הטכניון – מכון טכנולוגי לישראל
Technion – Israel Institute of Technology

ספריות הטכניון
The Technion Libraries

בית הספר ללימודי מוסמכים ע"ש ארווין וג'ואן ג'ייקובס
Irwin and Joan Jacobs Graduate School

קידוד אותות דיבור על-ידי פרוק וקטורי של
אות השארית מחזוי ליבארי

חבור על מחקר
לשם מילוי חלקי של הדרישות לקבלת התואר

מגיסטר למדעים
בהנדסת חשמל

מחסן

מאת
ארז עופר

הוגש לסנט הטכניון - מכון טכנולוגי לישראל
תשרי - תשמ"ז חיפה ספטמבר 1986

תודתי העמוקה נתונה לפרופסור דוד מלאך על הנחיתו
המועילה והמסורה ועל סבלנותו הרבה בכל שלבי המחקר

תודתי נתונה לחברת תדיראן - חטיבת התקשורת
שתמכה במחקר.

כמו כן ברצוני להודות לדוקטור אמיר דמבו
על הרעיונות הפוריים והשיחות המועילות.
לגברת זיוה אבני על העזרה בעבודת המחשב.
למהנדס המעבדה מר יורם אור חן על עזרתו
הרבה במעבדה ולכל שאר סגל המעבדה וחברי
שעזרו להשלים עבודה זו.

עמוד

ת ק צ י ר

בעבודה זו אנו מציגים שיטה חדשה לקידוד אות השארית מחזוי ליניארי: אות השארית מיוצג על-ידי קומבינציה ליניארית של מספר קטן של וקטורים הלקוחים מתוך אוסף נתון. אוסף הוקטורים נתון הן במשדר והן במקלט. הפרמטרים המשודרים הם אינדקסי הוקטורים שנבחרו והמקדמים הכופלים אותם בקומבינציה הלינארית.

טכניקות קידוד דיבור ספרתיות הקיימות כיום דורשות קצבי שידור בתחום 64 Kbps ל-2.4 Kbps . בראש התחום ניצבים מקודדי צורת גל (Waveform Coders) שהם בעלי איכות וחסינות לרעש גבוהים. הביצועים של מקודדים אלה נפגמים בצורה משמעותית כשקצב השידור יורד ל- 16 Kbps. בקצה השני של תחום הקצבים ניצבים מקודדי ה-LPC (pitch - excited LPC) שהם מקודדים המבוססים על מודל למערכת הפקת הדיבור האנושית. למקודדי ה-LPC איכות סינטטית והם אינם מתפקדים טוב בתנאי רעש רקע או ריבוי דוברים. הגדלת קצב השידור במקודדי LPC מעל הקצב האופייני של 2.4 Kbps אינה משפרת משמעותית את איכות הדיבור המתקבל.

ברור איפוא שקשה להשיג ביצועים טובים בתחום הקצבים 4.8-16 Kbps בעזרת מקודדי צורת גל או מקודדי LPC. שיטות שונות ומגוונות הוצעו לעבודה בתחום קצבים זה. משפחה של מקודדים שהוצעו לעבודה בתחום הקצבים הנ"ל הם מקודדי אות השארית (Residual Coders). מקודדי השארית מקודדים ומשדרים את אות השארית מחזוי ליניארי בנוסף למקדמי מסנן ה-LPC המשודרים במקודד LPC רגיל. בראש משפחה זו הן מבחינת קצב והן מבחינת איכות ניצבים מקודדים בהם מקודד כל דגם מאות השארית בצורה אינדיבידואלית, מערכת ה-APC (Adaptive Predictive Coding) היא דוגמא למערכת כזו. אולם, כשרוצים לרדת לקצב של 9.6 Kbps ומטה כבר לא ניתן לקודד כל דגם של אות השארית בנפרד. גישה אחת היא ה-Multi-Pulse LPC שבה אות השארית מיוצג על-ידי מספר דגמים חשובים. דגמים אלו נמצאים בעזרת תהליך איטרטיבי. גישה נוספת היא ה-CELP (Code Excited Linear Prediction), במקודדים אלה נעשה שימוש בבלוקים של רעש לבן גאוסי ליצוג אות השארית.

בעזרת הגישה החדשה המוצגת בעבודה זו אות השארית מקודד כקומבינציה ליניארית של מספר קטן של וקטורים. שיטה זו היא הכללה של הסכמאות שהוזכרו לעיל - Multi-Pulse ו-CELP.

בגישת הקידוד החדשה אנו נתקלים בשתי שאלות: ראשית, בהנתן אוסף הוקטורים כיצד נמצא את הקומבינציה האופטימלית ליצוג אות השארית? ושנית, מה יהיה אוסף וקטורים אופטימאלי?

פתרון השאלה הראשונה מהווה משימה קשה מכיון שאפילו בהנתן אוסף קטן של וקטורים מספר הקומבינציות הוא גדול. לכן, עלינו לבחון פתרונות תת-אופטימאליים. אחד הפתרונות האלה הוא איטרטיבי ופשוט ומוביל ל-Multi-Pulse LPC.
פתרון השאלה השניה מוביל לאלגוריתם לא-איטרטיבי הפועל בתחום התדר. למערכת שהתקבלה קראנו - Predictive Transform Coder (PTC). מערכת ה-PTC היא בעלת סיבוכיות נמוכה יותר מזו של הסכמאות האיטרטיביות בתחום הקצבים 9.6-16 Kbps.
עבור 16 Kbps האיכות המתקבלת טובה מזו המתקבלת במערכת האיטרטיבית.
בקצב 9.6 Kbps האיכות טובה אך ניתן להבחין באפקט של "low passing".

בעבודה מוצג תיאור מתמטי פורמאלי של סכמת הקידוד החדשה לאות השארית. ניתן פתרון אנליטי למציאת קומבינצית הוקטורים האופטימאלית. מוצג פתרון תת אופטימאלי איטרטיבי ומוצגת הסכמה החדשה של PTC. מוצגות בחינות אוביקטיביות וסוביקטיביות להערכת ביצועי מערכת ה-PTC והמערכת האיטרטיבית (Multi-Pulse) מבחינת איכות דיבור, חסינות לרעש רקע, חסינות לרעש ערוץ, tandeming והעברת אותות נתונים (Data).

רשימת סמלים

$H(z)$ - פונקצית התמסורת של מסנן הסינטזה ב-LPC.

G - הגבר אות הערור למסנן הסינטזה ב-LPC.

$A(z)$ - פונקצית התמסורת של המסנן ההפוך.

s_n - דגם אות דיבור מקורי.

$\hat{s}_n$ - דגם אות דיבור משוחזר.

a_k - מקדם של המסנן ההפוך.

u_n - דגם של אות הערור.

P - סדר מסנן ה-LPC.

e_n - דגם של אות השארית.

E - שגיאה ריבועית ממוצעת.

r_i - מקדם אוטוקורלציה.

k_i - מקדם החזרה - Reflection Coefficient.

$E(z)$ - התמרת z של אות השארית.

$S(z)$ - התמרת z של אות הדבור המקורי.

E_w - שגיאה ריבועית ממוצעת משוקללת.

V - אוסף הוקטורים בפרוק וקטורי של אות השארית.

$\underline{S}$ - וקטור אות דיבור מקורי.

$\underline{U}$ - וקטור אות הערור.

$\underline{v}_i$ - וקטור הלקוח מ-V.

x_i - המקדם הכופל את הוקטור $\underline{v}_i$.

w_n - התגובה להלם של המסנן המשוקלל.

I_n - מטריצה יחידה מסדר N.

$S(f)$ - התמרת פוריה של אות הדבור המקורי.

$\hat{S}(f)$ - התמרה פוריה של אות הדבור המשותזר.

$W(f)$ - התמרת פוריה של התגובה להלם של המסנן המשקלל.

f_s - תדר הדגימה.

$W(z)$ - פונקצית התמסורת של המסנן המשקלל.

h_n - התגובה להלם של מסנן הסינטזה.

l_n - leftover, האות הנוצר כתוצאה מזכרון המסנן.

s'_n - אות הדבור המקורי לאחר הפחתת l_n.

h'_n - התגובה להלם של מסנן סינטזה משוקלל.

$H'(z)$ - התמרת z של התגובה להלם של מסנן סינטזה משוקלל.

$\underline{e}$ - וקטור אות השארית.

$\underline{x}$ - וקטור המקדמים הכופלים את הוקטורים המרכיבים את הערור.

Q - מטריצת הוקטורים שנבחרו מהאוסף V.

F - מטריצה NxN מכילה בשורה n, בעמודה i את האבר h'_{n-i}.

$\underline{L}$ - וקטור המהווה את אות ה-leftover (l_n).

$E(\ )$ - תוחלת.

$\tilde{F}$ - מטריצת פוריה.

$\underline{e}'$ - וקטור אות שארית צבוע.

$\hat{\underline{e}}'$ - וקטור אות שארית צבוע משוחזר.

γ - מקדם שקלול השגיאה.

$Re[\ ]$ - חלק ממשי של ביטוי קומפלקסי.

$Im[\ ]$ - חלק מדומה של ביטוי קומפלקסי.

$\sigma^2(k)$ - ווריאנס הרכיב ה-k-י.

$\hat{\sigma}^2(k)$ - ווריאנס משוערך של הרכיב ה-k-י.

$\Delta(k)$ - צעד קוונטיזציה עבור הרכיב ה-k-י.

$b(k)$ - מספר סיביות מוקצות לרכיב ה-k-י.

$\bar{B}$ - קצב ממוצע.

m - גודל הבלוק עליו מחושבים מקדמי אוטוקורלציה באנליזת LPC.

M - אינטרוול עדכון מקדמי ה-LPC.

B - מספר הסיביות הנתון לקידוד ההתמרה ב-PTC.

n_b - מספר סיביות מקסימאלי הנתון לקידוד מקדם יחיד ב-PTC.

L - מספר הנקודות השונות מ-0 ב-h'_n.

J - מספר הנקודות המחושב לצורך אינטרפולציה ב-PTC.

P_e - הסתברות השגיאה לסיבית בודדת.

רשימת קיצורים

LPC	-	Linear Prediction Coding.
Kbps	-	K bit per second.
V/UV	-	Voiced / Un Voiced.
APC	-	Adaptive Predictive Coding.
RELP	-	Residual Excited Linear Prediction.
MRM	-	Maximum Residual Magnitude.
CELP	-	Code Excited Linear Prediction.
HFR	-	High Frequency Regeneration.
ATC	-	Adaptive Transform Coding.
PTC	-	Predictive Transform Coder.
DFT	-	Discrete Fourier Transform.
IDFT	-	Inverse DFT.
DCT	-	Discrete Cosine Transform.
GMRM	-	Generalized MRM.
SNR	-	Signal to Noise Ratio.
SNRSEG	-	Segmented Signal to Noise Ratio.
BER	-	Bit Error Rate.
VBD	-	Voice Band Data.
SF	-	Scale Factor.
MPLPC	-	Multi-Pulse LPC.
LAR	-	Log Area Ratio.
MAD	-	Multiply and Add.
DSP	-	Digital Signal Processor.
P.D.	-	Positive Definite.

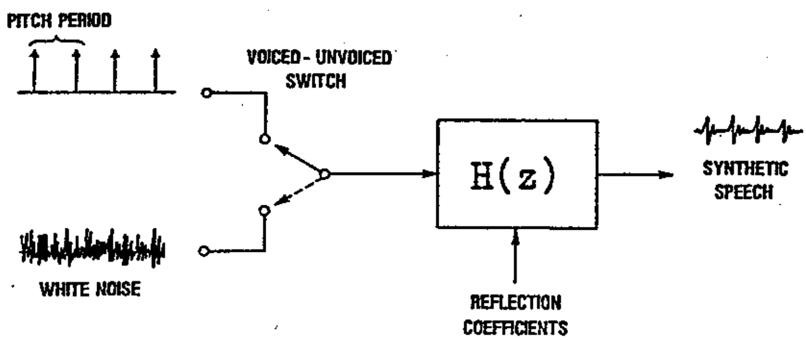
פרק 1. מ ב ו א

קידוד אותות דיבור בעזרת חיזוי ליניארי (LPC) היא שיטת קידוד מקובלת לקצבי שידור נמוכים - 2.4 Kbps . השיטה מבוססת על המודל המוצג בציור 1.1 . המסנן H(z) הוא מסנן all-pole המייצג את פונקצית התמסורת של המעבר הקולי (Vocal Tract). את הצלילים הנוצרים על ידי מערכת הקול האנושית ניתן לחלק בצורה גסה לשני סוגים:

(i) צלילים קוליים הנוצרים על-ידי ערור מחזורי של המעבר הקולי. הערור המחזורי נוצר על-ידי ריטוט מיתרי הקול והמעבר הקולי מעורר במספר תדרי תהודה הנקראים פורמנטים (Formants).

(ii) צלילים אל-קוליים. צלילים אלה לא נוצרים על-ידי ריטוט מיתרי הקול, אלא על-ידי מעבר אויר במעבר הקולי והתקלותו במכשולים (שיניים, שפתיים וכו').

