Downloaded from jame.iut.ac.ir at 19:55 IRST on Tuesday March 2nd 2021

استفاده از روش FFI[1] و چندی کننده‌های برداری[2] در طراحی کد کننده‌های صحبت کیفیت بالا با نرخ BPS ۱۲۰۰

ابوالقاسم صیادیان*

دانشکدهٔ مهندسی برق، دانشگاه صنعتی امیرکبیر

(دریافت مقاله: ۱۳۷۶/۷/۲۶ - دریافت نسخهٔ‌نهایی: ۱۳۷۷/۴/۲۲)

چکیده - ذخیره‌سازی و یا ارسال سیگنال صحبت با کیفیت بالا با نرخهای بیت خیلی پایین یکی از مسائل تحقیقاتی مورد توجه برای مدمهای پیشرفته (که باید صحبت ، تصویر و داده‌ها را به طور همزمان در یک کانال KHZ ۴ ارسال کند)، سیستمهای پست صوتی و بی‌سیمهای HF[3] و ... است. در طی این تحقیق با استفاده از درونیابهای تصادفی بین فریمی[4] و همچنین چندی کننده‌های برداری درون فریمی[5]، قادر به کدکردن سیگنال صحبت با نرخ BPS ۱۲۰۰ و با کیفیت خیلی خوب شدیم. کیفیت سیگنال بازسازی شده قابل رقابت با کدکننده‌های BPS ۴۸۰۰ مطرح (مانند CELP) بوده است. البته کاهش نرخ به BPS ۱۲۰۰ همراه با افزایش تأخیر الگوریتم کدینگ از ms ۲۰ به ms ۴۰ و همچنین افزایش اندکی دربار محاسباتی حاصل شده است.

Using FFI Interpolator and VQ Quantization for Designing of High Quality 1200 BPS Speech Vocoder

A. Sayadian

Department of Electrical Engineering, Amirkabir University of Technology

ABSTRACT- ***Storaging or transmission of speech signals at very low bit rate is a hot area in the field of speech processing. We used stochastic inter-frame interpolators and vector quantization (VQ) as a new method for developing a high quality 1200 BPS speech vocoder. The objective and subjective test results show that performance of the new vocoder is compairable with 4800 BPS standard vocoders (as CELP).***

۱ - مقدمه

فشرده‌سازی سیگنال گفتار با نرخهای خیلی پایین که به صورت زمان واقعی قابل پیاده‌سازی باشند، یکی از زمینه‌های تحقیقاتی اساسی پردازش گفتار است. برای مثال در مدمهای نسل آینده، ارسال همزمان صوت و تصویر و داده‌ها (ارتباطات چند رسانه‌ای) در یک پهنای باند KHZ ۴ مد نظر است. اگرچه با پیشرفت

* استادیار

تکنولوژی، نرخ ارسال اطلاعات در این مدمها با بهره‌گیری از روشهایی مانند بهبود S/N کانال انتقال ، استفاده از روشهای حذف نویز و همشنوایی و همانسازها و همچنین استفاده از روشهای مدولاسیون چند سطحی تلفیقی (فاز، دامنه و فرکانس) و ... افزایش یافته است. با این وجود نرخ ارسال اطلاعات در یک پهنای باند ۴ KHZ محدود است. یک روش اساسی آن است که قبل از ارسال داده‌های خام مربوط به سیگنال گفتار یا تصویر، با روشهای مناسب افزونگیهای آنها را حذف کرده و سپس اطلاعات اصلی فشرده شده را ارسال کنیم. همچنین برای ذخیره‌سازی پیغامهای گفتاری افراد در سیستمهای پست الکترونیکی فعلی، سیگنال گفتار با نرخ ۶۴kbps ذخیره می‌شوند. در نتیجه اگر حجم پیغامها زیاد بوده و یا تعداد پیغامها بالا باشد، محدودیت نگهداری وذخیره‌سازی حجم زیاد داده‌های گفتاری به علت محدودیت حجم حافظهٔ دیسکها (و همچنین افزایش قیمت آنها)، یکی از مشکلات این سیستمهاست. همچنین در سیستمهای ارتباطی که از باند HF برای تبادل اطلاعات و ارسال پیغامهای گفتاری به صورت امن (رمزشده) استفاده می‌کنند، به علت مشکلات عدیدهٔ کانال HF، مجبور هستیم که از نرخهای پایین (حدود ۱۲۰۰BPS) استفاده کنیم. بنا به نیازهایی که بعضاً ذکر شد، ضرورت دارد که کدکننده‌های صحبت با نرخهای بیت خیلی پایین (کمتر یا مساوی ۱۲۰۰BPS) مورد توجه و تحقیق قرار گیرند. کدکننده‌هایی که در این نرخ بیت تاکنون طراحی شده‌اند صرفاً به قابل فهم بودن[6] پیغام ارسالی توجه داشته‌اند (زیرا نوعاً برای کاربردهای نظامی طراحی شده‌اند). آنچه در این تحقیق مورد توجه است، طراحی کدکننده‌های کیفیت بالا (هم از نظر فهم، هم از نظر کیفیت و تشخیص گوینده و هم از نظر وضوح) با نرخ ۱۲۰۰BPS است، به قسمی که برای کاربردهای ذکر شده قابل استفاده باشند.

