده: در این مقاله طراحی یک کد کننده دیکد کننده جدید در نرخ بیت ۳۲ kb/s برای سیگنال صحبت باند وسیع و صوت بررسی می‌شود. این کدر جایگزین خوبی برای کدرهای باند وسیع قبلی مثل استاندارد G721 با نرخ بیت ۳۲ kb/s و G722 با نرخ بیت ٦٤ kb/s و MOS=٤/٢ می‌باشد. فشرده‌ساز یا کدر ما شامل قسمتهای کدگذار تبدیلی، مدل روان شنیداری، چندی‌کننده و قسمت کدگذار با طول متغیر است. در قسمت کدگذار تبدیلی از بسته موجکی که دارای باندهای خروجی نزدیک به باندهای بحرانی است استفاده شده است. تفاوت این قسمت با کارهای مشابه در استفاده از هسته تبدیل موجک توسعه یافته پارامتری جدید و نیز روشی است که شاخه های WP را گسترش داده‌ایم تا انطباق بیشتری با باندهای بحرانی شنوایی داشته باشند. فکر استفاده از مدل روان شنیداری را از MPEG1-Audio گرفته‌ایم اما به جای استفاده از طیف توان برای محاسبه نسبت سیگنال به ماسک S/M مستقیماً از داده‌های خروجی بسته موجکی استفاده کرده‌ایم. به این ترتیب، علاوه بر تطبیق مناسب خروجی‌های بسته موجکی با مدل روان شنیداری، از میزان محاسبات نیز کاسته شده است. در چندی‌کننده با توجه به تعداد بیت‌های هر باند بحرانی که قبلاً توسط مدل روان شنیداری حساب شده است به چندی کردن خروجی‌های بسته موجکی می‌پردازیم. در قسمت VLC، از روش کدگذاری آنتروپی استفاده کرده‌ایم. برای این کار از جداول دوباره کد کننده استاندارد JPEG استفاده شده است. اما تغییراتی برای تطبیق هر چه بهتر با شرایط سیگنال صحبت اعمال نموده‌ایم. کدر قابلیت استفاده وفقی از هسته موجک پارامتری را داراست. کدر با تغییر نسبت S/M قابلیت کم کردن نرخ بیت و کاهش کیفیت در حد کیفیت مورد نیاز را دارد. بنابراین، در جاهایی که احتیاج به نرخ بیت ثابتی باشد با تغییر S/M در اطراف نقطه کاری نرخ بیت به میزان خواسته شده می‌رسد. در نهایت این کدر با نرخ بیت ۳۲ kb/s کیفیت بسیار خوبی دارد که به راحتی از سیگنال PCM ورودی با نرخ نمونه برداری ١٦ kHz تعداد بیت ١٦ در هر نمونه قابل تشخیص نیست.

***کلید واژه:* فشرده‌سازی صحبت، بسته موجکی، مدل روان شنیداری، باند بحرانی، کدگذاری آنتروپی.**

۱- مقدمه

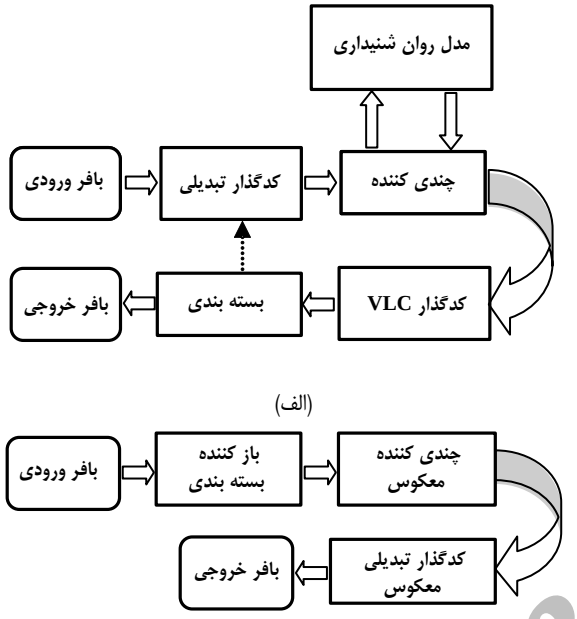
امروزه با ترافیک سنگین ارتباطات چند رسانه‌ای از طریق خطوط اینترنت و موبایل و حجم زیاد فایل‌های صوتی و تصویری و از طرف دیگر با محدودیت پهنای باند کانال و حجم حافظه روبرو هستیم. فشرده‌سازی داده‌های چندرسانه‌ای با در نظر گرفتن محدودیت کانال تنها راه حل برای حل این مشکلات است.

