



دانشگاه زنجان

دانشکده مهندسی

گروه برق

پایان نامه کارشناسی

گرایش: مخابرات

عنوان:

روش های کد کردن صحبت Voice coder

استاد راهنما: دکتر مصطفوی

نگارش: حسام یوسفی

# پایان نامه کارشناسی

تقدیم به

پدر و مادر

عزیزه مهربانم به پاس محبت‌های بیدری‌بخشان.

## چکیده

کدک صحبت ، یک کدک کم تاخیر است که صحبت با کیفیت عالی را در نرخ بیت ۱۶ kbps ارائه می دهد و برای شبکه های تلفن ماهواره ای و اینترنت و موبایل که به تاخیر زیاد حساس هستند ، مناسب است. در این رساله به اینکدر و دیکدر G.۷۲۸ بصورت دوطرفه کامل ( Full Duplex ) می پردازیم .

## کلمات کلیدی

کدینگ و فشرده سازی صحبت ، پیاده سازی بلادرنگ ،

## فهرست

### ۱- مقدمه

### فصل ۱: بررسی و مدل سازی سیگنال صحبت

۳..... معرفی سیگنال صحبت

۷..... ۱-۲ مدل سازی پیشگویی خطی

۸..... ۱-۲-۱ پنجره کردن سیگنال صحبت

۱۰..... ۱-۲-۲ پیش تاکید سیگنال صحبت

۱۱..... ۱-۲-۳ تخمین پارامترهای LPC

### فصل ۲: روش ها و استانداردهای کدینگ صحبت

۱۴..... مقدمه

۱۵..... ۱-۲ روش های کدینگ

۱۷..... ۱-۲-۱ کدرهای شکل موج

۱۸..... ۱-۲-۲ کدرهای صوتی

۱۹..... ۱-۲-۳ کدرهای مختلط

۱۹..... الف- کدرهای مختلط حوزه فرکانس

۲۳..... ب- کدرهای مختلط حوزه زمان

### فصل ۳: کدر کم تاخیر LD-CELP

۳۰..... ۱-۳ مقدمه

۳-۲-۲- بررسی کدر کم تاخیر LD-CELP..... ۳۱

۳-۲-۱- LPC معکوس مرتبه بالا..... ۳۴

۳-۲-۲- فیلتر وزنی شنیداری..... ۳۷

۳-۲-۳- ساختار کتاب کد..... ۳۸

۳-۲-۳-۱- جستجوی کتاب کد..... ۳۹

۳-۲-۴- شبه دیکدر..... ۴۰

۳-۲-۵- پست فیلتر..... ۴۱

مراجع..... ۴۴

## مقدمه -

امروزه در عصر ارتباطات و گسترش روزافزون استفاده از شبکه های تلفن، موبایل و اینترنت در

جهان و محدودیت پهنای باند در شبکه های مخابراتی، کدینگ و فشرده سازی صحبت امری

اجتناب ناپذیر است. در چند دهه اخیر روشهای کدینگ مختلفی پدید آمده اند ولی بهترین و

پرکاربردترین آنها کدک های آنالیزاسنتز هستند که توسط Atal & Remede در سال ۱۹۸۲

معرفی شدند [۲]. اخیراً مناسبترین الگوریتم برای کدینگ صحبت با کیفیت خوب در نرخ بیت

های پائین و زیر ۱۶ kbps، روش پیشگویی خطی باتحریک کد (CELP) می باشد که در سال

۱۹۸۵ توسط Schroeder & Atal معرفی شد [۸] و تا کنون چندین استاندارد مهم کدینگ صحبت

بر اساس CELP تعریف شده اند.

در سال ۱۹۸۸ CCITT برنامه ای برای استانداردسازی یک کدک ۱۶ kbps با تاخیراندک و

کیفیت بالا در برابر خطاهای کانال آغاز نمود و برای آن کاربردهای زیادی همچون شبکه PSTN،

ISDN، تلفن تصویری و غیره در نظر گرفت. این کدک در سال ۱۹۹۲ توسط Chen et al. تحت

عنوان LD-CELP معرفی شد [۶] و بصورت استاندارد G.۷۲۸ در آمد [۹] و در سال ۱۹۹۴ بان

مشخصات متمیز ثابت این کدک توسط ITU ارائه شد [۱۰]. با توجه به کیفیت بالای این کدک که

در آن صحبت سنتز شده از صحبت اولیه تقریباً غیرقابل تشخیص است و کاربردهای آن در شبکه

های تلفن و اینترنت و ماهواره ای در این گزارش به پیاده سازی این کدک می پردازیم.