לפי שני סוגי הערור הנ"ל הערור למסנן H(z) הוא אחד משני סוגים: ערור מחזורי לקטעי דיבור קוליים או רעש לבן לקטעי דיבור אל-קוליים.



ציור 1.1 - מודל יצירת הדיבור ב-LPC

Fig. 1.1 Speech production model in LPC

התדר היסודי בו מעורר המעבר הקולי בזמן צלילים קוליים נקרא תדר ה-Pitch.

כדי להשיג איכות דיבור גבוהה תוך שימוש ב-LPC יש לשדר מידע נוסף לגבי אות הערור ולא להסתפק בערור פשטני מהסוג הנ"ל (המכונה BUZZ/HISS). למעשה, שידור מדויק של אות השארית מחזוי לינארי, המתקבל על-ידי העברת אות הדיבור דרך המסנן ההפוך ל-$H(z)$, היה מאפשר שחזור מושלם של האות המקורי. אולם, שידור כזה דורש מספר רב של סיביות.

לכן הוצעו שיטות שונות לקידוד אות השארית בצורה יעילה יותר.

בפרק 2 נסקור את שיטת ה-LPC ובפרק 3 נסקור מקצת מהשיטות שהוצעו לקידוד אות השארית, במירב השיטות האלה הביצועים הופכים גרועים במהירות כשיורדים בקצב לתחום ה-9.6 Kbps.

בפרק 4 נציע שיטה חדשה לקידוד אות השארית - פירוק וקטורי (Vector expansion), לפיה אנו מרכיבים את אות הערור למסנן הסינטזה כקומבינציה לינארית של וקטורים הלקוחים מתוך אוסף נתון במשדר ובמקלט. האינפורמציה המשודרת תכיל את אינדקסי הוקטורים וערכי המקדמים הכופלים אותם. שיטה זו היא יצוג כללי למספר שיטות קידוד שהוצעו בשנים האחרונות, ביניהם: Multi-Pulse LPC [13], CELP [27] ו-Regular excitation [23].

לאחר הצגת השיטה נתמודד עם שתי שאלות:

(i) בהנתן אוסף הוקטורים כיצד נמצא את הקומבינציה הלינארית האופטימלית של וקטורים?

(ii) מהו אוסף וקטורים אופטימאלי?

נראה שפתרון מדויק לבעית מציאת הקומבינציה הלינארית אינו מעשי וכך יש צורך לפנות לפתרונות תת-אופטימליים.

אחד הפתרונות התת-אופטימליים מוליך לפתרון איטרטיבי פשוט, ניתוח מפורט של האלגוריתמים האיטרטיביים הנובעים מפתרון זה יובא בפרק 5.

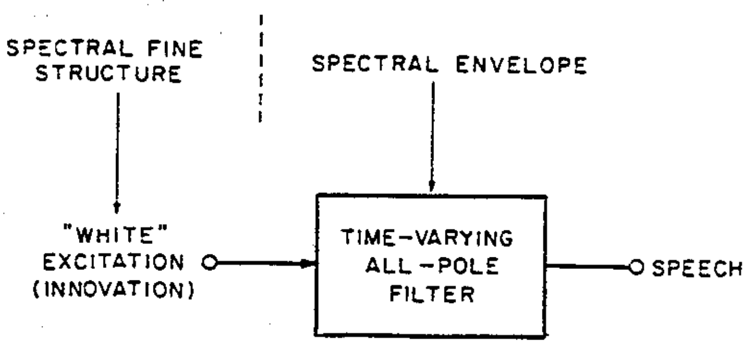
השאלה לגבי מציאת אוסף וקטורים אופטימלי תוביל אותנו לפתרון לא איטרטיבי למציאת הקומבינציה הלינארית. פתרון זה ופתרון לא איטרטיבי נוסף יידונו בפרק 6. לסיכום, נשווה בפרק 7 בין האלגוריתמים האיטרטיביים והלא איטרטיביים, נבחן את עמידותם לשגיאות ערוץ, את ביצועיהם ב-tandeming ואת יכולת ההעברה של אותות נתונים (Voice band data).

פרק 2. חיזוי ליניארי של דיבור (LPC)

2.1 מבוא

מקודדי הדיבור המקובלים כיום מתחלקים לשתי קבוצות עיקריות [1]:
הקבוצה הראשונה - מקודדי צורת גל (Wave Form Coders).
הקבוצה השניה - מקודדי מקור (Source Coders) או Vocoders.

מקודדי צורת הגל [2] כשמם כן הם, צורת הגל מקודדת תוך נסיון לשחזור כמה שיותר מדויק במקלט. בקצבים הגבוהים מקודדים אלה מתוכננים להיות בלתי תלויים באות ולכן ניתן לשדר בעזרתם סוגים שונים של אותות. מקודדים אלו חסינים בדרך-כלל לשגיאות בערוץ ולרעש המתווסף לאות הדיבור, הם פשוטים יותר ונוחים בדרך כלל למימוש בזמן אמיתי. חסרונם הבולט הוא הקצב הגבוה יחסית הנדרש להשגת איכות דיבור גבוהה: 16 Kbps - 64 Kbps והנפילה המהירה באיכות הדיבור כשיורדים בקצב.
הקבוצה השניה של המקודדים - Vocoders, תלויה בידע מוקדם על אופן יצירת האות במקור. ידע זה מנוצל כדי לתאר את אות הדיבור בצורה יעילה - מגדירים מודל למבנהו של האות ולאופן יצירתו והמקודד מסתפק בהעברת מידע על המודל בלבד.
המודל המקובל לתהליך יצירת אות הדיבור הוא המודל המוצג בציור 2.1 . הוא מורכב משני חלקים שמניחים שהם בלתי תלויים - אות ערור ומסנן ליניארי המייצג את המעבר הקולי. שימוש במודל כזה מקטין את קצב השידור, אך תמורת הקטנה זו אנו משלמים באיכות נמוכה יותר ובהקטנת החסינות לרעש או שגיאות ערוץ.
אחת השיטות המקובלות והפשוטות למימוש המודל שתואר לעיל היא החיזוי הליניארי - Linear Prediction נקדיש את המשך הפרק לדיון בשיטה זו.



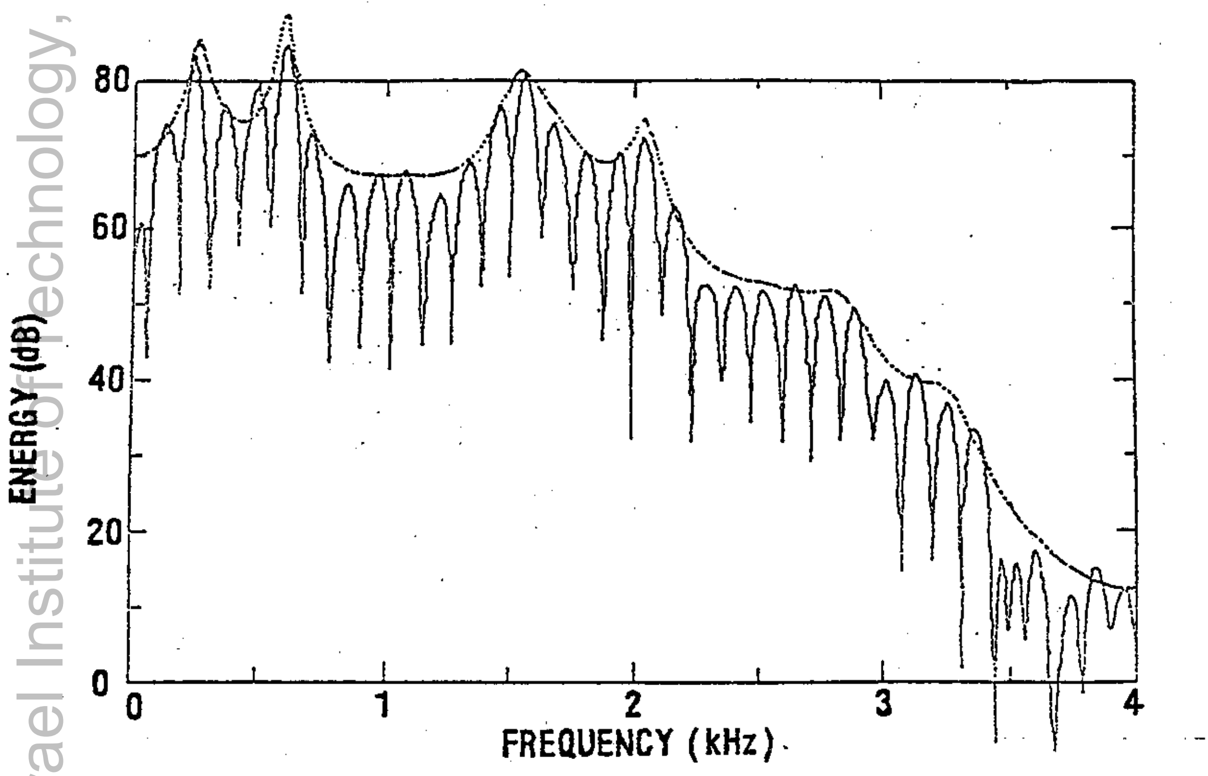
ציור 2.1 מודל יצירת הדיבור

Fig. 2.1 - Speech production model

2.2 תאור מודל ה-LPC

כל אותות הדיבור מופקים על ידי ערור המעבר הקולי האנושי בערור אקוסטי הנוצר במיתרי הקול בקטעי דיבור קוליים, או על-ידי מערבולות אויר הנוצרות במעבר הקולי בקטעי דיבור אל-קוליים. במערכת Vocoder המעבר הקולי מיוצג על-ידי מסנן לינארי.

הרעיון העומד מאחורי קידוד בעזרת חיזוי לינארי - (Linear Prediction Coding) [5], [4], [3] הוא שכל דגם של אות הדיבור ניתן ליצוג כקומבינציה לינארית של דגמי אות דיבור קודמים ושל הדגם הנוכחי של אות הערור. יצוג זה מוביל לכך שהמסנן המייצג את המעבר הקולי הוא מסנן all-pole, מסנן זה תורם למערכת את המעטפת הספקטרלית לזמן קצר של האות (Short-time Spectral envelope). עידון המעטפת כדי להגיע לספקטרום האמיתי של האות (Fine structure) מתקבל על-ידי הערור. ספקטרום אופיני ומעטפת ספקטרלית של אות הדיבור מוצגים בציור 2.2 .



ציור 2.2 - ספקטרום לזמן קצר ומעטפת ספקטרלית שהתקבלה על-ידי חיזוי לינארי

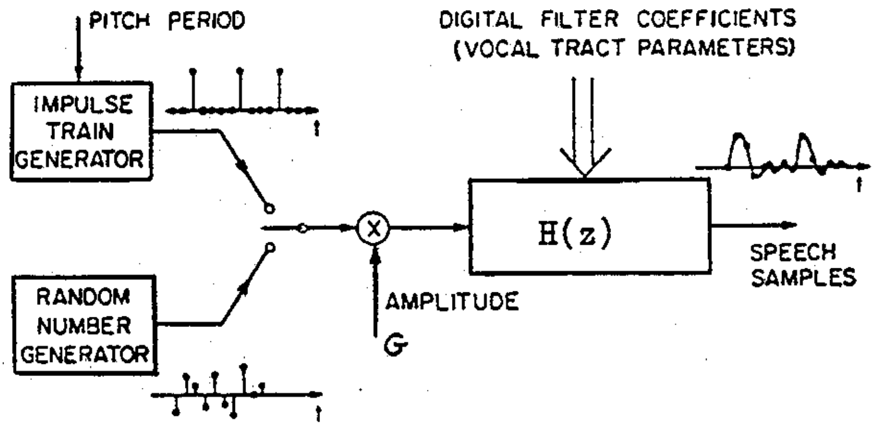
Fig. 2.2 - Short-time spectrum and spectrum envelope obtained by linear prediction

אות הערור ניתן לחלוקה גסה לשתי קבוצות:
בזמן הפקת אותות קוליים (Voiced Sounds) המעבר הקולי מעורר על-ידי ערור כמעט מחזורי. בזמן הפקת אותות לא-קוליים (Unvoiced sounds)אות הערור הוא רעש לבן.
ב-LPC Vocoder נעשות הנחות פשטות על הערור וניתנים שני סוגי ערור:

(i) סדרת הלמים במחזור ה-Pitch, שהוא התדר היסודי של אות הדיבור, מתארת את אות הערור בקטעי דיבור קוליים.

(ii) רעש לבן מתאר את אות הערור בזמן קטעי דיבור לא קוליים.

המערכת המתקבלת מוצגת בציור 2.3:



ציור 2.3 - מודל יצירת הדיבור ב-LPC

Fig. 2.3 - Speech production model in LPC

במודל מניחים שבפרקי זמן קצרים אות הדיבור סטציונרי. בכל מסגרת זמן כזו נקבעים: מקדמי מסנן ה-all-pole,קבוע הגבר G בו יש להכפיל את הערור, מחזור ה-Pitch אם המסגרת היא קולית והחלטה אם המסגרת היא קולית או לא-קולית (V/UV).
במערכת כזו ניתן לרדת לקצבי שידור של 2.4 Kbps אולם האיכות המתקבלת היא איכות סינטטית, האיכות לא תשתפר אם נעלה בקצב.