سیگنال صحبت شباهت زیادی با سیگنالهای صوتی دارد اما استفاده

این مقاله در تاریخ ۳۰ آذر ماه ۱۳۸۲ دریافت و در تاریخ ۱ آذر ماه ۱۳۸۳ بازنگری شد.

در زمان نگارش مقاله طه مرتضوی دانشجوی کارشناسی ارشد و در حال حاضر کارمند مرکز تحقیقات جهاد می‌باشد.

محمد حسن ساوجی، گروه الکترونیک، دانشکده برق و کامپیوتر، دانشگاه شهید بهشتی، اوین، تهران ۱۹۸۳۹۶۳۱۱۳ (e-mail: mh_savoji@yahoo.com).



شکل ۱: بلوک دیاگرام کلی کدر، (الف) انکدر، (ب) دیکدر.

وسیع از آن در خطوط اینترنت و موبایل باعث شده است که کدرهای مخصوصی برای آن طراحی شود. صدای انسان بر خلاف سایر امواج صوتی مثل موسیقی (که احتیاج به پهنای باند ۱۵ تا ۲۲ kHz برای داشتن کیفیت خوب دارند) با پهنای باند در حد ۷ kHz کیفیت مطلوبی دارد ـ کیفیتی در حد کنفرانس ویدئویی ـ بنابراین کدرهایی که مخصوص صحبت هستند توانایی فشرده‌سازی بیشتری دارند. به عنوان مثال در برخی از این کدرها برای آواهای صدای انسان مدلهایی درست شده است و با تغییر پارامترهای این مدلها که بسته به صدای هر شخص متفاوت است می‌توان صدای هر فردی را بازتولید کرد [۱] و [۲].

در ادامه این قسمت بلوک دیاگرام کلی کدر طراحی شده را بررسی می‌کنیم. در بخش‌های بعدی نحوه طراحی این قسمت‌ها و تفاوت آنها را با کارهای مشابه نشان می‌دهیم. شکل ۱ بلوک دیاگرام کلی کدر ما را نشان می‌دهد. اکثر کدرها برخی یا همه قسمت‌های شکل ۱ را دارا هستند. کدرهایی که مبتنی بر مدل آواهای صدای انسان هستند دارای سیستمی متفاوت از شکل ۱ هستند و در این مقاله بررسی نمی‌شوند.

هر کدری دارای دو قسمت انکدر (کدکننده) و دیکدر (دیکدکننده) است. انکدر وظیفه فشرده‌سازی سیگنال یا فایل ورودی را بر عهده دارد و دیکدر سیگنال بازسازی شده را از اطلاعات فشرده شده استخراج می‌کند.

مرحله اول در طراحی، تهیه سیگنال ورودی مناسب است. سیگنال ورودی در کدر ما داده‌های ۱۶ بیتی است که با فرکانس نمونه‌برداری

Archive of SID
www.SID.ir

۱۶ kHz از صدای با پهنای باند ۷ kHz نمونه‌برداری شده‌اند. به این ترتیب صدای ورودی دارای مشخصات صحبت باند وسیع و با کیفیت بالا است.