در فصل اول به معرفی و آنالیز سیگنال صحبت پرداخته می شود و در فصل دوم روش ها و

استانداردهای کدینگ بیان می شوند. در فصل سوم کدک LD-CELP را بیشتر بررسی می کنیم.



# فصل ۱

## بررسی و مدل سازی سیگنال صحبت

### ۱-۱- معرفی سیگنال صحبت

صحبت در اثر دمیدن هوا از ریه ها به سمت حنجره و فضای دهان تولید می شود. در

طول این مسیر در انتهای حنجره، تارهای صوتی<sup>۱</sup> قرار دارند. فضای دهان را از بعد از

تارهای صوتی، لوله صوتی<sup>۲</sup> می نامند که در یک مرد متوسط حدود ۱۷cm طول دارد. در

تولید برخی اصوات تارهای صوتی کاملاً باز هستند و مانعی بر سر راه عبور هوا ایجاد

نمی کنند که این اصوات را اصطلاحاً اصوات بی واك<sup>۳</sup> می نامند. در دسته دیگر اصوات،

تارهای صوتی مانع خروج طبیعی هوا از حنجره می گردند که این باعث به ارتعاش در آمدن

تارها شده و هوا به طور غیر یکنواخت و تقریباً پالس شکل وارد فضای دهان می شود. این

دسته از اصوات را اصطلاحاً باواك<sup>۴</sup> می گویند.

فرکانس ارتعاش تارهای صوتی در اصوات باواك را فرکانس Pitch و دوره تناوب

ارتعاش تارهای صوتی را پریود Pitch می نامند. هنگام انتشار امواج هوا در لوله صوتی،

<sup>۱</sup> Vocal Cords

<sup>۲</sup> Vocal Tracts

<sup>۳</sup> Unvoiced

<sup>۴</sup> Voiced

طیف فرکانس این امواج توسط لوله صوتی شکل می‌گیرد و بسته به شکل لوله، پدیده تشدید در فرکانس‌های خاصی رخ می‌دهد که به این فرکانس‌های تشدید فرمنت<sup>۵</sup> می‌گویند.

از آنجا که شکل لوله صوتی برای تولید اصوات مختلف، متفاوت است پس فرمنت‌ها برای اصوات گوناگون با هم فرق می‌کنند. با توجه به اینکه صحبت یک فرآیند متغییر با زمان است پس پارامترهای تعریف شده فوق اعم از فرمنت‌ها و پریود Pitch در طول زمان تغییر می‌کنند به علاوه مد صحبت به طور نامنظمی از باواک<sup>۶</sup> به بی واک<sup>۷</sup> و بالعکس تغییر می‌کند. لوله صوتی، همبستگی‌های زمان-کوتاه، در حدود ۱ ms، درون سیگنال صحبت را در بر می‌گیرد. و

بخش مهمی از کار کدکننده‌های صوتی مدل کردن لوله صوتی به صورت یک فیلتر زمان-

کوتاه می‌باشد. همان‌طور که شکل لوله صوتی نسبتاً آهسته تغییر می‌کند، تابع انتقال این فیلتر مدل‌کننده هم نیاز به تجدید<sup>۸</sup>، معمولاً در هر ۲۰ms یکبار خواهد داشت.

در شکل (۱-۱ الف) یک قطعه صحبت باواک که با فرکانس ۸KHz نمونه برداری شده است دیده می‌شود. اصوات باواک دارای تناوب زمان بلند به خاطر پریود Pitch هستند که نوعاً بین ۲ms تا ۲۰ms می‌باشد. در اینجا پریود Pitch در حدود ۸ms یا ۶۴ نمونه است. چگالی طیف توان این قطعه از صحبت در شکل (۱-۱ ب) دیده می‌شود [۳].

اصوات بی واک نتیجه تحریک نويز مانند لوله صوتی هستند و تناوب زمان-بلند اندکی را در بر دارند، همانگونه که در شکل‌های (۱-۱ ج) و (۱-۱ د) دیده می‌شود و لی همبستگی زمان

کوتاه به خاطر لوله صوتی در آنها هنوز وجود دارد.