2.3. קביעת פרמטרי LPC

היתרון הגדול של מודל all-pole למערכת יצור הדיבור נעוץ בכך שכדי למצוא את מקדמי המסנן יש לפתור מערכת משוואות ליניארית.
במודל all-pole נקבל שדגם המוצא ה-n-י יהיה:

$$s_n = -\sum_{k=1}^{p} a_k s_{n-k} + Gu_n \qquad (2.3.1)$$

כאשר u_n הוא הדגם ה-n-י של אות הערור. מקובל להניח שהערור לבן (למרות שבזמן קטעים קוליים אין הדבר כך) ואז השערוך הטוב ביותר לדגם ה-n-י בהתבסס על הדגמים הקודמים יהיה:

$$\hat{s}_n = -\sum_{k=1}^{p} a_k s_{n-k} \qquad (2.3.2)$$

נגדיר את שגיאת השיערוך כהפרש בין דגם אות דיבור מקורי ודגם אות משוחזר:

$$e_n = s_n - \hat{s}_n \qquad (2.3.3)$$

שגיאה ריבועית ממוצעת על פני המסגרת תהיה:

$$E = \frac{1}{N} \sum_{n=o}^{N-1} \left(s_n + \sum_{k=1}^{p} a_k s_{n-k}\right)^2 \qquad (2.3.4)$$

מקדמי החזאי האופטימלי נבחרים כך שהשגיאה הריבועית הממוצעת תהיה מינימלית. דבר זה מתבצע על-ידי גזירת E לפי a_i, $i=1...p$ והשוואה ל-0.
מתקבלת מערכת משוואות ליניאריות אותה יש לפתור למציאת המקדמים a_i:

$$\sum_{k=1}^{p} \Phi_{jk}\, a_k = -\Phi_{jo} \qquad j=1,2...p \tag{2.3.5}$$

$$\Phi_{jk} = \sum_{n} s_{n-j}\, s_{n-k} \tag{2.3.6}$$

תחום הסכימה ב-(2.3.6) קובע שתי שיטות אנליזה:
בשיטת הקווריאנס:

$$\Phi_{jk} = \sum_{n=p}^{N-1} s_{n-j}\, s_{n-k} \tag{2.3.7}$$

כלומר המינימיזציה מתבצעת רק על פני האינטרוול [p,N-1].
בשיטת האוטוקורלציה [6]:

$$\Phi_{jk} = \sum_{n=-\infty}^{\infty} s_{n-j}\, s_{n-k} \tag{2.3.8}$$

כשמניחים: $s_n=0$ עבור $n<0$ ו- $n \geq N-1$
במקרה זה נקבל:

$$\Phi_{jk} = r_{|j-k|} \tag{2.3.9}$$

כאשר r_i הוא מקדם אוטוקורלציה.

שתי השיטות השונות מובילות לשתי צורות פתרון שונות, לסיבוכיות שונה ולתכונות שונות של הפתרון [5].

בשיטת האוטוקורלציה יציבות תמיד מובטחת (תיאורטית - כלומר, ללא שגיאות חישוב). המטריצה המתקבלת במערכת המשוואות הלינאריות היא מטריצה של אברי אוטוקורלציה - מטריצת Toeplitz סימטרית וקיים פתרון פשוט על-ידי האלגוריתם של - Levinson Durbin [5].

בשיטת הקווריאנס לא מובטחת תמיד יציבות ויש לבדוק את המסנן המתקבל $A_{(z)}$ לוודא שכל שורשיו נמצאים בתוך מעגל היחידה. שיטת הפתרון במקרה זה יותר מסובכת ומתבססת בדרך כלל על פתרון מערכת המשוואות על-ידי פירוק Cholesky [5].

להשלמת מודל ה-LPC יש להחליט לגבי כל קטע אם הוא קולי או לא קולי ויש לחשב את תדר ה-Pitch עבור קטעים קוליים. שתי בעיות אלה קשות לפתרון מדויק ובמערכות קיימות הן גוזלות את מירב משאבי החישוב. בעבודה זו לא ניגע כלל בנושאים אלה.

2.4 קוונטיזציה של מקדמי ה-LPC

בבואנו לבצע קוונטיזציה של מקדמי ה-LPC עלינו לטפל בשתי בעיות:

(i) כמה סיביות יש להקצות לכל פרמטר משודר?

(ii) כיצד נבטיח יציבות למסנן המתקבל לאחר קוונטיזצית הפרמטרים?

קבוצת פרמטרים מקובלת לצרכי שידור האינפורמציה על מודל ה-LPC הם מקדמי ההחזרה (PARCOR).
המעבר ממקדמי המסנן $a_i,\ i=1...p$ למקדמי ההחזרה $k_i,\ i=1...p$ נעשה בצורה פשוטה:

בהנתן $a_i,\ i=1,2,...p$ ניתן לעבור ל- $k_m,\ m=1,2,...p$ על ידי p איטרציות והקשר הבא:

$$a_{m-1,i} = \frac{a_{mi} - k_m\, a_{m,m-i}}{1-k_m^2} \tag{2.4.1}$$

$k_m = a_{mm}$ ו-

עבור $m=p,p-1...1$ ו- $i=0,1,...m-1$

בהנתן $k_m,\ m=1,2,...p$ ניתן לעבור ל- $a_i,\ i=1,2,...p$
על-ידי p איטרציות והקשר הבא:

$$a_{m,i} = \begin{cases} a_{m-1,i} & i=0 \\ a_{m-1,i} + k_m\, a_{m-1,m-i} & i=1,2,...m-1 \\ k_m & i=m \end{cases} \tag{2.4.2}$$

עבור $m=1,2,...p$ $a_{oo}=1$

בין מקדמי ההחזרה קיים סדר כך שניתן להחליט על סכמת חלוקת סיביות בהתאם למספר המקדם [7]. כמו כן, ניתן לבדוק בעזרתם בצורה מאד פשוטה את היציבות - תנאי הכרחי ומספיק לכך שהמסנן יהיה יציב הוא: $|k_i|<1, \ i=1...p$ כאשר p הוא סדר המודל. הקצאת הסיביות במהלך עבודה זו נעשתה בהתאם להמלצה של LPC-10 [8] כך שעבור מודל של 10 מקדמים אנו זקוקים ל-41 סיביות לשידור סט מקדמים.

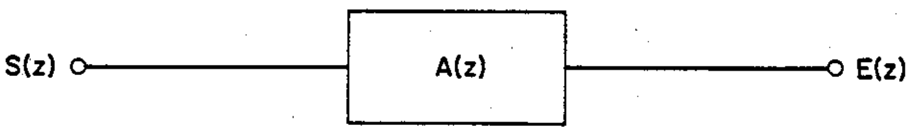
אות השארית מכיל את המידע שנותר על אות הדיבור לאחר שמיצינו ממנו את המידע לגבי המעטפת הספקטרלית, מידע הנמצא במסנן הסינטזה. כלומר, ניתן להשתמש באות השארית להשלמה מדויקת של המודל לקבלת אות דיבור משוחזר בצורה מושלמת.

אם מסנן הסינטזה הוא $\frac{1}{A_{(z)}}$ כאשר:

$$A_{(z)} = 1 + \sum_{k=1}^{p} a_k z^{-k} \qquad (2.5.1)$$

אזי אות השארית e_n מתקבל באמצעות המסנן ההפוך $A(z)$:

$$E_{(z)} = A_{(z)}\, S_{(z)} \qquad (2.5.2)$$

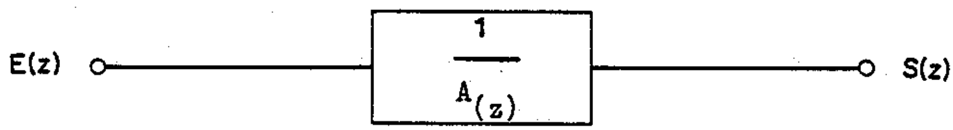


ציור 2.4 - קבלת אות השארית

Fig. 2.4 Generating the residual signal

שימוש באות השארית המתקבל כאות הערור למסנן ה-all-pole מוביל לשחזור של האות המקורי:

$$S_{(z)} = \frac{E_{(z)}}{A_{(z)}} = E_{(z)}\, H_{(z)} \tag{2.5.3}$$



ציור 2.5 - שחזור אות הדיבור מאות השארית

Fig. 2.5 Generating the speech signal using the residual signal

המידע הטמון באות השארית עשיר מזה הנמצא במודל ערור פשטני BUZZ/HISS. לכן נצפה לקבל תוצאות טובות יותר אם נצליח לקודד בצורה יעילה את אות השארית, זה ייעשה כמובן תמורת תוספת בקצב השידור.

2.6 חסרונות שיטת ה-LPC [9]

כאמור, במקודד LPC ניתן לרדת לקצבי שידור נמוכים מאד, בסביבות ה-2.4 Kbps. אולם, לשיטה כמה חסרונות בולטים, ננסה בסעיף זה לעמוד על כמה מהם ולהבין מאין הם נובעים.
הבעיות בקידוד LPC נובעות בעצם משתי סיבות עיקריות:

(i) חוסר התאמה מדויקת של מודל ה-all-pole למערכת המקורית וקוונטיזציה מקדמי ה-LPC.

(ii) ערור פשטני מדי - BUZZ/HISS ושגיאות בחישוב Pitch או החלטה V/UV.

חוסר ההתאמה של מודל ה-all-pole נובע מכמה גורמים:

1. מכיוון שהמודל הוא all-pole אפסים לא מיוצגים היטב. צלילים הנובעים גם מחלל האף, לגביהם ידוע שדרושים אפסים כדי ליצר אותם, לא מופקים בצורה טובה.

2. מיקום הפורמנטים מושפע על-ידי הרמוניות התדר היסודי וכך נכנס אי דיוק בשערוך הספקטרום.

3. מקדמי המסנן משתנים עם שינוי מקום החלון בזמן האנליזה.

4. בקצבי שידור נמוכים מקדמי המסנן עוברים קוונטיזציה בצורה שמכניסה שגיאה ניכרת.

כעקרון, ניתן לתלות את כל הירידה באיכות בפישוט של אות הערור. זאת מכיון שאילו היינו משדרים את אות השארית במדויק היינו מקבלים שחזור מדויק של האות המקורי.
ההרעה בביצועים עקב פישוט הערור ל-BUZZ/HISS נובעת מכמה סיבות:

1. מודל ה-BUZZ/HISS משאיר מחוץ למערכת אינפורמציה חשובה לגבי אמפליטודה ופאזה בספקטרום אות השארית.

2. קיימת בעיה לגלות באופן מושלם את מחזור ה-Pitch או להחליט באופן מושלם לגבי קטע דיבור באם הוא קולי או לא קולי.

כל הגורמים שהוזכרו חוברים יחד וגורמים לכך שאיכות הדיבור המתקבל היא לא טבעית. והירידה באיכות היא מהירה כשמתווסף רעש לאות או כאשר יש ריבוי דוברים.

כאמור, היינו מקבלים שחזור מדויק של האות המקורי אילו היינו משדרים במדויק את אות השארית. מכיוון ששידור מדויק יוביל לקצב שידור גבוה מדי הוצעו שיטות שונות לקידוד אות השארית כך שעדיין יישמר קצב שידור נמוך. בפרק הבא נסקור מקצת שיטות אלו.

פרק 3. מקודדי שארית

3.1 מבוא

בפרק הקודם ראינו שבעזרת מקודדי LPC קלאסיים ניתן להגיע לקצבים של 2.4 Kbps אולם האיכות המתקבלת היא איכות סינטטית. קידוד של אות השארית ושידורו בנוסף לאינפורמציה על מסנן ה-LPC תעלה את קצב השידור אך נצפה לקבל איכות טובה יותר. קידוד אות השארית יכול להעשות בצורות שונות וקמה משפחה שלמה של מקודדים שהמשותף להם היא העובדה שהם מקודדים את אות השארית בנוסף לאינפורמצית מסנן ה-LPC, למקודדים אלה קוראים - מקודדי שארית.

בראש הרשימה הן מבחינת סיב והן מבחינת קצב נמצאים המקודדים בהם מקודד אות השארית על בסיס של דגם - דגם, כלומר, כל דגם באות השארית עובר קוונטיזציה בצורה אינדיבידואלית. קצבים טיפוסיים לסוג מקודד זה הם 24 Kbps - 9.6 Kbps.
מקודד ה-APC (Adaptive Predictive Coding) [11], [12] הוא דוגמה לקבוצה זו של מקודדים.
כדי לרדת לתחום הקצבים של 9.6 Kbps ומטה לא ניתן כבר לבצע קוונטיזציה על בסיס של כל דגם ויש להשתמש בשיטות אחרות.

מקודדי ה-RELP או Baseband vocoder הם משפחה של מקודדי שארית בהם לא מועבר כל רוחב הסרט של אות השארית, עיקר המאמץ מושקע בהעברת תחום התדרים הנמוכים של האות (Base band), במקלט מבוצעת רגנרציה ליצירת תחום התדרים הגבוהים באות. מקודדים אלו הוצעו במקור [28] לעבודה בתחום הקצבים 4.8-9.6 Kbps.