اولین بلوک کدر بلوک کدگذار تبدیلی است. در اینجا از تبدیلی استفاده می‌شود تا سیگنال را از حوزه زمان به حوزه دیگری نگاشت کند تا در این حوزه به طور مؤثری از نظر فشرده‌سازی سیگنال توصیف شود. یکی از این تبدیل‌ها تبدیل موجک است که سیگنال را به تعداد محدودی ضرایب پر انرژی و تعداد زیادی ضرایب کوچک و نزدیک به صفر تبدیل می‌کند [۳] تا [۵]. اگر فقط از ضرایب بزرگ و پرانرژی در هنگام بازسازی سیگنال استفاده شود سیگنال بازسازی شده دارای کیفیتی نزدیک به سیگنال اولیه خواهد بود زیرا قسمت اعظم انرژی سیگنال در همین تعداد محدود ضرایب بزرگ نهفته است. بلوک بعدی بلوک چندی کننده است. چندی کننده ضرایب حقیقی خروجی کدگذار تبدیلی را به اعداد صحیح تبدیل می‌کند تا با تعداد محدودی بیت بیان شوند. روشن است که در این قسمت مقداری از اطلاعات از بین می‌رود و ضرایب کوچک کدگذار تبدیلی به صفر تبدیل می‌شوند. وظیفه تخصیص‌دهی بیت به ضرایب خروجی کدگذار تبدیلی بر عهده مدل روان شنیداری است. این مدل از خواص شنوایی گوش انسان مثل پوشیدگی تُن‌های ضعیف مجاور تُن‌های قوی‌تر صدا و باندهای بحرانی شنوایی استفاده می‌کند [۶] و [۷]. مدل روان شنیداری، منحنی ماسک[1] را در باندهای بحرانی شنوایی به دست می‌آورد. با استفاده از این منحنی سطح نویز حاصل از چندی کردن که گوش انسان قادر به شنیدن آن نیست به دست می‌آید. مدل روان شنیداری با محاسبه نسبت سیگنال به ماسک در باندهای مختلف میزان بیت اختصاص داده شده به هر نمونه در هر باند بحرانی را به دست آورده و در اختیار بلوک چندی کننده قرار می‌دهد. داده‌های خروجی چندی کننده وارد قسمت کدگذاری VLC می‌شوند. وظیفه این قسمت اختصاص کدهایی با طول متغیر به اعداد مختلف برای فشرده‌سازی بیشتر آنهاست. در این قسمت بدون از دست دادن اطلاعات به فشرده‌سازی بیشتری دست پیدا می‌کنیم. روش‌های مختلفی برای پیاده‌سازی این قسمت وجود دارد [۴] و [۸]. همه این روش‌ها به کدهایی که احتمال اتفاق افتادن بیشتری دارند کلمات کد کوچکتر از حالت معمول و به آنهایی که به ندرت اتفاق می‌افتند کدهایی بزرگتر از حالت معمول اختصاص می‌دهند. به این ترتیب کاهش زیادی در حجم کلی اطلاعات ایجاد می‌شود. علاوه بر این همیشه با تعداد زیادی نمونه صفر مواجه هستیم. برخی از کدگذارهای VLC به جای کد کردن تک تک نمونه‌های صفر به تعداد نمونه‌های صفر قبلی و خانواده عدد غیر صفر که باید کد شود کلمه کدی اختصاص می‌دهند. این خاصیت که Run_Length نام دارد باعث فشرده‌سازی بیشتری می‌شود. در نهایت کدهای VLC به همراه یک سری اطلاعات کناری که در هنگام بازسازی سیگنال لازم است به صورت داده‌های باینری در کنار هم قرار داده شده و در یک بافر یا فایل واسطه ریخته می‌شوند. سپس اطلاعات بسته‌بندی شده برای بازسازی به دیکدر فرستاده می‌شود. دیکدر عکس کارهای قسمت انکدر را انجام داده و سیگنال بازسازی شده را تولید می‌کند.

۲- کدگذاری تبدیلی

کدگذار تبدیلی در کدر ما تبدیل بسته موجکی است. دو دلیل عمده برای انتخاب تبدیل بسته موجکی عبارتند از:

جدول ۱: گستره فرکانسی ۱۹ باند بحرانی.

Band Number	Frequency range (Hz)
۱	۰–۱۲۵
۲	۱۲۵–۲۵۰
۳	۲۵۰–۳۷۵
۴	۳۷۵–۵۰۰
۵	۵۰۰–۶۲۵
۶	۶۲۵–۷۵۰
۷	۷۵۰–۸۷۵
۸	۸۷۵–۱۰۰۰
۹	۱۰۰۰–۱۲۵۰
۱۰	۱۲۵۰–۱۵۰۰
۱۱	۱۵۰۰–۱۷۵۰
۱۲	۱۷۵۰–۲۰۰۰
۱۳	۲۰۰۰–۲۲۵۰
۱۴	۲۲۵۰–(۲۵۰۰)–۲۷۵۰
۱۵	۲۷۵۰–(۳۰۰۰)–۳۱۲۵
۱۶	۳۱۲۵–(۳۲۵۰–۳۵۰۰)–۳۷۵۰
۱۷	۳۷۵۰–(۴۰۰۰)–۵۰۰۰
۱۸	۵۰۰۰–(۶۰۰۰)–۶۵۰۰
۱۹	۶۵۰۰–(۷۰۰۰)–۸۰۰۰

۱- این تبدیل تمام خواص تبدیل موجک را برای توصیف مؤثر سیگنال‌های طبیعی که ماهیت غیر ایستان دارند را داراست.

۲- برای استفاده از خواص باندهای بحرانی شنوایی لازم است که شاخه‌های خروجی تبدیل موجک بر این باندها منطبق باشند. اما تبدیل موجک گسترش ثابتی داشته و از الگوی گفته شده پیروی نمی‌کند. بنابراین از تبدیل بسته موجکی که گسترش شاخه‌های خروجی آن انعطاف‌پذیر است استفاده کرده‌ایم [۶] و [۹]. جدول ۱ گستره فرکانسی باندهای بحرانی نتیجه بسته موجکی ما را نشان می‌دهد.