در بخش دو، اصول کدکننده‌های پارامتریک صحبت برای وصول به نرخهای خیلی پایین بحث خواهد شد. در بخش سوم، روشهای درونیابی بین فریمی که برای پایین‌آوردن نرخ بیت مورد توجه است بیان خواهد شد. در بخش چهارم روشهای استفاده از چندی کنندهٔبرداری برای اهداف OBA[7] مورد مطالعه قرار خواهد گرفت. در بخش ۵ نتایج شبیه‌سازیهای انجام شده برای ارزیابی روش پیشنهادی توضیح داده خواهد شد.

۲- کدکننده‌های پارامتریک صحبت برای نرخهای خیلی پایین

روش اصولی برای وصول به نرخهای بیت خیلی پایین (کمتر از ۴۸۰۰BPS) استفاده از کدکننده‌های پارامتریک صحبت است (در مقابل کدکننده‌های شکل موج[8]) . بدین لحاظ در یک بازهٔ زمانی ۱۰تا ۲۰ میلی ثانیه که می‌توان سیگنال گفتار را ایستان فرض کرد، دودسته پارامتر استخراج می‌شود. پارامترهای دستهٔ اول نمایندهٔ دامنهٔ پوش طیف، و پارامترهای دسته دوم نمایندهٔ نوع و نحوهٔ اعمال سیگنال تحریک‌اند.

اطلاعات مربوط به فاز طیف درگیرنده، با فرض حداقل فاز بودن مدل سیستم تولید گفتار، به طور ضمنی (مانند روش LPC[9]) و یا به طور صریح (مانند مدل سینوسی یعنی روشهای MBE[10] و HV[11]) تولید می‌شوند. با توجه به اینکه هدف این تحقیق، طراحی کدکننده‌های صحبت با کیفیت بالاست (علاوه بر قابل فهم بودن)، لذا برای مدلسازی پارامتریک از روش تحلیل و سنتز MBE [۱-۷] استفاده کردیم. تحلیل و سنتز MBE در حوزهٔ فرکانس انجام می‌پذیرد. از ویژگیهای عمدهٔ تحلیل و سنتز پارامتریک MBE، مقاوم بودن در مقابل نویزهای محیطی، افزایش وضوح و طبیعی بودن صدا و بهبود کیفیت اطلاعات گوینده، علاوه بر حفظ فهم کلام است. مبانی مدلسازی پارامتریک MBE در مراجع متعدد از جمله [۱-۳] بیان شده است. خروجی تحلیل کننده MBE عبارت‌اند از : الف) گین فریم، فرکانس پیچ، یک الگوی ۱۲ بیتی (متناظر با باصدا بودن یا بی‌صدا بودن ۱۲ باند طیفی)، مجموعاً به عنوان پارامترهای تحریک ب) نمونه‌های دامنهٔ پوش طیف (۵۰ تا ۸۰ نمونه برای فرکانس نمونه‌برداری ۸ KHZ) و یا ۱۶ تا ۳۲ ضریب تبدیل DCT[12] یا ضریب کپسترم[13]، مجموعاً به عنوان پارامترهای نمایندهٔ پوش طیف. برای وصول به یک کیفیت قابل قبول، ضرورت دارد که فرکانس پیچ و گین فریم توسط ۸ بیت چندی شوند. (در مقیاس لگاریتمی بهتر است). همچنین برای ارسال اطلاعات V/UV باندهای طیفی از ۱۲ بیت استفاده کنند. بنابراین برای ارسال پارامترهای تحریک به ازای هر فریم به حدود ۲۸ بیت نیازمندیم. حال اگر طول هر فریم را برابر ۲۰ms در نظر بگیریم، ضرورت دارد که فرایند کدینگ ۵۰ بار در ثانیه انجام پذیرد. با ملاحظات بالا، مشاهده می‌شود که برای کدکردن صرفاً پارامترهای قسمت تحریک