بطور کلی سیگنال صحبت دارای افزونگی<sup>۹</sup> زیادی است که ناشی از عوامل ذیل هستند:

- وابستگی‌های زمان-کوتاه: این وابستگی‌ها عمدتاً به کندی تغییرات صحبت با زمان و

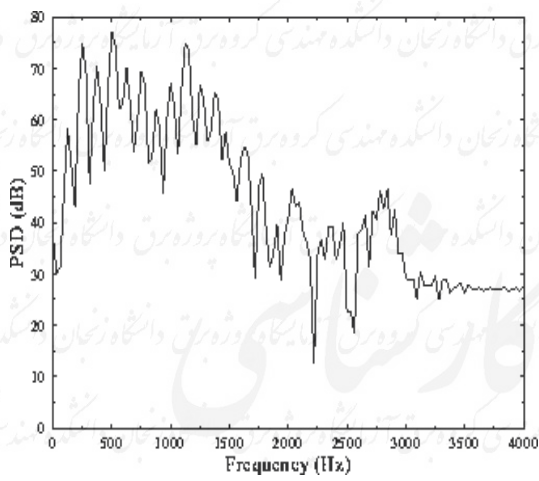
## ساختار

<sup>۵</sup>Formant

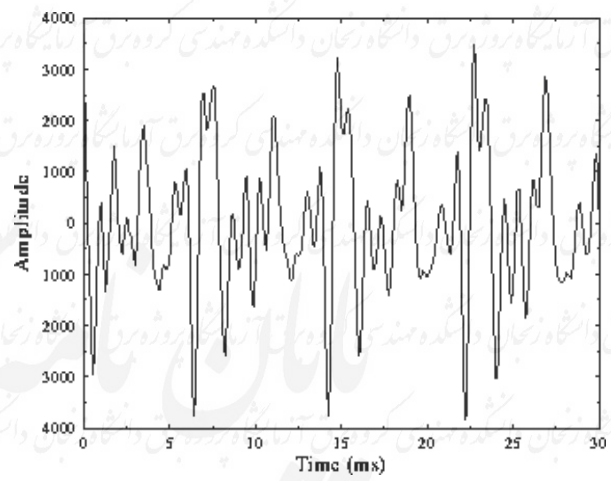
<sup>۶</sup>Update

<sup>۷</sup>Redundancy

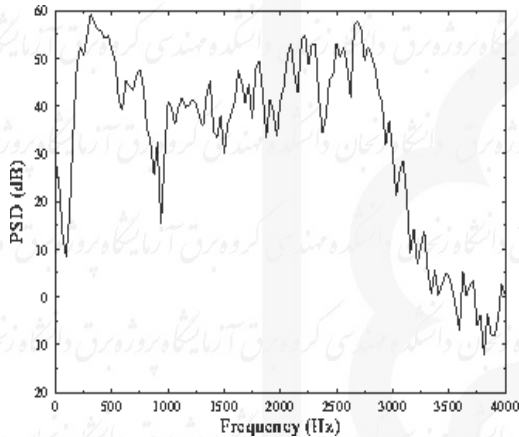




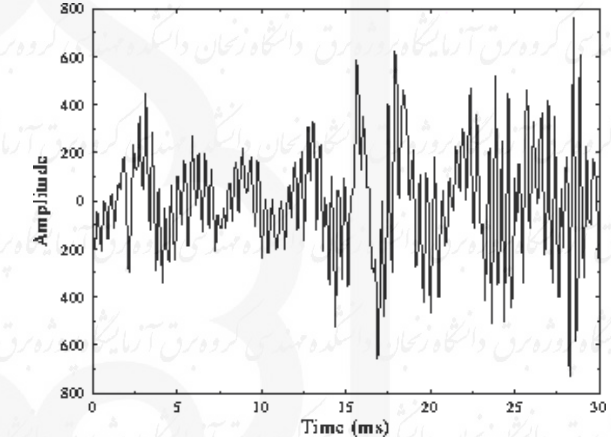
(ا)



(ب)



(ج)



(د)

شکل (۱-۱) : مقایسه اصوات باواک و بی واك. (الف) و (ب) : باواك ، (ج) و (د) : بی واك.

نسبتاً منظم فرمنت ها مربوط می شوند.

- وابستگی های زمان- بلند : که عمدتاً از طبیعت نیمه متناوب اصوات با واك و تغییرات آرام

پریود Pitch ناشی می شوند.