این جدول بیانگر وجود ۱۹ باند خروجی است. اینها باندهایی از محدوده شنوایی هستند که در پهنای باند ورودی ۷ kHz قرار می‌گیرند. تفاوت روش ما با کارهای مشابه در روش گسترش شاخه‌های درخت بسته موجکی است. در برخی باندهای بالا که امکان انطباق باند یک شاخه با باند بحرانی وجود ندارد از ضرایب دو یا سه شاخه که الزاماً در یک سطح گسترش قرار نگرفته‌اند استفاده می‌شود. به این ترتیب انطباق خوبی بین باندهای بحرانی شنوایی و کدگذاری وقفی که در اثر آن حداقل تخصیص‌دهی بیت‌ها بر حسب سیگنال ورودی انجام می‌گیرد به وجود آمده است.

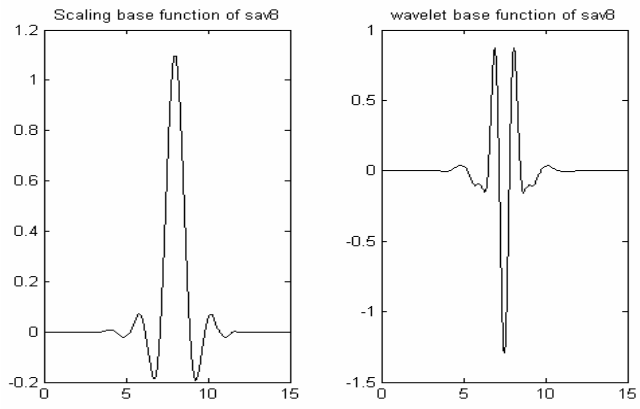
مرحله دوم انتخاب هسته یا تابع مادر مناسبی برای تبدیل بسته موجک است. کدر ما علاوه بر امکان استفاده از توابع موجک مشهور مثل دوبشی[2]، از یک هسته پارامتری جدید نیز استفاده می‌کند. این هسته از تقریب FIR تابع پایین گذر Raised Cosine به دست می‌آید. رابطه (۱) بیانگر تابع Raised Cosine است

$$T(f)=\begin{cases}1 & |f|\le (r-\beta)k \\ \cos\left[\dfrac{\pi}{4\beta}\left(\dfrac{f}{k}-r+\beta\right)\right] & |f|\le (r+\beta)k \\ 0 & |f|>(r+\beta)k\end{cases} \qquad (1)$$

1. Masking Curve

2. Daubechies

Archive of SID
www.SID.ir



شکل ۲: توابع پایه موجک و مقیاس 'sav8'.

که در آن $k = N/۲ \; ; k = ۰,...,$ و درجه فیلتر N می‌باشد.

ابتدا به خاطر اینکه فیلترها نصف باند[1] هستند پارامتر r برابر مقدار ثابت ۰/۵ قرار داده می‌شود. سپس این تابع برای قابل استفاده بودن در الگوریتم موجک سریع در حوزه فرکانس نمونه‌برداری شده و با گرفتن IFFT ضرایب فیلتر FIR تقریب یافته به دست می‌آیند. این ضرایب هسته موجک ما را تشکیل می‌دهند و ما آنها را با 'savN' نشان می‌دهیم. این ضرایب پس از نرمالیزه شدن ضرایب فیلتر پایین‌گذر ترکیب (Lo_R) در بانک فیلتری را ایجاد می‌کنند. با استفاده از روابط فیلترهای متعامد[2] فیلترهای دیگر به دست می‌آیند [۱۰]

$$Lo_R = \frac{w}{norm(w)} \qquad (۲)$$

$$Hi_R(k) = (-۱)^k \, Lo_R(۲N - ۱ - k) \quad k = ۰,... N/۲ \qquad (۳)$$

$$Hi_D(k) = Hi_R(N - k + ۱) \;\; ,k = ۰,...N/۲ \qquad (۴)$$

$$Lo_D = qmf\,(Hi_D) \quad (\text{qmf : Quadrature Mirror Filter}) \qquad (۵)$$

در روابط فوق k به جای f در رابطه (۱) قرار گرفته است.