הצעות נוספות שהופיעו לקידוד ה-Residual:

- Subband Coding [29] - הפעלת טכניקה של קידוד בפסי תדר נפרדים על אות השארית.

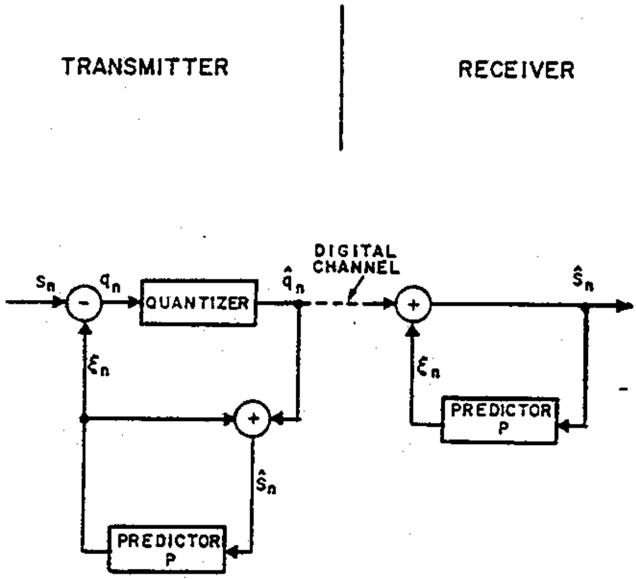
- Multi-Pulse LPC [13] - יצוג אות השארית על ידי מספר קטן של דגמים חשובים וקידודים. עיקר חסרונה של שיטה זו היא סיבוכיותה הגבוהה, בעקבותיה הופיעו הצעות נוספות לקידוד אות השארית בעזרת מספר קטן של דגמים תוך נסיון לשמור על סיבוכיות נמוכה. לדוגמא, Maximum Residual Magnitude (MRM) [16], [17].

- CELP - Code Excited Linear Prediction [27] - הצעה זו שנראית מבטיחה מאד מבוססת על החזקת מילון ובו קטעי אות אקראיים. הקידוד מתבצע על-ידי חפוש הקטע הקרוב ביותר לקטע שיש לשדר מבחינת שגיאה ריבועית ממוצעת בין אות המקור והאות המשוחזר. הפרמטר שמשודר עבור כל מסגרת הוא האינדקס של וקטור הערור המתאים. החסרון העיקרי של אלגוריתם זה היא סיבוכיות גבוהה מאד (אף יותר מזו של Multi-Pulse LPC) ולאחרונה הוצעו שיטות להקטין סיבוכיות זו [30].

בהמשך הפרק נתאר בקצרה את הסוגים השונים של מקודדי השארית.

[12] ,[11] APC - Adaptive Predictive Coder 3.2

הסכמה הבסיסית של מקודד APC מוצגת בציור 3.1 .



ציור 3.1 סכמת בלוקים של predictive coder

Fig. 3.1 Block diagram of a predective coder

הקונטיזר Q והחזאי P בסכמה הזאת שניהם אדפטיביים.
בערוץ משודרת אינפורמציה הן על אות השארית והן על הפרמטרים המשתנים של החזאי והקונטיזר (Side information). אות השארית דורש בדרך כלל את רוב הסיביות המוקצות לשידור בהשוואה לאינפורמצית הצד. לדוגמא, עבור אות דיבור דגום בקצב 8khz וקונטיזצית אות השארית ל-1 bit/sample ידרשו 8 Kbps לצורך העברת אות השארית ו-3-5 Kpbs לצורך העברת אינפורמצית הצד.
כדי לרדת במערכת זו לקצבים נמוכים יש לקודד את אות השארית ביעילות, עבור קצבים נמוכים מ-10 Kbps נדרש לקודד את אות השארית בפחות מסיבית אחת לדגם. קוונטיזציה גסה זו היא הגורם העיקרי לשגיאה באות המשוחזר.

<u>בחירת החזאי</u>

החזאי P מורכב משני חזאים נפרדים:

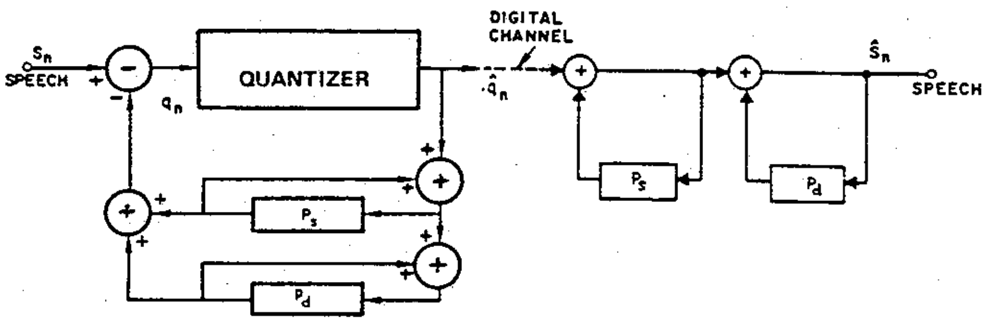
P_s - חזאי המבוסס על המעטפת הספקטרלית לזמן קצר.

P_d - חזאי Pitch.

החזאי P_s הוא חזאי מהסוג שנתקלנו בו בפרק 2 (LPC) כשמספר המקדמים בו משתמשים בדרך כלל הוא 6 או יותר.

P_d - הוא חזאי בעל 1-3 מקדמים, תפקידו להוציא מהאות את מבנה ה-Pitch.

שני החזאים משולבים במערכת שניתן לראות בציור 3.2 .



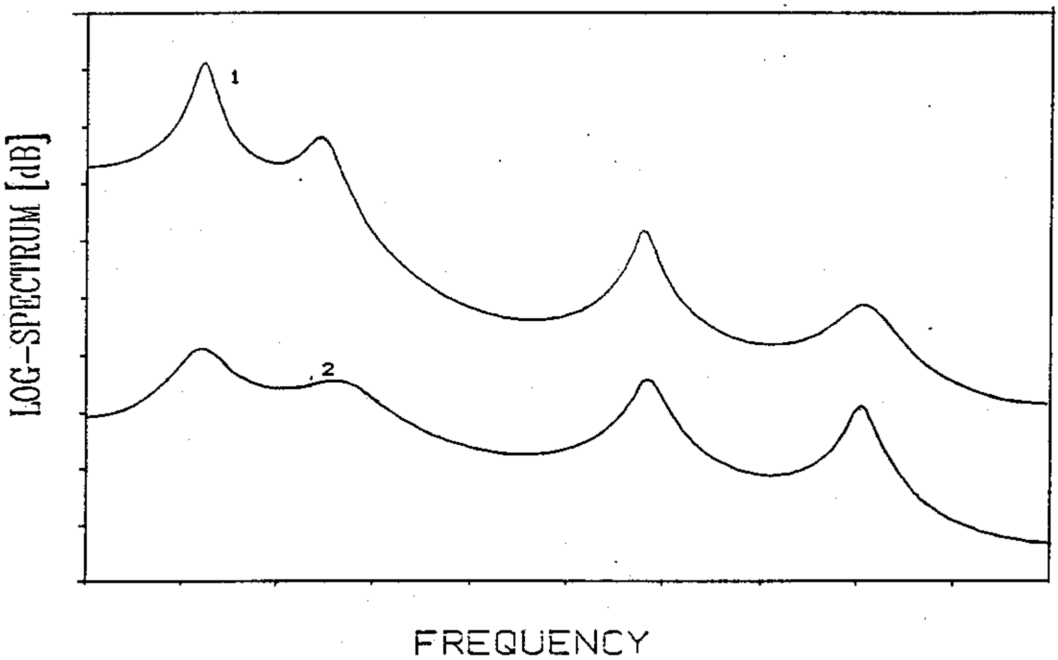
ציור 3.2 הוספת חזאי pitch ל-predictive coder

Fig. 3.2 Predictive Coder with pitch predictor

<u>APC עם צביעת שגיאה</u>

במערכת שתיארנו לעיל הרעש המתקבל הוא בקירוב לבן (ככל שהקוונטיזציה גסה יותר קירוב זה סוב פחות), והמערכות הראשונות מסוג APC היו בנויות לפי הסכמה הנ"ל. כיום ידוע שניתן לנצל את תכונת המיסוך של האוזן כדי להשיג איכות סוביקטיבית טובה יותר על-ידי "צביעת" הרעש. הרעש באזורי הפורמנטים, אזורים בהם מרוכזת אנרגיה גדולה, ממוסך על-ידי אות הדיבור ולכן נוכל להרשות לרכיבי תדר באות השגיאה באזורי הפורמנטים להיות בעלי אנרגיה גבוהה יותר יחסית לרכיבי התדר של הרעש באזורי התדר בין הפורמנטים.

כמו כן קיים אפקט של מסוך התדרים הגבוהים על-ידי הפורמנט הראשון והשני שהם בדרך כלל בעלי אנרגיה גדולה, לכן, אפשר להרשות יחסי SNR נמוכים יותר באזורי התדר הגבוה. צביעה כזאת של הרעש מודגמת בצורה איכותית בציור 3.3 .



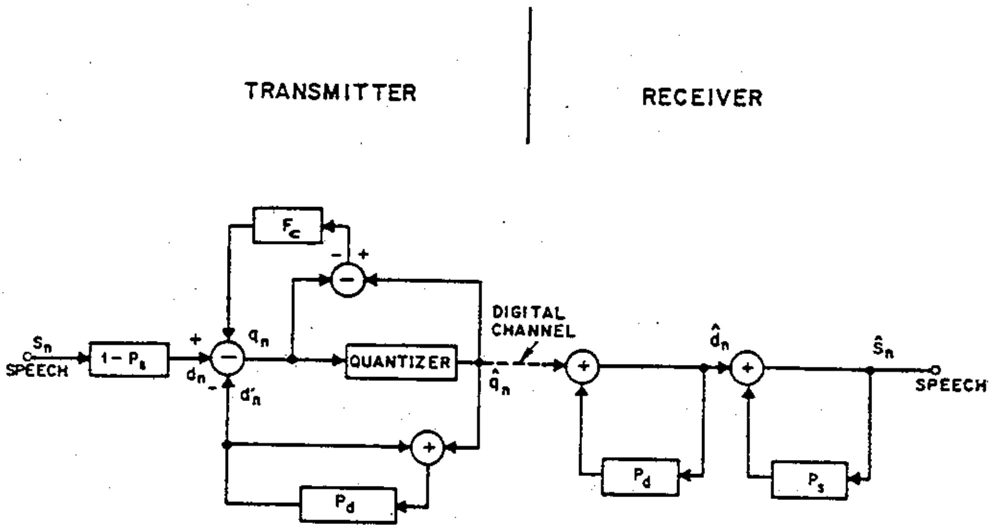
1 - Speech spectral envelope

2 - Error spectral envelope

ציור 3.3 "צביעת" הרעש

Fig. 3.3 Spectrum of colored noise

בציור 3.4 מוצגת סכמה בה יש שליטה על מעטפת ספקטרום השגיאה בעזרת המסנן F_c.



ציור 3.4 APC עם שליטה על מעטפת ספקטרום הרעש

Fig. 3.4 APC with Noise Spectrum Control

בפרק 4 נרחיב את הדיבור על "צביעת" רעש מכיון שנשתמש בטכניקה זו גם במערכות שנציע בהמשך.

3.3 מקודדי RELP (Base band) [28], [32]

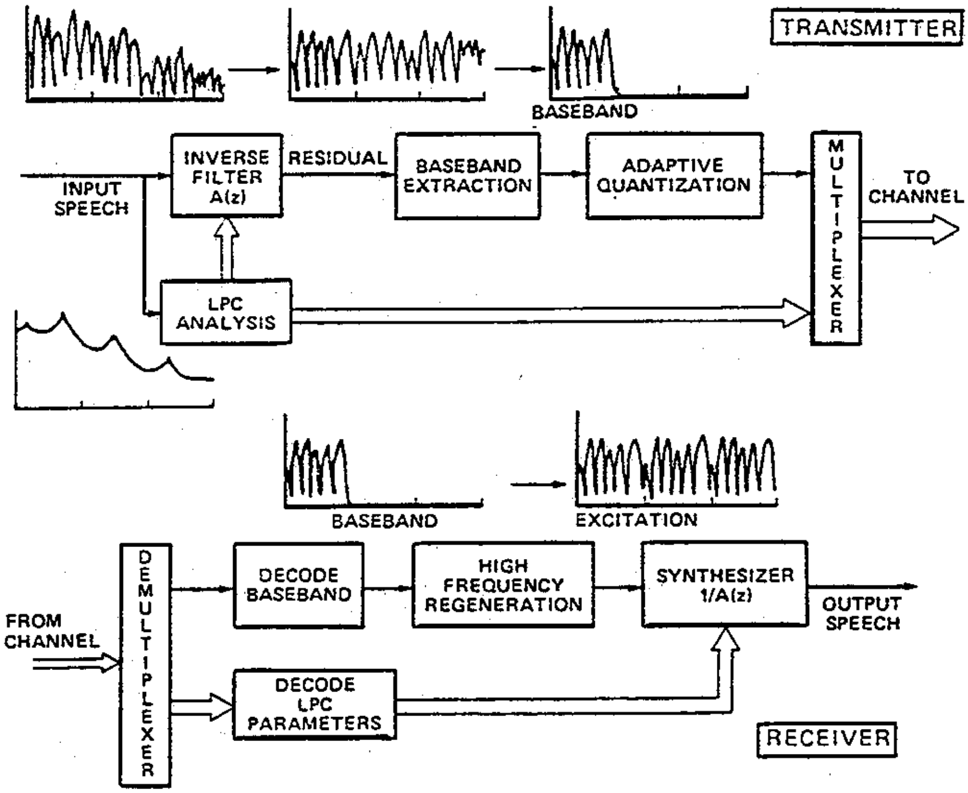
הרעיון הבסיסי העומד מאחורי מקודדי ה-RELP הוא להשקיע את מירב המשאבים בקידוד אזור התדר הנמוך באות השארית (base band). במקלט יוצרים מחדש את התדרים הגבוהים בעזרת פעולה לא לינארית (High Frequency Regeneration). מקודד ה-RELP המקורי שהוצע [28] יועד לעבודה בתחום הקצבים 4.8kbps-9.6kbps.