تابع چگالی احتمال صحبت : علیرغم پیچیدگی آماری صحبت می‌توان آن را با توابع چگالی

احتمال شناخته شده تقریب زد. شکل لوله صوتی و مد تحریک آن به صورت نسبتاً آرام تغییر

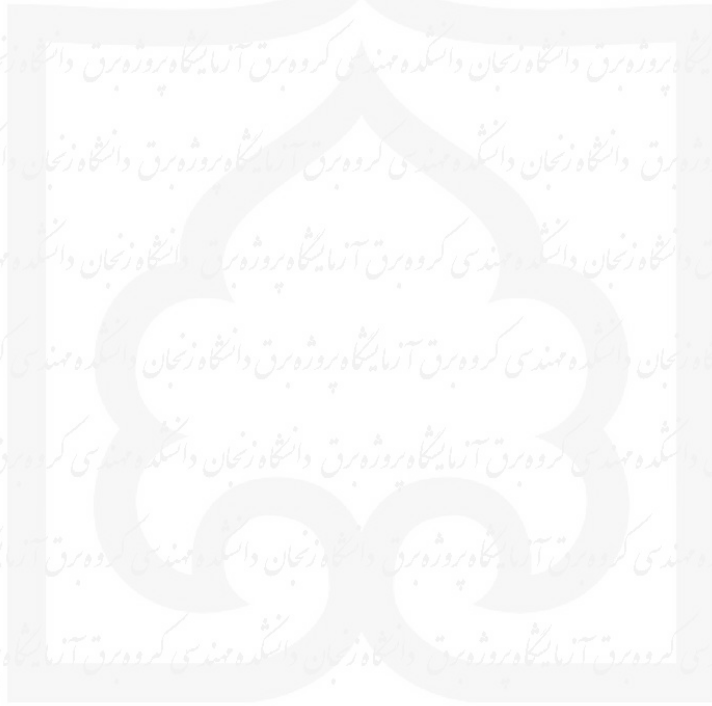
می‌کند و بنابراین صحبت را می‌توان به صورت شبه ایستان در دوره های کوتاه زمانی

(حدود ۲۰ms) در نظر گرفت و با یک فرآیند تصادفی ارگادیک در یک قطعه زمانی کوچک

مدل نمود و طیف مشخصی برای آن در این قطعه زمانی بدست آورد.

علاوه بر افزونگی های فوق عامل مهم دیگری که کاهش نرخ داده سیگنال صحبت را ممکن

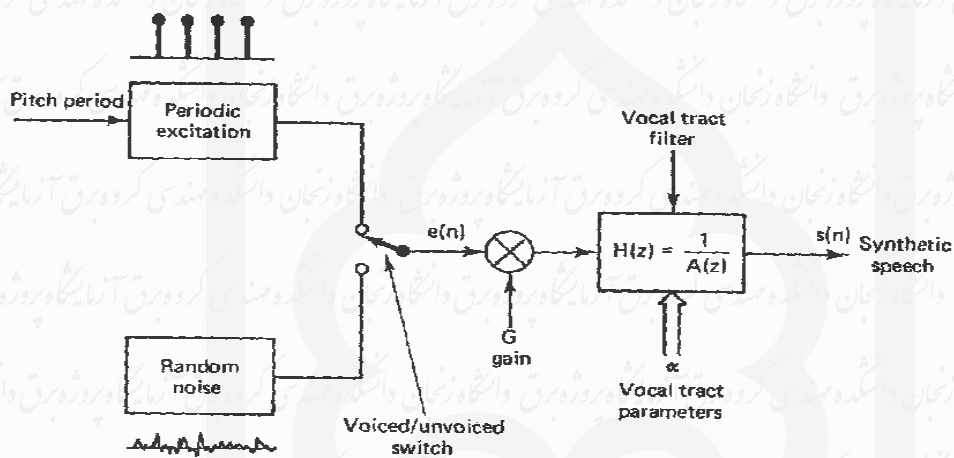
می‌سازد، طبیعت غیر حساس گوش انسان نسبت به بسیاری از ویژگیهای این سیگنال می‌باشد.



## ۲-۱- مدل سازی پیشگویی خطی

روش کدینگ پیشگویی خطی (LPC<sup>a</sup>) مبتنی بر مدل تولید صحبت در کد کننده های صوتی می باشد که در اینجا در شکل (۲-۱) نشان داده شده است. برای استفاده از مدل لازم

است که معلوم شود سیگنال با واك است یا بی واك و اگر با واك است پریود Pitch مجاسبه گردد. تفاوت اصلی بین LPC و سایر کدکننده های صوتی در مدل کردن لوله صوتی است. در تحلیل LPC ، لوله صوتی به صورت يك فیلتر دیجیتال تمام قطب در نظر گرفته می شود. [۴،۱].