به این ترتیب هسته موجک پارامتری به وجود می‌آید که توانایی کارکرد وفقی با تغییر پارامتر $\beta(۰\sim ۰/۵)$ را داراست. شکل ۲ توابع مادر موجک و تابع مقیاس[3] را که با استفاده از فیلترهای مرتبه ۸ یا 'sav8' با $\beta = ۰/۱۶۸$ به وجود آمده‌اند نشان می‌دهد.

بررسی‌های مختلفی که بر روی سیگنال‌های صحبت مختلف انجام شد نشان داد که هسته 'savN' در مرتبه مساوی با هسته‌های مشهور دیگر مثل 'dbN' به ازای برخی از βها فشرده‌سازی بهتری انجام می‌دهد. مثلاً 'sav8' و یا 'sav16' منجر به کاهش نرخ بیتی بین ۲۰۰ تا ۱۰۰۰ bit/s نسبت به 'db8' و 'db16' می‌شوند.

کدر طراحی شده با توجه به سرعت پردازش بهترین کرنل را در مرتبه مورد نظر از جدولی انتخاب نموده و در بسته موجکی استفاده می‌کند. کدر قابلیت استفاده وفقی از هسته جدید را نیز داراست به این ترتیب که برای هر قطعه یا segment سیگنال که در حال فشرده‌سازی است با استفاده از روش تجزیه از طریق ترکیب از بین چند هسته مرتبه ۱۶ با βهایی که قبلاً بهترین نتیجه را داده‌اند بهترین هسته را انتخاب می‌کند و کد آن هسته را نیز به دیکدر ارسال می‌کند.

1. Half Band
2. Orthogonal
3 Scaling Function

جدول ۲: خانواده های مختلف اعداد در جدول VLC استاندارد JPEG.

خانواده	جدول ضرایب
۰	۰
۱	۱، -۱
۲	۳،۲ -۲، -۳
۳	۴،...،۷ -۴،...،-۷
۴	۸،...،۱۵ -۸،...،-۱۵
۵	۱۶،...،۳۱ -۱۶،...،-۳۱
۶	۳۲،...،۶۳ -۳۲،...،-۶۳
۷	۶۴،...،۱۲۷ -۶۴،...،-۱۲۷
۸	۶۴،...،۲۵۵ -۶۴،...،-۲۵۵
۹	۲۵۶،...،۵۱۱ -۲۵۶،...،-۵۱۱
۱۰	-۵۱۲،...،-۱۰۲۳
A	۵۱۲،...،۱۰۲۳

۳- مدل روان شنیداری

مدل روان شنیداری ما شبیه مدل روان شنیداری MPEG1 Layer2 است، اما تغییراتی نیز نسبت به آن دارد [۱۱] و [۱۲]. در این مدل برای به دست آوردن میزان پوشندگی (masking) در باندهای مختلف مستقیماً از محاسبه انرژی ضرایب خروجی بسته موجکی استفاده می‌شود. به عبارت دیگر، از بلوکی موازی کدگذار تبدیلی برای به دست آوردن طیف سیگنال استفاده نمی‌شود. روش کار به این صورت است که منحنی‌های پوشندگی را در محور Bark[4] تقریب پاره‌خطی می‌زنیم [۱۳] و [۱۴]. این منحنی‌ها در محور Bark در باندهای مختلف رفتار منظم‌تری دارند. منحنی‌های تقریبی برای چند توان تُن در باند دهم به دست آمده‌اند و از انتقال این منحنی‌ها در محور Bark منحنی‌های پوشندگی در باندهای دیگر نیز به دست می‌آیند. برای به دست آوردن میزان تخصیص‌دهی بیت به ضرایب باندهای مختلف مراحل زیر انجام می‌گیرد:

۱- در هر باند بحرانی توان ضرایب موجک مربوط به آن باند را به دست می‌آوریم. سپس با فرض اینکه انرژی سیگنال در یک تن قوی در وسط این باند متمرکز شده است با استفاده از مدل تقریبی خود اثر پوشندگی این تُن را در باند فعلی و باندهای دیگر حساب می‌کنیم.

۲- مرحله ۱ را در مورد باندهای دیگر تکرار کرده و با جمع اثرات پوشندگی‌های باندهای مختلف و نیز تأثیر دادن اثر منحنی آستانه شنوایی منحنی پوشندگی نهایی را که نشانگر مقدار پوشندگی در ۱۹ باند مختلف است حساب می‌کنیم.