בציור 3.5 מוצגת סכמה עקרונית של מקודד Base band. איכות האות המתקבל תלויה באופן ניכר בצורת השחזור של התדרים הגבוהים במקלט.
השיטות ל-HFR (High Frequency Regeneration) משתמשות בעובדה שהאינפורמציה לגבי אות הדיבור מצויה בעיקר בתדרים נמוכים ובעובדה שהספקטרום של אות השארית הוא בקירוב לבן. פעולת ה-HFR מרחיבה את ה-baseband לספקטרום שטוח בכל רוחב הסרט.
שתי שיטות פופלאריות לביצוע ה-HFR הן:

(i) Rectification & Spectral Flattening - הרבה מהסכמאות שהוצעו ל-baseband vocoders השתמשו בשיטה זו: אות השארית הנוצר מה-baseband (האות עובר דצימציה ואינטרפולציה לצורך השידור) עובר יישור (rectification), לאחר פעולה לא לינארית זו מבצעים Spectral Flattening כלומר, מנסים להגיע לספקטרום לבן ככל האפשר כתוצאה סופית.

(ii) שיטה שניה לרגנרציה של התדרים הגבוהים היא Spectral Duplication - כלומר, משכפלים את ה-baseband להשלמת הספקטרום בכל רוחב הסרט. שיטה זו מתבססת על העובדה שעבור אותות קוליים לספקטרום אות השארית צורה בקירוב מחזורית ועבור אותות לא קוליים צורה אקראית כשבשני המקרים המעטפת שטוחה. את השכפול מבצעים בשתי צורות - Spectral Folding או-Spectral Translation [31].

קידוד אות השארית לאחר דצימציה יכול להעשות בשיטות שונות:
Adaptive quantization, APC, ATC, Sub-band Coding וכו' בספרות מופיעות הצעות שונות למערכת עם צורות קידוד שונות לאות השארית.

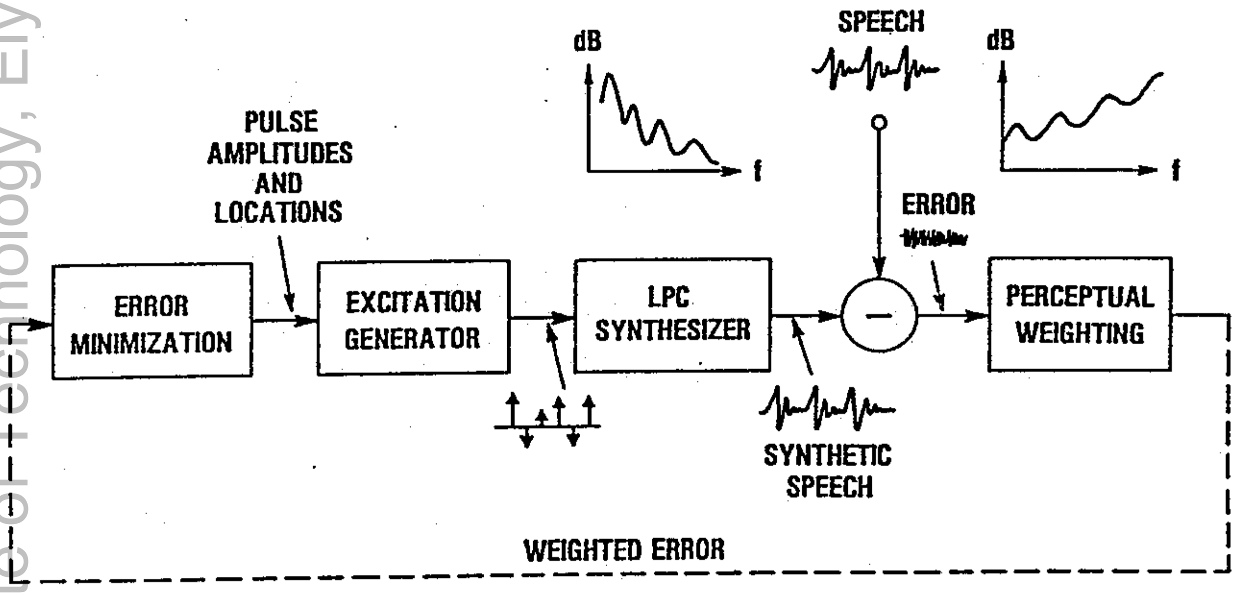


Baseband vocoder מערכת - 3.5 ציור

Fig. 3.5 Baseband vocoder

Multipulse LPC 3.4

שיטה מבטיחה לקידוד אות השארית שהוצגה בשנים האחרונות היא שיטת ה-Multipulse LPC [13]. בשיטה זו אות הערור המשודר למקלט מורכב ממספר קטן של פולסים חשובים, האינפורמציה המשודרת בנוסף לפרמטרי מסנן ה-LPC היא מיקום הפולסים והאמפליטודות שלהם. מציאת מקומות הפולסים וקביעת האמפליטודות נעשית בתהליך איטרטיבי המוצג בציור 3.6:



ציור 3.6 - אנליזת Multipulse

Fig. 3.6 Multipulse Analysis

הערור מורכב כך שמתקבלת שגיאה ריבועית משוקללת ממוצעת מינימלית. שקלול השגיאה נעשה כדי להתחשב בתכונת המיסוך של האוזן.

התהליך האיטרטיבי הוא כדלקמן: בתחילה, ללא כל פולסי ערור, האות המשוחזר נוצר מזכרון מסנן הסינטזה, התרומה של זכרון זה לאות המשוחזר מופחתת מהאות המקורי ונקבעים מקום ואמפליטודה של פולס בודד כך שמקבלים מינימום שגיאה, כעת מחושבת שגיאה חדשה על-ידי הפחתת השפעת הפולס החדש ושוב התהליך חוזר על עצמו עד שנמצאים מספר הפולסים הדרוש או עד שהשגיאה יורדת לערך נמוך שנקבע מראש.

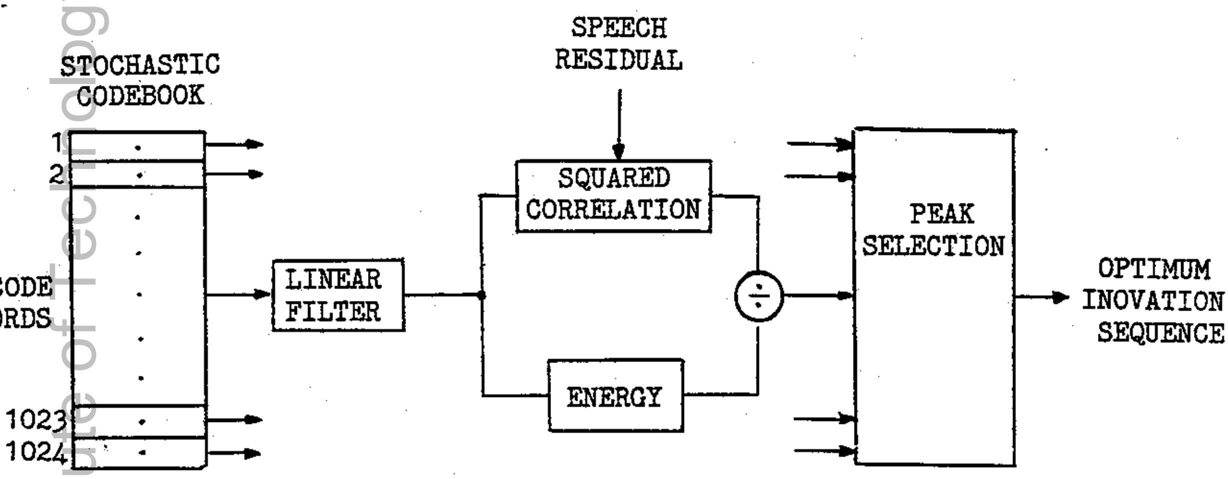
הפוטנציאל בצורת קידוד זו טמון בעובדה שמספיק מספר קטן של פולסים להשגת איכות דיבור טובה - מספיקים כ-4 פולסים במסגרת של 5msec, כלומר, כ-800 פולסים לשניה.

אלגוריתם ה-Multipulse הוא שהוביל לעבודה זו ובפרק הבא נציג אלגוריתם כללי לקידוד אות השארית ששיטת ה-Multipulse היא מקרה פרטי שלו.

3.5 Code Excited Linear Prediction (CELP) [27]

במקודד ה-CELP שנודע גם כן כ-Stochastic Coder הערור האופטימלי נבחר מתוך מילון של סדרות אקראיות לבנות מפולגות גאוסית. הערור נבחר כך שיתקבל מינימום שגיאה ריבועית משוקללת (Subjective error criterion) בהתבסס על תכונת המיסוך של האוזן.

החסרון העיקרי של השיטה הוא בחיפוש המייגע הנדרש על-פני מילון הסדרות האקראיות. סכמת החפוש למציאת הסדרה האקראית האופטימלית מתוארת בציור 3.7 . סכמת חיפוש כזו דומה לסכמת החפוש בה משתמשים ב-Multipulse LPC והיא תהיה ברורה יותר לאחר הדיון ב-Multipulse LPC בפרקים 4 ו-5.



ציור 3.7 - סכמת החפוש למציאת הערור האופטמלי ב-CELP

Fig. 3.7 Search procedure for determining the best stochastic code in CELP

פרק 4. קידוד אות השארית על-ידי פרוק וקטורי

4.1 מבוא

בפרק זה נציע שיטה חדשה לקידוד אות השארית: אות השארית יורכב מקומבינציה ליניארית של מספר קטן של וקטורים הלקוחים מתוך אוסף נתון. האינפורמציה המשודרת לגבי אות השארית תהיה כעת אינדקסי הוקטורים שנבחרו (האוסף ממנו נבחרים הוקטורים נמצא הן במקלט והן במשדר) וערכי המקדמים הכופלים אותם.
נראה שכמה ממקודדי השארית שסקרנו בפרק הקודם הם מקרים פרטיים של צורת קידוד זו. נבחן אלגוריתמים איטרטיביים ולא איטרטיביים למציאת קומבינצית הוקטורים בהנתן אוסף הוקטורים וננסה למצוא אוסף וקטורים אופטימאלי.

בסוף הפרק תוצג סכמה חדשה למקודד הפועל בתחום התדר ומקצה סיביות באופן דינאמי להתמרה של אות השארית - PTC - Predictive Transform Coder.

בפרקים הבאים יובא תאור מדויק של האלגוריתמים שיוצגו בפרק זה, תוצגנה תוצאות סימולציה והערכות סיבוכיות.

4.2 קידוד אות השארית על-ידי פרוק וקטורי

הצגת צורת הקידוד [35].

יהי נתון V אוסף בן M וקטורים מאורך N כל אחד. בהנתן וקטור אות הדיבור $\underline{S}$ ומקדמי מסנן הסינטזה $A_{(z)}$ נרכיב את אות הערור $\underline{U}$ כקומבינציה לינארית של מספר קטן של וקטורים מ-V :

$$(4.2.1) \qquad \underline{U} = \sum_{i=1}^{k} x_i \underline{v}_i \quad , \quad \underline{v}_i \in V \quad , k \ll M$$

הוקטורים $\underline{v}_i$ ייבחרו באופן שיבטיח מינימום שגיאה ריבועית משוקללת בין אות המקור והאות המשוחזר (4.2.2).

קריטריון השגיאה הריבועית המשוקללת הוא:

$$(4.2.2) \qquad E_w = \sum_n [(s_n - \hat{s}_n) * w_n]^2$$

כאשר:

s_n - אות הדבור המקורי

$\hat{s}_n$ - אות הדיבור המשוחזר

w_n - תגובת המסנן המשקלל

$*$ - קונבולוציה.

בסעיף 4.2.1 נסביר בדיוק את שקלול השגיאה.