شکل (۲-۱): مدل تولید صحبت در LPC

با شرکت دادن بهره  $G$  در این فیلتر داریم:

$$H(Z) = \frac{G}{1 + a_1 Z^{-1} + \dots + a_p Z^{-p}} = \frac{S(Z)}{E(Z)}$$

<sup>a</sup> Linear Predictive Coding

که در آن  $p$  مرتبه فیلتر است. اگر  $S(n)$  خروجی فیلتر مدل صحبت و  $e(n)$  تحریک ورودی باشد، معادله فوق را در حوزه زمان به صورت زیر می‌توان نوشت:

$$S(n) = Ge(n) - a_1S(n-1) - \dots - a_pS(n-p)$$

به عبارت دیگر هر نمونه صحبت به صورت ترکیب خطی از نمونه های قبلی قابل بیان است و این دلیل نام گذاری کدینگ پیشگویی خطی (LPC) می‌باشد.

### ۱-۲-۱- پنجره کردن سیگنال صحبت

روش LPC هنگامی دقیق است که به سیگنال‌های ایستاد<sup>۱</sup> اعمال شود، یعنی به

سیگنال‌هایی که رفتار آنها در زمان تغییر نمی‌کند. هر چند که این موضوع در مورد صحبت

صادق نیست، اما برای اینکه بتوانیم روش LPC را بکار ببریم، سیگنال صحبت را به قسمت

های کوچکی بنام "فریم" تقسیم می‌کنیم که این فریم‌ها شبیه ایستاد هستند. شکل (۱-۳) مثالی از

قسمت بندی سیگنال صحبت را نشان می‌دهد. این قسمت بندی با ضرب کردن سیگنال صحبت

$S(n)$ ، در سیگنال پنجره  $W(n)$  انجام می‌شود.



شکل (۱-۳): قسمت بندی سیگنال صحبت

معروف ترین انتخاب برای پنجره، پنجره همینگ (Hamming) به صورت زیر است:

$$W(n) = 0.54 - 0.46 \cos \frac{2\pi n}{N}, \quad 0 \leq n \leq N-1$$

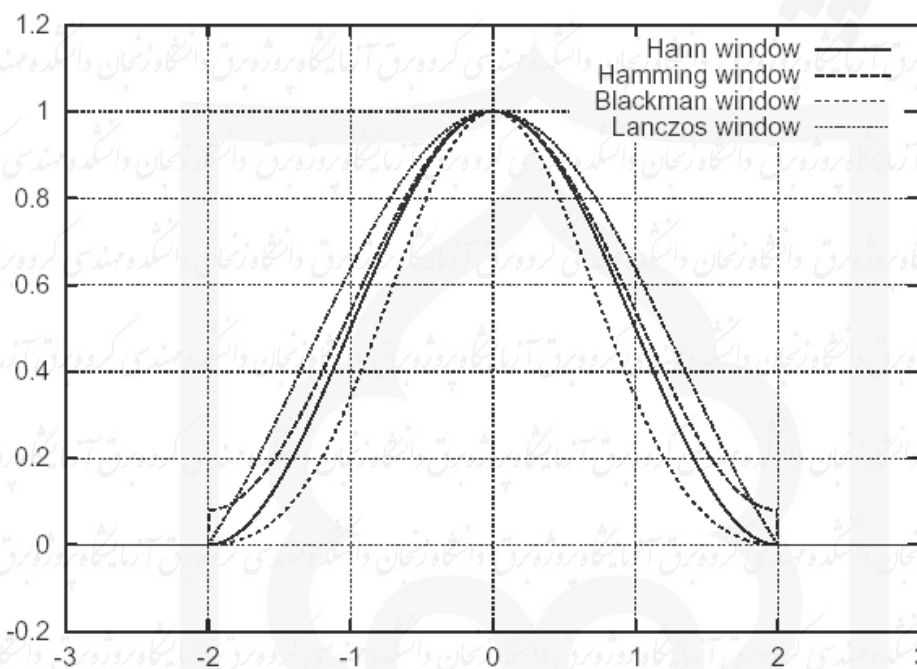
$$W(n) = 0, \quad \text{otherwise}$$

<sup>۱</sup> Stationary

در اینجا  $N$  ، طول پنجره دلخواه به نمونه و عموماً در محدوده ۱۶۰-۳۲۰ انتخاب می‌گردد

که ۲۴۰ يك مقدار نوعي می‌باشد . در شکل (۱-۴) چند پنجره معروف نشان داده شده است.