Downloaded from jame.iut.ac.ir at 19:55 IRST on Tuesday March 2nd 2021

به حدود ۱۴۰۰BPS نیازمندیم. یعنی با روشهای معمولی و با محدودیت نرخ ۱۲۰۰BPS، حتی قادر نیستیم که اطلاعات تحریک را برای گیرنده ارسال کنیم. مضافاً اینکه بیشترین حجم بیت موردنیاز در کدکننده‌ها نوعاً مربوط به پارامترهای نمایندهٔ پوش طیف است (اگر از روش چندی کردن اسکالر استفاده کنیم.)

۳- استفاده از درونیابهای بین فریمی FFI برای کاهش نرخ ارسال FRPS[۱۴]

همان طوری که در بخش قبل بیان شد، با کدکردن پارامترهای پوش طیف و سیگنال تحریک به ازای هر ۲۰ms، امکان وصول به نرخ ۱۲۰۰BPS وجود ندارد. برای فائق آمدن بر این مشکل، نوعاً پارامترهای دو فریم متوالی را با همدیگر به صورت مشترک کد می‌کنند. یعنی اگر فرض کنیم $X_1(n)$ بردار نمایندهٔ پارامترهای فریم nام و $X_2(n)$ بردار نمایندهٔ پارامترهای فریم n+۱ام باشند، در این صورت بردار $X(n)= [X_1(n), X_2(n)]$ را به صورت مشترک و به طریق مناسب کد کرده و ارسال می‌کند. روش کلاسیک برای کد کردن بردار تعمیم یافتهٔ X(n)، استفاده از چندی کننده‌های ماتریسی است. وقتی که پارامترهای داخل بردارها دارای ارزش یکسان بوده و همچنین دارای خواص چندی شدن مناسب باشند، استفاده از روش بالا مناسب خواهد بود. متأسفانه بعضی از پارامترها (به ویژه پترن ۱۲ بیتی V/UV سیگنال تحریک) قابل چندی شدن به روش معمول نیستند. علاوه بر آن حجم داده‌های آموزشی مورد نیاز برای طراحی کتاب کد ماتریسی فوق‌العاده بالاست. روشی که ما در طی این تحقیق مورد توجه قرار دادیم استفاده از درونیابهای بین فریمی به‌شرح زیر است. فرض می‌کنیم X(n) و X(n+۱) و X(n+۲) بردار پارامترهای مربوط به سه فریم متوالی هر کدام به طول ۲۰ms باشند. در روش درونیابی بین فریمی FFI، پارامترهای مربوط به X(n+۱) ارسال نمی‌شود. گیرنده با استفاده از پارامترهای بردارهای X(n) و X(n+۲) و ضرایب درونیابی (که متعاقباً نحوهٔ تخمین آن بیان خواهد شد)، پارامترهای بردار حذف شدهٔ X(n+۱) را بازسازی می‌کنند. با این روش ما عملاً به ازای هر ۴۰ms یکبار پارامترها را در فرستنده کد کرده و برای گیرنده ارسال می‌کنیم (به علاوه تعدادی بیت برای ارسال ضرایب درونیاب). بنابراین به ازای هر فریم ۴۰ms برای کدکننده با نرخ ۱۲۰۰BPS تعداد ۴۸ بیت در اختیار خواهیم داشت ($۴۸ = \frac{۱۲۰۰}{۲۵}$).