בצורת קידוד כמו שהצגנו לעיל נשאלות שתי שאלות:

(i) בהנתן האוסף V כיצד תיבחר קומבינצית הווקטורים?

(ii) מהו האוסף V האופטימלי?

צורת קידוד כזאת לאות השארית אטרקטיבית מכיוון שכמות האינפורמציה המשודרת קטנה יחסית:

(i) אינדקסי הווקטורים שנבחרו (האוסף V נמצא במשדר ובמקלט).

(ii) המקדמים הכופלים את הווקטורים x_i, $i=1...k$.

בשנים האחרונות הוצעו כמה מקודדי שארית שהם מקרים פרטיים של צורת הקידוד על-ידי פרוק וקטורי:

1) MULTI-PULSE LPC [13] - במקודד זה נבחר $V = I_N$ מטריצת יחידת מסדר N. הדבר שקול ליצירת אות הערור ממספר קטן של פולסים כאשר מיקומיהם מתאימים לאינדקסי הוקטורים שנבחרו והאמפליטודות שלהם הם $i=1...p$ x_i.
עם בחירה זו של V, החפוש אחר קומבינצית הוקטורים נעשה בצורה איטרטיבית וקטור אחר וקטור.

2) CELP - CODE EXCITED LINEAR PREDICTION [27] - במקודד זה V הינו אוסף וקטורים אקראיים ו-k=1. כלומר, נבחר וקטור יחיד מהאוסף עבור כל מסגרת ומחושב מקדם הגבר יחיד.

3) REGULAR EXCITATION [23] - האוסף V מכיל וקטורי "מסרק" בהם יש 1-ים במרווחים קבועים בפאזה שונה.

e.g.

$$V_1 = 0100000100000100...00100$$

$$V_2 = 0010000010000010...00010$$

וכו'.

ה"מסרק" הנבחר מהאוסף קובע את מיקום הפולסים בערור ויש צורך באיטרציה נוספת לקביעת האמפליטודות של הפולסים האלו.

4) SELF EXCITED VOCODER [33] - במקודד זה מוצע לכלול באוסף V גם ווקטורים אקראיים כמו ב-CELP וגם וקטורי יחידה כמו ב-MPLPC ואליהם להוסיף וקטורים מההסטוריה של הערור, חלק זה של האוסף יהיה דינאמי ויתעדכן כל הזמן.

4.2.1 שקלול אות השגיאה

לפני שנתקוף את שתי השאלות שהוצגו לעיל נקדיש כמה מלים לבחינת קריטריון הטיב בו נשתמש.

מדד השגיאה הריבועית הוא מדד אוביקטיבי שאינו מתחשב בגורמים סובייקטיביים כמו תכונות האוזן. כדי להגדיר מדד טוב יותר להבדל בין האות המקורי והמשוחזר נגדיר שגיאה משוקללת בתדר: (Frequency weighted error):

$$E_w = \int_0^{f_s} |S_{(f)} - \hat{S}_{(f)}|^2 W_{(f)} df \qquad (4.2.3)$$

כאשר $S_{(f)}$ ו-$\hat{S}_{(f)}$ הם התמרות פוריה של האות המקורי והמשוחזר בהתאמה, $W_{(f)}$ היא פונקצית המשקל ו-f_s הוא תדר הדגימה.

בגלל הריכוז הגדול יחסית של אנרגיה באזורי הפורמנטים לעומת האזורים בין פורמנטים בספקטרום אות הדיבור, האוזן יכולה לסבול שגיאה גדולה יותר באזורי הפורמנטים - בוצר מיסוך של הרעש בתדירויות אלה.

נרצה איפוא לבחור $W_{(f)}$ כך שספקטרום השגיאה המתקבלת יהיה צבוע באופן יחסי כלשהו לספקטרום האות המקורי.

המסנן ההפוך $A_{(z)} = 1+P_{(z)}$ נתון על-ידי מקדמיו:

$$A_{(z)} = 1 + \sum_{k=1}^{p} a_k z^{-k} \qquad (4.2.4)$$

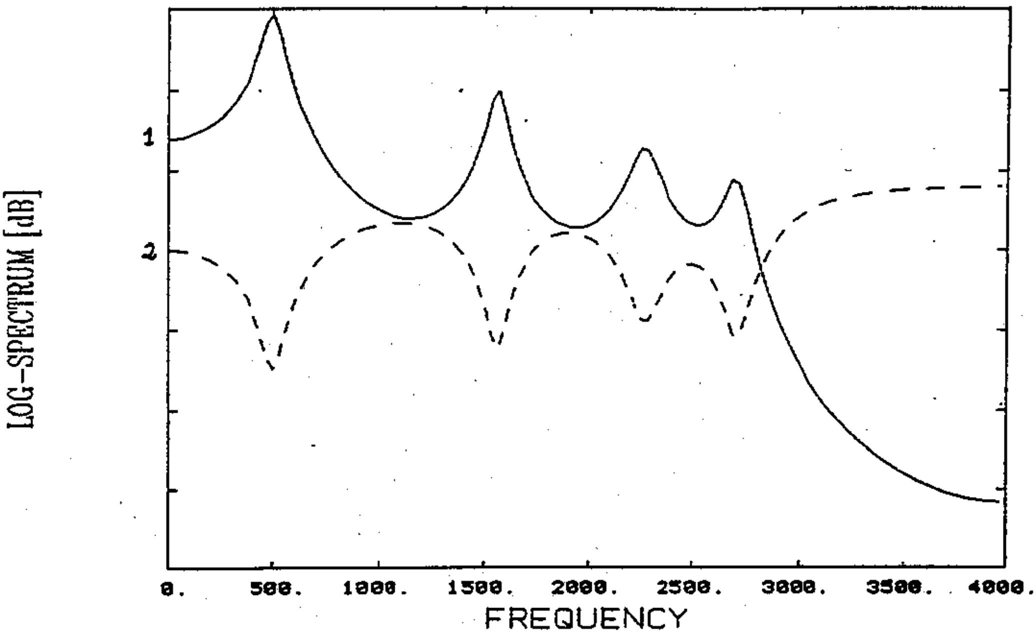
אם נסמן את פונקצית התמסורת של המסנן המשקלל ב-$W_{(z)}$ אזי בחירה נוחה עבור המסנן תהיה [12]:

$$W_{(z)} = \frac{1 + \sum_{k=1}^{p} a_k z^{-k}}{1 + \sum_{k=1}^{p} a_k \gamma^k z^{-k}} \qquad (4.2.5)$$

כאשר המקדם γ, $0 \le \gamma \le 1$, שולט על מידת ההדגשה של השגיאה באזורי הפורמנטים. הבחירה $0<\gamma<1$ היא מקרה ביניים בין שתי בחירות קיצוניות:
הבחירה הראשונה היא לקבוע $\gamma=1$ המביאה ל- $W_{(z)}=1$. ספקטרום השגיאה איננו משוקלל ובצפה לכן לקבל ספקטרום שגיאה לבן. במקרה זה נקבל ערכי SNR גבוהים באזורי הפורמנטים ו-SNR נמוך באזורים שבין הפורמנטים בהם אנרגית האות נמוכה. הבחירה הקיצונית השניה היא $\gamma=0$ המביאה ל- $W_{(z)} = 1 + P_{(z)}$. במקרה זה ספקטרום השגיאה יהיה צבוע לפי ספקטרום האות. כפשרה בין שתי בחירות אלו נבחר $0<\gamma<1$.

כאמור, הערך של γ נקבע לפי הדרגה בה נרצה להדגיש את הרעש באזורי הפורמנטים והערך הסופי צריך להקבע בעזרת מבחני האזנה. נמצא [13] שערך אופיני ל-γ צריך להיות בתחום 0.7-0.8 ואיננו קריטי כל כך.

בציור 4.1 מובאת דוגמא לתגובת התדר של המסנן המשקלל לעומת מעטפת ספקטרום אות הדיבור המקורי עבור $\gamma=0.8$. ובציור 4.2 מובאת דוגמא למעטפת ספקטרום השגיאה המתקבלת לעומת מעטפת ספקטרום האות המקורי עבור $\gamma=0.8$.

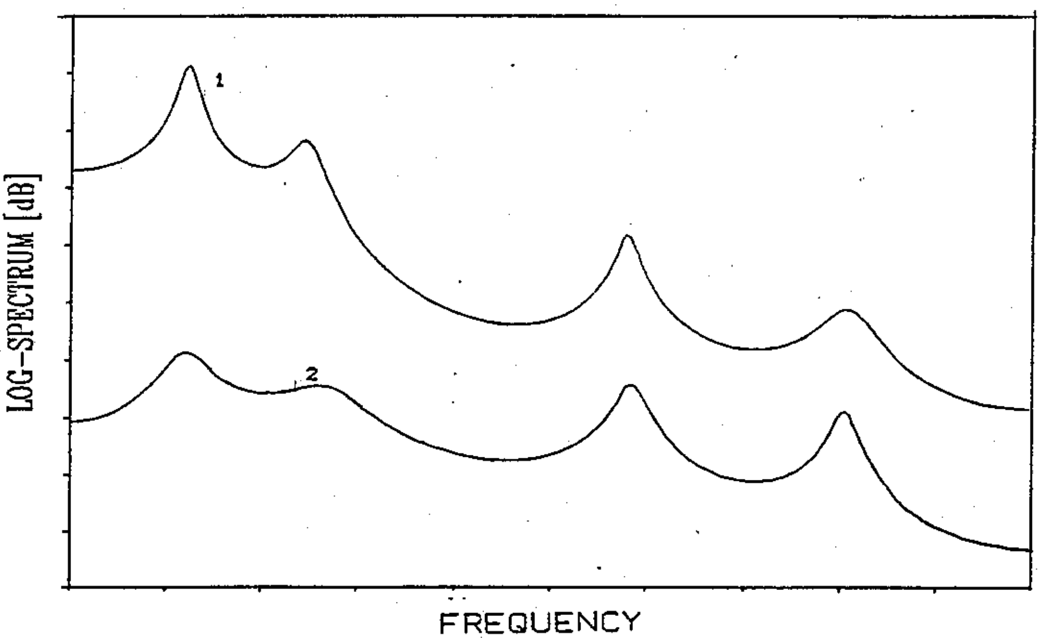


1 - Speech spectral envelope

2 - $W_{(z)}$ Frequency response

ציור 4.1 תגובת המסנן המשקלל לעומת מעטפת ספקטרום אות הדבור

Fig. 4.1 Spectrum of $W_{(z)}$ vs. Spectrum envelope of speech signal



1 - Speech spectral envelope

2 - Error spectral envelope

ציור 4.2 מעטפת ספקטרום השגיאה לעומת מעטפת ספקטרום הדבור

Fig. 4.2 Error spectral envelope vs. speech spectral envelope

4.2.2 מציאת קומבינצית הוקטורים

נתמקד תחילה בשאלת מציאת קומבינצית הוקטורים האופטימלית כאשר האוסף V נתון.
אנו מחפשים מינימום לשגיאה הריבועית המשוקללת שהוגדרה ב- (4.2.2).
אות הדיבור המשוחזר מתקבל על-ידי:

$$\hat{s}_n = u_n * h_n + l_n \tag{4.2.6}$$

כאשר:

u_n - אות הערור
h_n - תגובת מסנן הסינטזה
l_n - אות הנוצר מזכרון המסנן כתוצאה מ"היסטורית" האות (מהמסגרת הקודמת - הקידוד מתבצע במסגרות.

הצבה של (4.2.6) ל-(4.2.2) תתן:

$$\begin{aligned} E_w &= \sum_n [(s_n - u_n * h_n - l_n) * w_n]^2 \\ &= \sum_n [(s'_n - u_n * h_n) * w_n]^2 \end{aligned} \tag{4.2.7}$$

כאשר:

$$s'_n = s_n - l_n \tag{4.2.8}$$

את s'_n ניתן לרשום כ-

$$s'_n = e_n * h_n \tag{4.2.9}$$

כאשר e_n הוא אות "שארית" - הוא מתקבל על-ידי החסרת l_n מ-s_n והעברת האות המתקבל - s'_n דרך המסנן ההפוך $A_{(z)}$ עם זכרון מאופס.

הצבה של (4.2.9) ל-(4.2.7) תתן:

$$(4.2.10)\quad E_w = \sum_n [(e_n - u_n) * h_n * w_n]^2 = \sum_n [(e_n - u_n) * h'_n]^2$$

כאשר h'_n היא תגובת מסנן סינטזה משוקלל:

$$(4.2.11)\quad h'_n = h_n * w_n$$

ומכאן:

$$(4.2.12)\quad H'_{(z)} = H_{(z)}\, W_{(z)} = \frac{1}{A_{(z)}}\;\frac{A_{(z)}}{A_{(z/\gamma)}} = \frac{1}{A_{(z/\gamma)}}$$

נגדיר מטריצה R מימד NxN שהאבר r_{ij} בה הוא:

$$(4.2.13)\quad r_{ij} = \sum_n h'_{n-i}\, h'_{n-j} \qquad 0 \le i \le N-1, \quad 0 \le j \le N-1$$

ונוכל לרשום בצורה מטריצית:

$$(4.2.14)\quad E_w = (\underline{e} - \underline{U})^T R(\underline{e} - \underline{U})$$

כאשר:

$\underline{e}$ - וקטור השארית

$\underline{U}$ - וקטור הערור.