۳- مقدار سیگنال به ماسک را در باندهای مختلف بر حسب dB به دست آورده و از رابطه (۶) مقدار تخصیص‌دهی بیت را در باندهای مختلف به دست می‌آوریم

$$b_i = \frac{(S/M)_{dB}}{X}, \; i = ۱, ..., ۱۹ \qquad (۶)$$

b_i میزان بیت تخصیص داده شده به هر ضریب در باند i ام، $(S/M)_{dB}$ نسبت سیگنال به ماسک در باند iام و X عددی است که $(S/M)_{dB}$ بر آن برای به دست آوردن b_i تقسیم می‌شود. در صورتی که با محدودیت پهنای باند کانال روبرو باشیم این عدد در یک فرآیند تجزیه از طریق ترکیب تغییر داده شده و باعث تغییر در نرخ بیت نهایی و نیز کیفیت سیگنال بازسازی می‌شود. چندی کننده با استفاده از اطلاعات بردار **b** به چندی کردن ضرایب بسته موجکی می‌پردازد.

4. Critical Band Rate

Archive of SID

www.SID.ir

جدول ۳: نرخ بیت و SEGSNR در چند فایل صحبت نمونه در حالت ایستا با استفاده از هسته 'SAV16' و (الف) طول قطعه ۱ ثانیه، (ب) طول قطعه ۶۴ میلی ثانیه و (ج) طول قطعه ۳۲ میلی ثانیه.

الف: طول قطعه ۱ ثانیه

فایل ورودی	طول فایل (ثانیه)	SEGSNR (dB)	جنسیت گوینده	نرخ بیت متوسط (bit/s)
Test1	۵	۲۰/۹۰	مذکر	۲۱۲۳۷
Test2	۲	۲۷/۴۹	مذکر	۲۵۲۹۰
Test3	۴	۲۲/۴۷	مذکر	۲۱۲۵۲
Test4	۵	۱۹/۹۹	مؤنث	۱۸۱۸۵
Test5	۲	۲۰/۵۱	مؤنث	۲۰۸۲۶
Test6	۱	۱۹/۲۸	مؤنث	۱۹۳۹۴

ب: طول قطعه ۶۴ میلی ثانیه

فایل ورودی	طول فایل (ثانیه)	SEGSNR (dB)	جنسیت گوینده	نرخ بیت متوسط (bit/s)
Test1	۵	۲۰/۰۲	مذکر	۲۳۸۳۱
Test2	۲	۲۷/۰۹	مذکر	۲۸۷۴۱
Test3	۴	۲۱/۶۸	مذکر	۲۳۸۴۲
Test4	۵	۱۹/۱۵	مؤنث	۲۱۲۲۳
Test5	۲	۱۹/۷۸	مؤنث	۲۴۰۲۱
Test6	۱	۱۸/۹۴	مؤنث	۲۲۸۰۱

ج: طول قطعه ۳۲ میلی ثانیه

فایل ورودی	طول فایل (ثانیه)	SEGSNR (dB)	جنسیت گوینده	نرخ بیت متوسط (bit/s)
Test1	۵	۲۰/۱۵	مذکر	۲۷۰۶۱
Test2	۲	۲۶/۷۳	مذکر	۳۲۴۸۹
Test3	۴	۲۱/۷۴	مذکر	۲۷۷۱۹
Test4	۵	۱۹/۰۵	مؤنث	۲۴۲۷۲
Test5	۲	۱۹/۸۲	مؤنث	۲۷۱۶۰
Test6	۱	۱۸/۷۶	مؤنث	۲۵۶۴۵

٤- کدگذاری VLC

در قسمت کدگذار با طول متغیر یا VLC از کدهای استاندارد JPEG استفاده کرده‌ایم. این کدها دارای دو خاصیت اصلی کدگذاری آنتروپی و کدگذاری Run_Length هستند و باعث فشرده‌سازی خوبی می‌شوند. روش کار به این صورت است که اعداد به خانواده‌های[1] مختلف تقسیم شده‌اند. جدول ۲ یازده عنصر اول و خانواده‌های آنها را که در ساخت پیش کدها بکاررفته‌اند نشان می‌دهد. هر پیش کد JPEG به تعداد صفرهای قبلی و خانواده عدد غیر صفر کد شونده اشاره می‌کند. کدها علیرغم اینکه دارای طول متغیری هستند به خاطر دارا بودن خاصیت Prefix به هنگام استخراج رقم‌های باینری اشتباهاً به جای یکدیگر استخراج نمی‌شوند. حداکثر تعداد صفری که این کدها قبول می‌کنند ۱۵ است. این جدول ۱۱ عنصر اول جدول JPEG را که مورد استفاده قرار گرفته است، نشان می‌دهد. (البته یک کد خاص برای ۱۶ صفر یعنی عدد صفری که قبل از آن ۱۵ صفر قرار دارد وجود دارد که در اینجا مورد استفاده قرار نگرفته است.)