۳-۱ تخمین ضرایب درونیاب تصادفی

فرض می‌کنیم X(n) و X(n+۱) و X(n+۲) سه بردار متناظر با پارامترهای طیفی سه فریم متوالی (هر فریم به طول ۲۰ms) باشند. همان طوری که در ابتدای بخش بیان شد، در فرستنده پارامترهای مربوط به فریم X(n) و X(n+۲) کد شده و ارسال می‌شود، ولی از ارسال پارامترهای X(n+۱) صرف نظر می‌شود. پارامترهای مربوط به فریم X(n+۱) باید از روی پارامترهای فریم X(n) و X(n+۲) به نحو مقتضی بازسازی شود. ساده‌ترین روش بازسازی X(n+۱) آن است که آن را به عنوان میانگین X(n) و X(n+۲) در نظر بگیریم. یعنی پارامترهای فریم X(n+۱) را به صورت $X(n+۱)= ۰/۵\, X(n) + ۰/۵\, X(n+۲)$ درگیرنده تولید کنیم. با این روش به ارسال اطلاعات اضافی برای تولید پارامترهای X(n+۱) نیازی نخواهیم داشت. متأسفانه این روش درونیابی برای فواصل زیر ۲۰ms تا حدودی قابل قبول بوده ولی برای فواصل ۴۰ms کیفیت صدای تولید شده را خراب می‌کند. مرجع [۱۰] برای درونیابی پارامترهای X(n+۱) از روی X(n) و X(n+۲) روشی به شرح زیر پیشنهاد کرده است:

$$S_۱ : \hat{X}(n+۱) = X(n)$$

$$S_۲ : \hat{X}(n+۱) = X(n+۲)$$

$$S_۳ : \hat{X}(n+۱) = ۰/۵\, X(n) + ۰/۵\, X(n+۲)$$

$$S_۴ : \hat{X}(n+۱) = ۰/۲۵\, X(n) + ۰/۷۵\, X(n+۲)$$

در فرستنده خطای MSE[۱۵] بین هر یک از حالات S_i (i = ۱ تا ۴) و X(n+۱) را به دست می‌آورند. آن گاه حالتی که کمترین خطای MSF را ایجاد کند، انتخاب و شمارهٔ آن را به عنوان اطلاعات مربوط به ضرایب درونیاب برای گیرنده ارسال می‌کند. روش بالا در طی این تحقیق پیاده‌سازی شد. نتایج تجربی حاصل نشان می‌دهد که برای تولید صحبت قابل فهم می‌توان از روش بالا استفاده کرد ولی کیفیت آن بالا نیست (به ویژه در نواحی گذرا و

Downloaded from jame.iut.ac.ir at 19:55 IRST on Tuesday March 2nd 2021

Downloaded from jame.iut.ac.ir at 19:55 IRST on Tuesday March 2nd 2021

همچنین در واجهای ب، پ، ت، ج، گ، خ، د و ....).
برای حل مشکل ذکر شده از نظریه تخمین خطی به شرح زیر استفاده کردیم. فرض می‌کنیم $X(n) = [x_1(n), \dots, x_p(n)]$ و $X(n+1)=[x_1(n+1), \dots, x_p(n+1)]$ و $X(n+2) = [x_1(n+2), \dots, x_p(n+2)]$ بردار پارامترهای سه فریم متوالی (هر فریم به طول ۲۰ms) باشند. p تعداد پارامترهای طیفی است. حال می‌خواهیم $X_i(n+1)$ ها را بر حسب $X_i(n)$ ها و $X_i(n+2)$ ها (به عنوان داده‌ها) به شرح زیر تخمین خطی بزنیم:

$$\hat{X}_i(n+1) = \alpha_1 . X_i(n) + \alpha_2 . X_i(n+2) \quad ; \quad i = 1 \text{ تا } p \tag{2}$$

$$e_i(n+1) = X_i(n+1) - \hat{X}_i(n+1) \quad ; \quad i = 1 \text{ تا } p \tag{3}$$