בכוונה לא הגדרנו עד שלב זה את תחום הסכימה בחישוב E_w. אם סכימה זו תיעשה על פני N דגמים נקבל אנליזת "קווריאנס" שבה לא נעשות שום הנחות על האות מחוץ ל-N הדגמים. המטריצה R המתקבלת תהיה מטריצה סימטרית חיובית מוגדרת של ביטוי קווריאנס.

אם גבולות הסכימה יהיו ∞- ל-∞ ונניח ש-h'_n מתאפס החל מנקודה מסוימת נקבל אנליזת "אוטוקורלציה". המטריצה R המתקבלת במקרה זה תהיה מטריצת Toeplitz סימטרית חיובית מוגדרת של אברי אוטוקורלציה. בשלב זה אין לסוג הניתוח השפעה על פיתוח הפתרון אולם בהמשך נראה כיצד סוג האנליזה משפיע על האלגוריתמים המתקבלים.

וקטור הערור נבנה כקומבינציה לינארית של וקטורים מתוך V:

$$(4.2.15) \quad \underline{U} = \sum_{i=1}^{k} x_i \underline{v}_i \qquad \underline{v}_i \in V$$

הצבה של (4.2.15) ל-(4.2.14) תתן:

$$(4.2.16) \quad E_w = (\underline{e} - \sum_{i=1}^{k} x_i \underline{v}_i)^T R(\underline{e} - \sum_{i=1}^{k} x_i \underline{v}_i)$$

נסמן ב-Q מטריצה שעמודותיה הם הוקטורים שנבחרו מ-V ונקבל:

$$(4.2.17) \quad E_w = (\underline{e} - Q\underline{x})^T R(\underline{e} - Q\underline{x})$$

כאשר $\underline{x}$ הוא וקטור המקדמים $x_i,\ i=1...k$.
נניח תחילה שהמטריצה Q נתונה ונמצא פתרון אופטימלי עבוא $\underline{x}$.

$$(4.2.18) \quad \underline{x}^T = [x_1 \ x_2 \ ...x_k]$$

נגזור את E_w ביחס ל-$\underline{x}$ ונשווה ל-$\underline{0}$.

$$(4.2.19) \quad \frac{\partial E_w}{\partial \underline{x}} = -2Q^T R(e - Q\underline{x})$$

ומתקבל:

$$(4.2.20) \quad \underline{x}_{opt} = (Q^T RQ)^{-1} Q^T R\underline{e}$$

כלומר, בהנתן Q - מטריצת הוקטורים שנבחרו מ-V - ניתן לחשב בעזרת (4.2.20) את וקטור המקדמים $\underline{x}$.
הצבה של (4.2.20) ל-(4.2.17) תתן את השגיאה E_w כשידוע שעבור כל Q שתיבחר נבחר $\underline{x}$ האופטימלי.

$$(4.2.21) \quad E_W = \underline{e}^T R \underline{e} - (\underline{e}^T RQ)(Q^T RQ)^{-1}(Q^T R\underline{e})$$

נסמן:

$$(4.2.22) \quad \Delta E = (\underline{e}^T RQ)(Q^T RQ)^{-1}(Q^T R\underline{e})$$

בכל מסגרת ניתן לחשב את R, ומכיוון ש-$\underline{e}$ נתון הבעיה מתמקדת במציאת המטריצה Q כזו שתתן ΔE מקסימלי ומכאן E_W מינימאלי.

נבחן כעת כמה אפשרויות למציאת המטריצה Q המבוקשת מתוך האוסף V:

1) <u>חפוש אופטימאלי</u> - מעבר על כל הקומבינציות האפשריות של k וקטורים מתוך V. זו גישה לא מעשית מכיוון שמספר קומבינציות טיפוסי הוא לדוגמא $\binom{80}{8}$ (בחירת 8 וקטורים מתוך אוסף של 80). ניתן לחסוך בחיפוש על-ידי שימוש באלגוריתמים מהסוג של Branch and Bound [34] לחיפוש על עץ, אולם, גם במקרה זה הסיבוכיות עדיין תהיה גבוהה מאד.

מכיוון שחיפוש אופטימלי כזה אינו מעשי נפנה לאלגריתמים תת-אופטימליים:

2) <u>Independent Selection</u> [55] - בחירת k וקטורים שכל אחד מהם נותן באופן בלתי תלוי את ערכי ΔE הגדולים ביותר. גישה זו פשוטה אולם מביאה לתוצאות גרועות.

3) <u>Sequential Forward Selection (SFS) [55]</u> - בחירה איטרטיבית. בכל שלב מחפשים וקטור יחיד בהתחשב באלה שכבר נבחרו באיטרציות הקודמות. כאן כבר מתחשבים בתלות בין הוקטורים אולם לא ניתן לוותר על וקטורים שכבר נבחרו.

4) <u>Sequential Backward Selection (SBS) [55]</u> - ויתור איטרטיבי על וקטורים. מתחילים עם כל הווקטורים ומוותרים בצורה איטרטיבית על וקטורים עד שנשארים עם ה-k הנבחרים. שיטה זו לא מעשית מכיוון שחישוב וקטור המקדמים $\underline{x}$ מתוך (4.2.20) יהיה מסובך מאד עבור קבוצות וקטורים גדולות.

5) Algorithm (l-r) [55] - זהו אלגוריתם שמשלב את SFS ו-SBS. בכל שלב משתמשים ב-SFS לבחור l וקטורים ואח"כ ב-SBS לוותר על r וקטורים.
אלגוריתם זה מתגבר על הבעיה של SFS - אי יכולת לוותר על וקטורים שכבר נבחרו, אולם, הדבר נעשה תמורת תוספות סיבוכיות.

בסעיפים הבאים נבחן אלגוריתמים ספציפיים למציאת קומבינציות הווקטורים: נציע אפשרויות שונות לבחירת V - בחירות שיובילו לפתרונות פשוטים יחסית ומעשיים ונראה מה יהיו פתרונות אלו. את הדיון נפריד לדיון באלגוריתמים איטרטיביים ולא איטרטיביים.

4.3 אלגוריתמים איטרטיביים למציאת קומבינציית הוקטורים

כדי להגיע לאלגוריתם איטרטיבי פשוט יחסית מהטיפוס של SFS מוצע לבחור $V=I_N$ (I_N - מטריצת יחידה מסדר N). כלומר, הערור מורכב מקומבינציה לינארית של k וקטורי יחידה - וקטורים שרק איבר אחד בתוכם שונה מ-0 וערכו שווה ל-1.
בחירה זו שקולה בעצם למיקום k פולסים בוקטור הערור כאשר וקטור המקדמים $\underline{x}$ מכיל את האמפליטודות שלהם.
גם עם בחירה זו של V חפוש מלא איננו מעשי כך שבחירת הוקטורים נעשית בצורה סדרתית. במהלך האלגוריתם ייבחר כל פעם וקטור אחד מתוך I_N ותחושב אמפליטודה מתאימה. השארית $\underline{e}$ תעודכן כך שנתחשב בוקטור החדש שנבחר ושוב נחזור על התהליך.
נסמן:

$\underline{e}^j$ - וקטור השארית לאחר שנמצאו j פולסים.

$\underline{v}_i$ - וקטור שנבחר מ- $V=I_N$.

בכל שלב באלגוריתם נבחר $Q=\underline{v}_i$ כאשר $\underline{v}_i$ הוא אחד הוקטורים מ-V. לפי (4.2.22) ולאחר שנמצאו j וקטורים (פולסים) יש למצוא i כזה כך ש-:

$$\Delta E_i = (\underline{e}^{jT} R \underline{v}_i) \frac{1}{r_{ii}} (\underline{v}_i^T R \underline{e}^j) \tag{4.3.1.}$$

יהיה מקסימלי.
בנוסף סימונים נוספים:

- $\underline{\alpha} = R\underline{e}^j$

- $\hat{\underline{\alpha}}$ יהיה וקטור ריבועי אברי $\underline{\alpha}$ כלומר:

$$\hat{\alpha}_i = \alpha_i^2 \qquad 0 \le i \le N-1$$

- $diag(1/r_{ii})$ תהיה מטריצה אלכסונית שעל אלכסונה בשורה i בעמודה i נמצא:

$(1/r_{ii})$ עבור $0 \le i \le N-1$.

כפי שנראה, הבחירה $V=I_N$ נוחה מכיוון שכפל וקטור ב-$\underline{v}_i$ פרושו בחירת האבר ה-i מהוקטור. וכפל מטריצה ב-$\underline{v}_i$ פרושו בחירת העמודה ה-i-ית מהמטריצה.

<u>האלגוריתם המתקבל:</u>

1. חישוב הוקטור:

$$\underline{\alpha} = R\underline{e}^0 \tag{4.3.2}$$

2. מתוך $\underline{\alpha}$ חשב את $\hat{\underline{\alpha}}$: $\hat{\alpha}_i = \alpha_i^2, \quad 0 \le i \le N-1$

3. חישוב הוקטור:

$$\underline{\Delta E} = \mathrm{diag}(1/r_{ii})\hat{\underline{\alpha}} \tag{4.3.3}$$

4. מצא איבר מקסימלי ב- $\underline{\Delta E}$, יהי אינדקסו l.

5. חשב את האמפליטודה:

$$x_l = (\underline{v}_l^T R \underline{v}_l)^{-1} (\underline{v}_l^T R \underline{e}^j) = \frac{\alpha_l}{r_{ll}} \tag{4.3.4}$$

6. אם נמצאו k וקטורים סיים.

7. עדכן את השארית $\underline{e}^j$:

$$\underline{e}^{j+1} = \underline{e}^j - x_l \underline{v}_l \tag{4.3.5}$$

למעשה עדכן את $\underline{\alpha}$:

$$\underline{\alpha}^{j+1} = R\underline{e}^{j+1} = \underline{\alpha}^j - x_l R\underline{v}_l \tag{4.3.6}$$

8. חזור ל-2.

בשלב זה ניתן לפצל את הדיון לשניים לפי אופי המטריצה R:

1. אם הסכימה לצורך חישוב R נעשית רק על N דגמים כפי שצריך להעשות אם רוצים לבצע ניתוח מדויק לכל מסגרת בנפרד, כלומר, לא נעשות שום הנחות לגבי האותות הקיימים מחוץ למסגרת, אזי המטריצה R המתקבלת תהיה מטריצה של ביטויי קווריאנס - מטריצה סימטרית חיובית מוגדרת במימד NxN. לסוג אנליזה כזה נקרא אנליזת קווריאנס.

2. אם מניחים שהסכימה בחישוב R היא מ-$-\infty$ ל-∞ תוך הנחה ש-h'_n מתאפס החל מנקודה מסוימת אזי R תהיה מטריצה של אברי אוטוקורלציה - מטריצת Toeplitz סימטרית חיובית מוגדרת ממימד NxN, לסוג אנליזה כזה נקרא אנליזת אוטוקורלציה.

בפרק 5 נציג את האלגוריתמים השונים המתקבלים ביתר פירוט.

מהאלגוריתם האיטרטיבי שהוצג ניתן בשינוי קל להגיע לאלגוריתם ה-Multi-Pulse LPC [13] שהוצע בשנים האחרונות:

אם נצא מ-אנליזת קווריאנס ניתן לפרק את המטריצה R ל-

$$R = F^T F \qquad (4.3.7)$$

כאשר F היא מטריצה NxN שבשורה ה-m-ית והעמודה ה-i-ית מכילה את האיבר h'_{n-i}:

$$F = \begin{bmatrix} h'_0 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ h'_1 & h'_0 & 0 & 0 & \cdot & \cdot & 0 \\ h'_2 & h'_1 & 0 & 0 & 0 & \cdot & 0 \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ \cdot & & & & & & \cdot \\ h'_{N-1} & h'_{N-2} & & \cdot & \cdot & \cdot & h'_0 \end{bmatrix} \qquad (4.3.8)$$

הצבה של (4.3.7) ל-(4.3.2) תתן:

$$\underline{\alpha} = R\underline{e}^j = F^TF\underline{e}^j = F^T\underline{\tilde{S}}^j \quad (4.3.9)$$

כאשר:

$$\underline{\tilde{S}}^j = F\underline{e}^j \quad (4.3.10)$$

כלומר, $\underline{\tilde{S}}^j$ הוא וקטור המתקבל על-ידי העברת וקטור השארית $\underline{e}^j$ דרך מסנן הסינטזה המשוקלל h'_n.