معمولاً پنجره هاي متوالي برروي هم همپوشاني دارند و فاصله بين آنها را پريود فریم می‌گویند. مقادير نوعي براي پريود فریم ۱۰-۳۰ms می‌باشد. این انتخاب به نرخ بیت و کیفیت صحبت دلخواه ما بستگی خواهد داشت. هر چه پريود فریم کوچکتر باشد، کیفیت بهتری خواهیم داشت.



شکل (۱-۴): نمایش چند پنجره معروف

### ۲-۲-۱- پیش تاکید سیگنال صحبت

شکل (۱-۵) يك توزیع طیفی نمونه سیگنال صحبت را برای اصوات باواک نشان

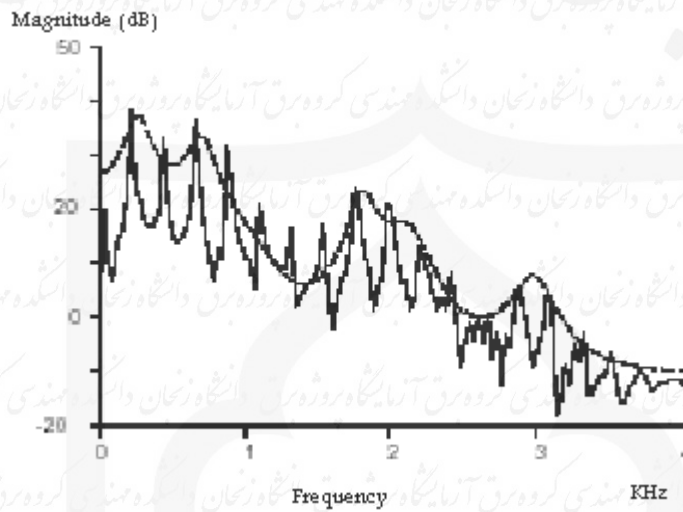
می‌دهد. با توجه به افت طیف در فرکانس هاي بالا و ضعيف بودن فرکانس هاي بالا در طیف صحبت ، تحلیل LPC در فرکانس هاي بالا عملکرد ضعيفي خواهد داشت. برای تقویت مؤلفه هاي فرکانس بالا صحبت ، آن را از يك فیلتر بالا گذر با تابع انتقال  $1 - az^{-1}$  که فیلتر پیش

تاکید نامیده می‌شود، عبور می‌دهیم. مقدار نوعی ضریب  $a$  معمولاً  $0.9375 = \frac{15}{16}$  در نظر

گرفته می‌شود.

اگر  $S(n)$  سیگنال ورودی باشد، سیگنال پیش تأکید شده  $S'(n)$  خواهد شد:

$$S'(n) = S(n) - 0.9375S(n-1)$$



شکل (۱-۵): پوشش طیفی نمونه اصوات باواک

### ۱-۲-۳- تخمین پارامترهای LPC

در اینجا لازم است که پارامترهای مدل LPC یعنی ضرایب  $a_i$  فیلتر و بهره  $G$  تعیین

گردند. اگر

$$\hat{S}(n) = -a_1 S(n-1) - \dots - a_p S(n-p)$$

تخمین  $S(n)$  از روی نمونه های قبلی باشد، ضرایب  $a_i$  را چنان تعیین می‌کنیم که خطای

$$\sum_n [s(n) - \hat{S}(n)]^2$$



## مراجع

۱- Panos E. Papamichalis, "Practical Approaches to Speech Coding ", Prentice-Hall

Inc. ,۱۹۸۷.

۲-B.S. Atal & R.Remde "A new model of LPC excited for producing natural-sounding speech at low bit rates", Proc.ICASSP pp.۶۱۴-۶۱۷ ۱۹۸۲.

۳-Jason P. Woodard , "Digital Speech Coding" , Mini-Thesis , Department of Electronics & Computer Science, University of Southampton, Jun ۱۹۹۴.

۴-J.Makhoul "Linear Prediction: A Tutorial Review " Proc. IEEE,vol ۶۳, No.۴,pp.۵۶۱-۵۸۰ Apr ۱۹۷۵.

۵- Kondoz A.M," Digital Speech , Coding of low bit rate communication Systems", Chichester Wiley ۲۰۰۰

۶- Chen,Cox & Lin,"A Low-Delay CELP Coder for the CCITT ۱۶ kb/s Speech Coding Standard",IEEE Jour. On Selected Area in Comm.,vol.۱۰,no.۵ ,June ۹۲.

۷- Schroeder & Atal , "Code-Excited linear Prediction (CELP):High quality speech