در معادلهٔ (۲)، $\hat{X}_i(n+1)$ تخمین خطی $X_i(n+1)$ بر حسب دو مقدار $X_i(n)$ و $X_i(n+2)$ است. در معادلهٔ (۳)، $e_i(n+1)$ خطای تخمین خطی $X_i(n+1)$ برحسب $X_i(n)$ و $X_1(n+2)$ است. با جایگزینی $\hat{X}_i(n+1)$ در معادلهٔ (۳)، خطای تخمین برابر مقدار زیر خواهد شد:

$$e_i(n+1) = X_i(n+1) - \alpha_1 . X_i(n) - \alpha_2 . X_i(n+2) \tag{4}$$

حال از اصل تعامد خطا برداده‌ها (در نظریه تخمین خطی) به شرح زیر استفاده می‌کنیم:

$$E\{e_i(n+1). X_i(n)\} = 0 \tag{5}$$

$$E\{e_i(n+1). X_i(n+2)\} = 0 \tag{6}$$

$E\{...\}$ نمایندهٔ متوسط‌گیری آماری[16] بر حسب اندیس پایین i ($i = 1$ تا p) است. با جایگزینی معادلهٔ (۴) در معادله‌های (۵) و (۶) و پس از مقداری عملیات ریاضی خواهیم داشت:

$$\alpha_1 . E\{X_i^2(n)\} + \alpha_2 . E\{X_i(n) . X_i(n+2)\} = E\{X_i(n) . X_i(n+1)\} \tag{7}$$

$$\alpha_1 . E\{X_i(n) . X_i(n+2)\} + \alpha_2 . E\{X_i^2(n+2)\} = E\{X_i(n+2) . X_i(n+1)\} \tag{8}$$

مقادیر $E\{X^2_i(n)\}$ و $E\{X^2_i(n+2)\}$ و $E\{X_i(n). X_i(n+2)\}$ و $E\{X_i(n) . X_i(n+1)\}$ و $E\{X_i(n+2) . X_i(n+1)\}$ از روی پارامترهای بردارهای $X(n)$ و $X(n+1)$ و $X(n+2)$ در فرستنده تعیین می‌شوند (در نتیجه معلوم فرض می‌شوند). با استفاده از معادله‌های (۷) و (۸) (دو معادله و دو مجهول)، مقادیر ضرایب درونیاب خطی یعنی (α_1 و α_2) را به دست می‌آوریم.

۳-۲ درونیاب تصادفی مضاعف

روش درونیابی FFI مرجع [۱۰] آزادی عمل مدل تغییرات زمانی پارامترها از فریم nام به n + ۲ام را بسیار محدود می‌کند (صرفاً توسط ۴ حالت قطعی). با استفاده از روش درونیابی تصادفی که در این بخش بیان شد، قادر به حذف محدودیت سرتاسری[17] ضرایب درونیاب شدیم. یعنی کلیه پارامترهای $X(n+1)$ به صورت یکسان و با آزادی عمل در حد دو پارامتر (α_1 و α_2) قادر به بازسازی از روی پارامترهای $X(n)$ و $X(n+2)$ هستند. وقتی که کانتور کلیه پارامترها، ضمن عبور از فریم n ام به n + ۲ ام دارای تغییرات یکسان باشند، روش درونیاب تصادفی تکی که در این بخش معرفی شد، مناسب خواهد بود. البته برای نواحی ایستان و همچنین برای طول فریمهای زمانی کمتر از ۲۰ms روش درونیابی تصادفی تکی قابل قبول است. ولی برای طول فریمهای طولانی (۴۰ms)، مشاهدات تجربی نشان می‌دهد که ایجاد آزادی عمل برای پارامترهای مختلف به نتایج مناسبتری منجر می‌شود. مطلب بالا این مسئله را بیان می‌کند که اگر برای هر جفت پارامتر طیفی، یک جفت ضریب درونیابی طراحی شود، به وضعیت ایده‌آل خواهیم رسید. متأسفانه استفاده از ایدهٔ بالا موجب افزایش غیرقابل قبول تعداد ضرایب درونیابی می‌شود، که در عمل نیز ضرورتی به آن احساس نشد. با توجه به نتایج شبیه‌سازیهای انجام شده در طی این تحقیق، ملاحظه کردیم که انتخاب دو جفت ضرایب درونیابی برای تولید سیگنال کیفیت بالا کاملاً مناسب است

(برای فواصل ۴۰ms). بدین منظور یک جفت ضریب درونیابی (α_1 و α_2) برای پارامترهای $\frac{p}{3}$ تا $i = 1$ و گین فریم و همچنین یک جفت ضریب درونیابی دیگر (β_1 و β_2) برای پارامترهای طیفی p تا $i = \frac{p}{3} + 1$ طراحی کردیم. روش بالا را درونیابی تصادفی مضاعف می‌نامیم.