بنابراین در حدود ۱۵۰ پیش کد مختلف در جدول JPEG پیش‌بینی شده است. بعد از هر پیش کد نمایش باینری عدد کد شونده قرار می‌گیرد. اگر چه این عد باینری دارای خاصیت Prefix نیست اما به خاطر اینکه خانواده عدد توسط کد Prefix قبل از آن مشخص می‌شود طول کلمه کد نهایی مشخص بوده و هنگام استخراج توسط دیکدر درست استخراج می‌شود. ما از دو روش دیگر برای توسعه این کلمات کد استفاده کرده‌ایم تا کدهای استاندارد JPEG انطباق بهتری با کدر ما داشته باشد:

۱- در حالتی که تعداد صفرها بیشتر از ۱۵باشد از یک کد ۱۴ بیتی برای نشان دادن تعداد صفرها استفاده می‌کنیم. برای تشخیص این نوع کلمه کد خاص از دیگر کدها قبل از آن کد ۴ بیتی '1010' را که همیشه برای نشان دادن انتهای کدگذاری قطعه سیگنال استفاده می‌شود قرار می‌دهیم. از آنجا که طول قطعه سیگنال به همراه دیگر اطلاعات کناری به دیکدر ارسال می‌شود در صورتی که دیکدر کد مزبور را در جایی غیر از انتهای قطعه سیگنال ببیند متوجه وجود عدد ۱۴ بیتی تعداد صفرها بعد از این کد می‌شود. در باندهای بالا که با تعداد زیادی صفر روبرو هستیم این روش کدگذاری کمک بزرگی در فشرده‌سازی بیشتر می‌کند.

۲- گاهی اوقات اعدادی بزرگ در ضرایب خروجی تبدیل بسته موجکی ایجاد می‌شوند که خارج از ۱۱ خانواده اولیه اعداد JPEG قرار می‌گیرند. برای رفع این مشکل پیش کدهایی برای یک خانواده اعداد دیگر (خانواده ۱۲) به وجود آورده و به جدول پیش کدهای قبلی اضافه کرده‌ایم.

به ترتیب زیر اطلاعات کدر کامل می‌شوند. با افزودن یک سری اطلاعات کناری مثل طول قطعه فعلی سیگنال، کد هسته‌ای که استفاده شده است، بردار تعداد بیت‌های تخصیص داده شده به نمونه‌های ۱۹ باند و ... داده‌ها بسته‌بندی می‌شوند. این داده‌ها توسط دیکدر باز شده و استخراج می‌شوند و با انجام عکس کارهایی که در قسمت انکدر صورت گرفته است قطعه سیگنال باز سازی می‌شود.

۵- نتیجه گیری

جدول ۳ حداکثر نرخ بیت، نرخ بیت متوسط و SEGSNR[2] را برای چند فایل سیگنال صحبت زن و مرد نشان می‌دهد. این نتایج برای طول قطعات متفاوت به دست آمده است و طول قطعه در محاسبه SEGSNR برابر ۱۰ ms در نظر گرفته شده است. در صورتی که طول قطعه‌های کد شده کوچکتر شود به علت افزایش تأثیر اطلاعات کناری نرخ بیت تا حدودی زیاد می‌شود. در بدترین شرایط حداکثر نرخ بیت در اطراف ۳۲ kb/s است که مقدار مناسبی برای سیگنال با کیفیت بالا است. کیفیت سیگنال بازسازی شده تا حدی است که سیگنال بازسازی شده به راحتی از سیگنال اولیه تشخیص داده نمی‌شود.

به عنوان ادامه این تحقیقات پیش‌بینی می‌شود که استفاده از قطعات ایستان سیگنال ورودی با طول متغیر و نیز استفاده از مدل روان شنیداری که به صورت بهتری از خواص تن‌های صدا استفاده کند باعث فشرده‌سازی بهتری خواهد شد. چند نکته در مورد نتایج حاصل قابل ذکر است. برای استفاده‌های مخابراتی (مثل مکالمات تلفنی دو طرفه) استفاده از طول قطعات ۳۲ ms مناسب است در حالیکه برای ذخیره‌سازی و آرشیو می‌توان از قطعات بزرگتر و نرخ اطلاعاتی کمتر استفاده نمود.