۴- استفاده از چندی‌کننده‌های برداری ضربی[18] PVQ برای کاهش نرخ بیت

در بخشهای قبلی نشان دادیم که به ازای هر ۴۰ms مجموعه پارامترهای زیر باید استخراج شده و برای گیرنده ارسال شوند: ۱) گین فریم ۲) فرکانس پیچ ۳) یک پترن ۱۲ بیتی شاخص V/UV بودن باندهای متمایز طیفی ۴) پارامترهای پوش طیف ۵) ۲ تا ۴ ضریب درونیاب تصادفی بین فریمی.

همان طوری که قبلاً بیان شد، ارسال این تعداد پارامتر با ۴۸ بیت با استفاده از چندی‌کننده‌های اسکالر به قسمی که سیگنال بازسازی شده دارای کیفیت بالا باشد، تقریباً ناممکن است. برای افزایش دقت و کیفیت نهایی کدکننده، ضرورت دارد که فرکانس پیچ و گین توسط ۸ بیت (در مقیاس لگاریتمی) کد شوند. در این صورت برای ارسال سایر پارامترها صرفاً باید از ۳۲ بیت استفاده کنیم. (که ۱۲ بیت آن برای ارسال پترن V/UV مصرف می‌شود). راه حل اساسی برای ارسال این تعداد از پارامترها (با تعداد بیت کم)، استفاده از گونه‌های مختلف چندی کننده‌های برداری VQ است [۴-۷]. فرض می‌کنیم $\alpha_1(n)$, $\alpha_2(n)$ ... $\alpha_4(n)$] و $Y(n) = [X_1(n),... , X_p(n)]$ بردار تعمیم‌یافته مربوط به فریم nام باشد. این بردار از مجموعه پارامترهای پوش طیف و ضرایب درونیاب تشکیل شده است. ابتدا توسط بردارهای آموزشی مناسب یک کتاب کد اصلی ۱۲ بیتی (با تعداد ۴۱۹۶ = N_1 بردار مرجع) طراحی می‌کنیم [۱۱-۱۴].

فرض می‌کنیم $C = (C_1 , ... , C_{N1})$ مجموعه بردارهای مرجع کتاب کد اصلی باشند. مجموعه بردارهای آموزشی باقیمانده را بصورت $Z(n) = Y(n) - C_k$ به دست می‌آوریم که C_k نزدیکترین بردار مرجع به $Y(n)$ از میان کلیه بردارهای مرجع واقع در کتاب کد اصلی C است. مجدداً با استفاده از بردارهای آموزشی باقیمانده $Z(n)$ ها، یک کتاب کد ۱۲ بیتی دیگر طراحی می‌کنیم. اسم کتاب کد اخیر را، کتاب کد فرعی می‌نامیم. کلیه بردارهای آزمون باقیمانده را با کتاب کد فرعی چندی می‌کنیم. بنابراین برای هر بردار آزمون تعمیم‌یافته ورودی در مرحلهٔ کدینگ، ۲ عدد ۱۲ بیتی خواهیم داشت. عدد اول شاخص آدرس نزدیکترین بردار مرجع در کتاب کد اصلی و دومی شاخص آدرس نزدیکترین بردار مرجع درکتاب کد فرعی خواهد بود. پس از طراحی کتابهای کد اصلی و فرعی، به هر بردار مرجع واقع در کتاب کد اصلی یک الگوی ۱۲ بیتی (که نمایندهٔ وضعیت V/UV باندهای طیفی متناظر با آن بردار است) نسبت می‌دهیم. اگر چند الگوی ۱۲ بیتی V/UV برای هر بردار مرجع داشته باشیم، الگویی که فریم متناظر با آن دارای بیشترین بهره باشد را در نظر می‌گیریم. در مرحله کدینگ، تغییرات الگوی ۱۲ بیتی V/UV بردار ورودی با بردار مرجع اصلی متناظر را با ۶ بیت کد کرده و ارسال می‌کنیم. در نتیجه با استفاده از این روش، هر بردار تعمیم یافته ورودی به همراه با الگوی V/UV ذیربط، توسط ۳۰ بیت کد می‌شود ۳۰ = (۶ + ۱۲ + ۱۲). دو بیت باقیمانده (از ۳۲ بیت) برای اهداف کنترلی مورد استفاده قرار می‌گیرند. بنابراین در مجموع با ۴۸ بیت کلیه پارامترهای مربوط به هر فاصلهٔ زمانی ۴۰ms را کد کرده و برای گیرنده ارسال می‌کنیم. بدین‌طریق قادر به طراحی یک کدکنندهٔ کیفیت بالا با نرخ ۱۲۰۰BPS شدیم.

۵- نتایج شبیه‌سازی

کلیه نتایج کمی و کیفی حاصل در طی این تحقیق با استفاده از داده‌های گفتاری مربوط به حدود ۱۰ دقیقه صحبت طبیعی ۱۰ نفر مرد و ۱۰ نفر زن و ۱۰ نفر بچه است. آزمون اول برای انتخاب و بررسی کارایی سه روش درونیاب ثابت، تصادفی تکی، تصادفی مضاعف، انجام پذیرفته است. این ارزیابی به صورت کمی و با استفاده از پارامتر SNR[19] بر حسب dB با معادلهٔ زیر محاسبه شده است.

$$SNR = \frac{1}{L} \sum_{n=1}^{L} 10 \log \frac{||X(n+1)||}{||X(n+1) - \hat{X}(n+1)||} \quad (9)$$

در معادلهٔ (۹)، L تعداد بردارهای آزمون برای انجام آزمون

Downloaded from jame.iut.ac.ir at 19:55 IRST on Tuesday March 2nd 2021

جدول ۱- نتایج SNR برای سه نوع درونیاب FFI

نوع درونیاب	ثابت	تصادفی تک	تصادفی مضاعف
مقدار SNR برحسب dB	۲۵/۷۴	۲۹/۶۵	۳۱/۸۶

است.) | | ... | | برابر انرژی بردار متناظر است. جدول (۱) مقادیر SNR (بر حسب dB را برای سه نوع درونیاب الف) ثابت ب) تصادفی ج) تصادفی مضاعف را نشان می‌دهد.

نتایج حاصل کارایی انکارناپذیر درونیاب تصادفی نسبت به درونیاب ثابت را نشان می‌دهد (افزایش کیفیتی در حدود dB ۶-۴). همچنین مشاهده می‌شود که عملکرد درونیاب مضاعف حدود ۲dB بهتر از درونیاب تصادفی تکی است. آزمون دوم برای ارزیابی کیفیت نهایی کدکنندهٔ ۱۲۰۰BPS طراحی شده به عمل آمده است. در این آزمون از DRT[۲۰] [۱۵] برای ارزیابی کیفی کدکننده استفاده کردیم. همچنین به منظور مقایسهٔ نسبی، آزمون بالا برای کدکنندهٔ CELP[۲۱] [۸ و ۹] با نرخ ۴۸۰۰BPS نیز به عمل آمده و نتایج حاصل در جدول (۲) درج شده است.

با ملاحظهٔ نتایج جدول (۲)، قابلیت کدکنندهٔ ۱۲۰۰BPS به روشنی مشخص می‌شود. زیرا با اینکه نرخ کدکنندهٔ CELP حدود ۴ برابر نرخ کدکنندهٔ FFI-MBE طراحی شده است، ولی کیفیت آنها تقریباً برابرند. تأخیر کد کنندهٔ FFI-MBE حدود دو برابر کدکنندهٔ

جدول ۲- نتایج آزمون DRT برای کدکنندهٔ طراحی شده و CELP

نوع کدکننده	کدکنندهٔ CELP ۴۸۰۰BPS	کدکنندهٔ FFI-MBE ۱۲۰۰BPS
نتایج آزمون DRT (برحسب درصد)	۹۳/۲۵	۹۲/۸۵

CELP بوده و در حدود ۱۲۰ms است (مجموع تأخیر مورد نیاز برای تحلیل و سنتز). کدکنندهٔ بالا بر روی رایانه پنتیوم (۲۰۰MHZ) به صورت زمان واقعی قابل پیاده‌سازی است.