معیار SEGSNR برای اینگونه کدرها که توان نویز کمتر کاهش یافته و کیفیت بهتر به علت استفاده از خواص روان-شنیداری حاصل شده است

1. Category

2. Segmental Signal to Noise Ratio

Archive of SID
www.SID.ir

معیار مناسبی نیست و فقط به عنوان مقایسه در اینجا به کار رفته است. استفاده از معیار MOS به راحتی امکان پذیر نیست ودر اینجا فقط به مقایسه نتایج حاصل از نظر کیفی با دیگر استانداردها مثل G722 توسط افرادی که با این استانداردها آشنائی داشته‌اند بسنده شده است. این مقایسه گویای این است که کیفیت صدای بازسازی شده در این روش بهتر یا همانند استاندارد G722 یا ویرایش جدید آن G726 است.

مراجع

[1] N. S. Jayant and P. Noll, *Digital Coding of Waveforms*, Prentice Hall, INC. Englewood Cliffs, NJ, 1984.

[2] N. Jayant, J. Johnston and R. Safranek, "Signal compression based on models of human perception," *Proc. of IEEE*, vol. 81, no. 10, pp. 1385-1422, Oct. 1993.

[3] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, 1992.

[4] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1999.

[5] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, 1993.

[6] P. Srinivasan and H. Jamieson, "High quality audio compression using an adaptive wavelet packet decomposition and psychoacoustic modeling," *IEEE Tran. Signal processing*, vol. 46, no. 4, pp. 1085-1093, Apr. 1998.

[7] E. Ambikairajah, A. G. Davis, and W. T. K. Wong, "Auditory masking & MPEG-1 audio compression," *Electronics & Communication Engineering Journal*, vol. 9, no. 4, pp. 165-175, Aug. 1997.

[8] ISO/IEC JTC 1/SC 29/WG 1, "Call for contributions- lossless compression of continuous-tone still pictures," *ISO Working Document ISO/IEC JTC1/SC29/WG1 N41*, Mar. 1995.

[9] D. Sinha and A. H. Tewfik, "Low bit rate transparent audio compression using adaptive wavelets," *IEEE Trans. Signal Processing*, vol. 41, no. 12, pp. 3463-3479, Dec. 1993.

[10] *Matlab 5.3 Wavelet Toolbox Users Guide*.

[11] K. Brandenburge and M. Bosi, "Overview of MPEG audio: current and future standards for low bit rate audio coding," *J. Audio Engineering Society*, vol. 45, no. 1-2, pp. 4-21, Jan./Feb. 1997.

[12] ISO/IEC Int'l Standard 11172-3, *Information Technology: Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to about 1.5 Mb/s- Part3: Audio*.

[13] Digitale Audioverarbeitung, WS 2000, 18.205

[14] D. Y. Pan, "A tutorial on MPEG/Audio compression," *IEEE Multimedia*, vol. 2, no. 2, pp. 60-74, Summer 1995.

سید طه مرتضوی در سال ۱۳۷۹ مدرک کارشناسی مهندسی برق خود را از دانشگاه صنعتی خواجه نصیر طوسی و در سال ۱۳۸۲ مدرک کارشناسی ارشد خود را در گرایش الکترونیک از دانشگاه شهید بهشتی دریافت نمود. زمینه های تحقیقاتی نامبرده شامل سیستم‌های سونار، پردازش سیگنالهای صوتی و الکترونیک قدرت می‌باشد. ایشان از سال ۱۳۸۰ تاکنون بعنوان کارشناس و کارشناس ارشد پژوهشکده جهاد در انجام پروژه‌های تحقیقاتی فعال است.

محمد حسن ساوجی تحصیلات خود را در مقطع کارشناسی و کارشناسی ارشد مهندسی برق در دانشگاه صنعتی شریف در سالهای ۱۳۵۱ و ۱۳۵۳ به اتمام رساند. وی در سال ۱۹۷۹ از انستیتو پلی تکنیک گرونوبل فرانسه مدرک دکتر–مهندسی خود را در رشته الکترونیک و مخابرات دریافت نمود و در سال ۱۹۸۱ دردانشگاه آکسفورد انگلستان دوره فوق دکترای خود را گذراند. ایشان بین سالهای ۱۹۸۱تا ۱۹۹۵ در دانشگاه‌ها و مراکز مختلف اروپائی به تدریس و تحقیق مشغول بود. دکتر ساوجی از سال ۱۳۸۰ تاکنون با سمت استاد الکترونیک و مخابرات در دانشگاه شهید بهشتی به فعالیت خود ادامه می‌دهد. برخی از زمینه های تحقیقاتی مورد علاقه ایشان عبارتند از: پردازش سیگنال، پردازش تصویر و ویدئو، پردازش سیگنال صحبت و فیلترهای وفقی و غیرخطی.

Archive of SID

www.SID.ir