۶- جمعبندی

در طی این تحقیق کدکنندهٔ کیفیت بالای FFI-MBE با نرخ ۱۲۰۰ BPS طراحی و پیاده‌سازی شد. اساس این کدکننده مبتنی بر روش تحلیل و سنتز با تحریک چند باندهٔ MBE است. عناصر اصلی برای وصول به نرخ پایین، استفادهٔ همزمان از درونیابهای تصادفی مضاعف و کدکننده‌های برداری درون فریمی PVQ بوده است. کیفیت کدکنندهٔ طراحی شدهٔ FFI-MBE در نرخ ۱۲۰۰ BPS با کیفت کدکنندهٔ استاندارد شدهٔ CELP با نرخ ۴۸۰۰BPS برابری می‌کند. تأخیر مورد نیاز کل الگوریتم (تحلیل و سنتز) حدود ۱۲۰ms است.

واژه نامه

1. frame fill interpolation
2. vector quanitization
3. high frequency
4. stochastic inter frame interpolation
5. intera frame VQ
6. perceptual inteligibility
7. optimum bit allocation
8. waveform vocoder
9. linear predictive coding
10. multi band excitation
11. harmonic vocoder
12. discrete cosine transform
13. cepstrum
14. frame rate per second
15. mean square error
16. expectation
17. global
18. product VQ
19. proportional signal to noise ratio
20. diagnostic rytheme test
21. code excited linear prediction

مراجع

1. Griffin, D. W., and Lim., J. S., "Multi-Band Excitation Vocoder," *IEEE Trans. on ASSP* , Vol . 36, Aug. 1988.
2. Hardwick, J. C., and Lim, J. S., "A 4.8 kbps Multi-Band Excitation Speech Coder, " *ICASSP,* pp. 374-377, Apr. 1988.
3. Hardwick, J. C., and Lim, J. S.,"A 4800 BPS Improved Multi-Band Excitation (IMBE) Speech coder," *IEEE Speech Coding Workshop*, Vancouver, B. C. Canada, Sep., 1989 .

Downloaded from jame.iut.ac.ir at 19:55 IRST on Tuesday March 2nd 2021

4. Meuse, P. C., "A 2400 BPS Multi-Band Excitation Vocoder, "*ICASSP* , pp. s1. 3-9, 1990.
5. Digital Voice System , "Inmarsat-M Voice Codec, Version.2, "Inmarsat-M Specification, Inmarsat, Feb. 1991.
6. Nishiquchi, M., Mastsumoto, J., Wakatsuki, R., and Ono, S., "Vector Quantizer MBE With Simplified V/UV Decision at 3. kbps, " *ICASSP*, pp. 151-159, 1993.
7. Das, A., Rao, A. V., and Gersho, A., "Variable Dimension Vector Quantization of Speech Spectra For Low Rate Vocoders, " *ICASSP*, pp. 420-430, 1994.
8. Kemp, D. P., and Sueda, R. A., and Tremain, T. E., "An Evaluation of 4800 BPS Coders," *Proc . of the Military and Government Speech Tech.,* 89, pp. 86-90, Arlington, VA, Nov. 13-15, 1989.
9. Campbell, J. P., Welsh , V. C., and Tremain, T. E., "The New 4800 BPS Coding Standard," *Proc. of the Military and Government Speech Tech* 89, pp. 64-70, Arlington, VA, Nov. 13-15, 1989.
10. Mcaulay, R. J., and Quatieri, T. F., "Low-Rate Speech Coding Based on the Sinusoidal Model," Advanced in Speech Signal Processing, S. Furui, Marcel Dekker, Inc., 1992.
11. Linde, Y., Bazo, A., and Gray, R. M., "An Algorithm For Vector Quantization Design, "*IEEE Trans.* Com-28, No. 1, pp. 84-95, 1980.
12. Gray, R. M., "Vector Quantization, " *IEEE ASSP Magazine* , Vol. 1, No. 2, pp. 4-29, April 1981.
13. Makhoul, J., Roucos, S., and Gish, H., "Vector Quantization in Speech Coding, " *Proc. of IEEE,* Vol. 73, No. 11, pp. 1551-1588, Nov. 1985.
14. Wang, Z., and Hanson, J. V., "Codebook Optimiza -tion by Couchy Annealing, Signal Processing Tech. and Applic.," *IEEE Technical Activities,* Board, 1995.
15. Voiers, W. D., "Diagnostic Acceptability Measure for Speech Communication Systems, "*IEEE Proc. on ICASSP,* pp. 204-207, 1977.

Downloaded from jame.iut.ac.ir at 19:55 IRST on Tuesday March 2nd 2021