עבור $\underline{\tilde{S}}^j$ j=0 יהיה אות דיבור משוקלל לאחר שהחסרנו ממנו את ה-leftover - $\underline{L}$ הנוצר מזכרון מסנן הסינטזה:

$$\underline{\tilde{S}}^o = F\underline{e}^o = (\underline{S} - \underline{L})_W \quad (4.3.11)$$

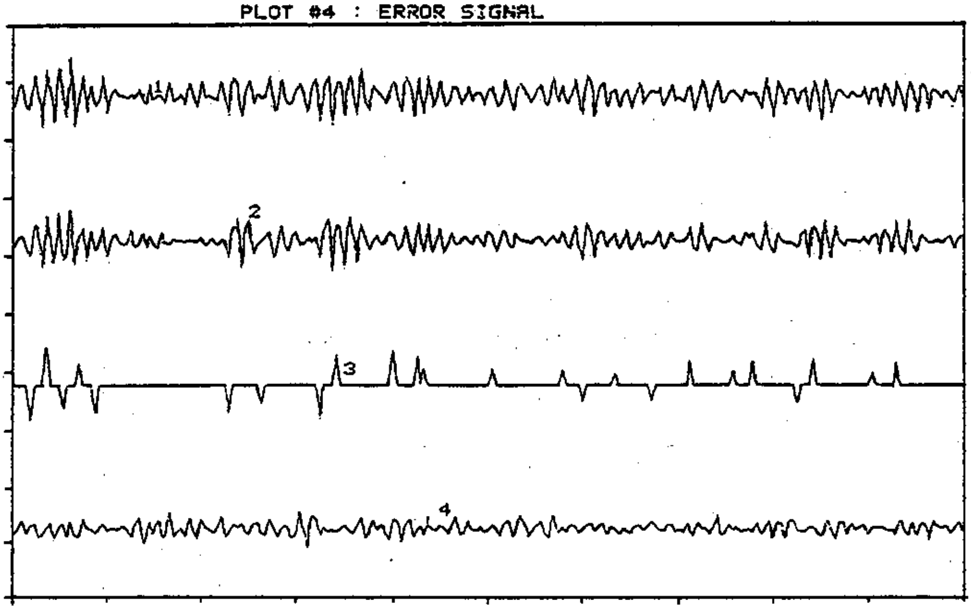
כאשר הסימון $(\ )_W$ מסמן שיקלול בעזרת המסנן המשקלל $W_{(z)}$.

הוקטור $\underline{\alpha}$ מהווה איפוא וקטור של מקדמי Crosscorrelation בין התגובה המשוקללת h'_n ואות הדיבור המשוקלל לאחר הפחתת השפעת זכרון מסנן הסינתזה - $\underline{\tilde{S}}$.

בעזרת (4.3.9) ניתן לחשב את $\underline{\alpha}$ ללא חישוב מוקדם של המטריצה R. צורה זו של אלגוריתם הוצעה במקור ב-[13] Multi-Pulse LPC והצגה דומה הובאה גם ב-[18] [20].

כפי שראינו וכפי שהוצג ב-[13] הבחירה $V = I_N$ ניתנת לפרוש כבחירה של פולסים באות הערור. בציור 4.3 מוצגים: האות המקורי, האות המשוחזר, אות הערור ואות השגיאה הנוצר בין אות המקור והשחזור עבור קטע דיבור קולי ולא-קולי. ניתן לראות שבקטע הקולי הערור מקבל מבנה מחזורי במחזור ה-Pitch ובקטע הלא-קולי הערור מקבל צורה אקראית דמויית רעש.

PLOT #1 : ORIGINAL SPEECH SIGNAL
PLOT #2 : SYNTHETIC SPEECH SIGNAL
PLOT #3 : EXCITATION SIGNAL
PLOT #4 : ERROR SIGNAL



PLOT #1 : ORIGINAL SPEECH SIGNAL
PLOT #2 : SYNTHETIC SPEECH SIGNAL
PLOT #3 : EXCITATION SIGNAL
PLOT #4 : ERROR SIGNAL

ציור 4.3 האותות המתקבלים באלגוריתם האיטרטיבי

Fig. 4.3 Signals obtained by the iterative algorithm

4.3.1 אופטימיזצית אמפליטודות [15]

באלגוריתם שהוצג המקדם של וקטור שנבחר איננו משתנה. יתכן לכן מצב בו נזדקק ליותר פולסים כדי לכפר על אי דיוקים שנכנסו במהלך האיטרציות. ניתן לשפר את האלגוריתם אם בסיום כל מסגרת נבצע חישוב מחדש של האמפליטודות בעזרת (4.2.20):

$$\underline{x} = (Q^T R Q)^{-1} (Q^T R \underline{e})$$

פעולה זו תהיה כרוכה בהיפוך מטריצה או לחליפין פתרון מערכת משוואות ליניאריות.

המטריצה $Q^T RQ$ ידועה - היא מתקבלת על ידי בחירת אברים מתאימים מ-R. הוקטור $Q^T R\underline{e}$

מתקבל על-ידי בחירת אברים מתאימים מהוקטור $\underline{\alpha} = R\underline{e}$ שחושב בתחילת האיטרציות. שיפור נוסף יתקבל אם נבצע אופטימיזצית אמפליטודות לאחר כל איטרציה באלגוריתם. בצורה זו וקטור השגיאה יישמר אורתוגונלי לוקטורי הערור בכל שלבי האלגוריתם ויימנע מצב בו עלול להבחר באחת האיטרציות וקטור שכבר נבחר, כפי שאכן קורה באלגוריתם שהוצג. גם כאן נשתמש ב-(4.2.20) אולם הסיבוכיות המתקבלת תהיה כמובן יותר גדולה.

ניתן לבחור בפתרונות ביניים בין אופטימיזציה יחידה בסוף מסגרת לבין אופטימיזציה לאחר כל איטרציה על-ידי חישוב האמפליטודות מחדש מספר קטן יותר של פעמים במשך המסגרת אך יותר מפעם אחת [22].

4.3.2 שיפורים נוספים שהוצעו בספרות [14],[15],[19],[21]

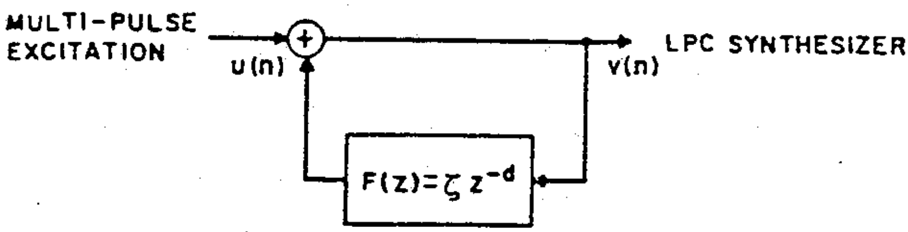
האלגוריתם האיטרטיבי שהצגנו הוא בעצם הצגה שונה של אלגוריתם ה-Multi-Pulse [13] ובשנים האחרונות הוצעו לו שיפורים ותוספות שונים.

נציג כאן כמה מהשיפורים היותר משמעותיים או מעניינים לאלגוריתם:

1. אופטימיזצית מקדמי המסנן בהנתן אות הערור [14] -
באנליזת LPC מניחים שהכניסה למסנן הסינטזה לבנה ומקדמי המסנן מחושבים על-ידי מינימיזציה של שגיאה ריבועית ממוצעת בין האות המקורי והאות המתקבל על-ידי חזאי לינארי מתוך הדגימות הקודמות. ניתוח כזה בדרך כלל לא מניב תוצאות מדויקות עבור אותות מחזוריים (קטעים קוליים).
כדי להתגבר על בעיה זו יש לשערך את הערור למסנן הסינטזה. ב-[14] הוצע שערור ה-Multi-Pulse יהווה שערוך לערור למסנן הסינטזה כדי לקבל מקדמי מסנן סינטזה מדויקים יותר. הנחה זו מובילה לסט משוואות דומות למשוואות נורמליות [14] אולם פתרון מערכת זו איננו אפשרי בצורה שתבטיח תמיד יציבות. ניתן להתגבר על בעיה זו על-ידי בדיקת מיקום הקטבים ושיקוף קטבים שנפלו מחוץ למעגל היחידה לתוכו. אולם, המסנן המתקבל כבר לא יתן מינימום שגיאה ריבועית ממוצעת. בבדיקות שמיעה שערכנו מסתבר שאופטימיזצית המקדמים אכן מביאה לשיפור קל באיכות הדיבור עבור קטעים שמובילים למסנן יציב, אולם, הסיבוכית הנוספת וחוסר היציבות המתקבל מעמידים בספק את כדאיות ביצוע האופטימיזציה.

2. הוספת חזאי Pitch [15] -
מסתבר שכ-8 פולסים דרושים לכל מחזור Pitch כדי לקבל אות משוחזר באיכות גבוהה, כלומר, ככל שתדר ה-Pitch יגדל הקצב בו נצטרך לשדר יגדל אף הוא - לקולות נשים נזדקק לקצב יותר גבוה כדי לשמור על אותה איכות.
כמו כן ניתן להבחין בקורלציה באות הערור ב-Multi-Pulse בין מחזור Pitch אחד למשנהו. ניתן להשתמש בקורלציה זו להקטנת מספר הפולסים על ידי הוספת חזאי Pitch למערכת.

ב-[15] מוצע חזאי פשוט מסדר ראשון:



ציור 4.4 הוספת חזאי pitch למערכת Multi-Pulse

Fig. 4.4 Adding a pitch predictor to a Multi-Pulse system

ζ הוא הגבר החזאי ו-d היא ההשהיה שלו.
השהית החזאי היא מסדר גודל של כמה מחזורי Pitch ובאופן כללי ארוכה מאורך מסגרת - N.
מציאת פרמטרי החזאי וערור ה-Multi-Pulse נעשית בשני שלבים נפרדים - תחילה מניחים שערור ה-Multi-Pulse מאופס ומחשבים את נתוני החזאי ובהנתן החזאי מחשבים את ערור ה-Multi-Pulse.

4.4 אלגוריתמים לא איטרטיביים למציאת קומבינצית הווקטורים

בסעיף הקודם הראנו שהבחירה $V = I_N$ מובילה לאלגוריתם איטרטיבי לבחירת הוקטורים $\underline{v}_i$ (Multi-Pulse) החסרון של האלגוריתמים מהסוג הזה (כפי שנפרט בפרק 5) הוא עדיין סיבוכיות גבוהה. בפרק זה נבחן בחירות אפשריות לאוסף V שיובילו לאלגוריתמים לא איטרטיביים מתוך כוונה להקטין את הסיבוכיות.
בסעיף הבא יוצג אלגוריתם ה-(GMRM) Generalized Maximum Residual Maguitude שהוא אלגוריתם פשוט אך נותן איכות ירודה ומגיעים אליו מתוך בחירה פשוטה של V. בהמשך נראה שניתן להגיע לאלגוריתם לא איטרטיבי נוסף הפועל דווקא בתחום התדר ואליו נגיע מתוך נסיון למצוא את האוסף V האופטימאלי.

4.4.1 Generalized Maximum Residual Maguitude

בחינה של הביטוי לו מחפשים מקסימום באלגוריתם:

$$\Delta E = (\underline{e}^T RQ)\ (Q^T RQ)^{-1} (Q^T R\underline{e})$$

תבהיר שהביטוי $(Q^T RQ)^{-1}$ נכנס כמטריצה משקללת והוא הגורם לצורך לחפש חיפוש מלא מייגע או לחליפין חפוש סדרתי תת אופטימלי.

אם נבחר V כזה שכל Q שתיבחר ממנו תלכסן את $Q^T RQ$ הרי שנוכל לפשט את האלגוריתם מכיוון שבחירת הפולסים במקרה זה תעשה במקביל.

ב-(4.3.7) פירקנו את R באופן הבא:

$$R = F^T F \qquad (4.4.1)$$

נבחר לכן:

$$V = F^{-1} \qquad (4.4.2)$$

Q היא תת מטריצת עמודות של V ולכן:

$$(4.4.3) \quad Q^T R Q = I_k$$

I_k - מטריצת יחידה מסדר k.

נקבל לכן:

$$(4.4.4) \quad \Delta E = (\underline{e}^T R Q)\ (Q^T R \underline{e})$$

כדי למצוא מקסימום לביטוי זה נגדיר וקטור $\underline{Y}$:

$$(4.4.5) \quad \underline{Y} = V^T R \underline{e}$$

ונבחר ממנו את k האברים המקסימאליים בערך מוחלט, האינדקסים שלהם יקבעו איזה וקטורים ייבחרו מ-V וירכיבו את Q. נציב את (4.4.1) ו-(4.4.2) ל-(4.4.5) ונקבל:

$$(4.4.6) \quad \underline{Y} = (F^{-1})^T F^T F \underline{e} = F \underline{e}$$

כלומר, $\underline{Y}$ יתקבל על-ידי העברת $\underline{e}$ דרך מסנן הסינטזה המשוקלל $\dfrac{1}{A_{(z/\gamma)}}$

האמפליטודות יתקבלו על ידי:

$$(4.4.7) \quad \underline{X}_{opt} = (Q^T R Q)^{-1} (Q^T R \underline{e}) = Q^T R \underline{e}$$

כלומר, האמפליטודה המתאימה לאינדקס i שנבחר תהיה בדיוק הערך y_i (האבר ה-i-י בוקטור $\underline{Y}$).
אות הערור ינתן על ידי:

$$(4.4.8) \quad \underline{U} = \sum_{i=1}^{k} x_i \underline{v}_i = Q\ \underline{x} = V \underline{\tilde{Y}}$$

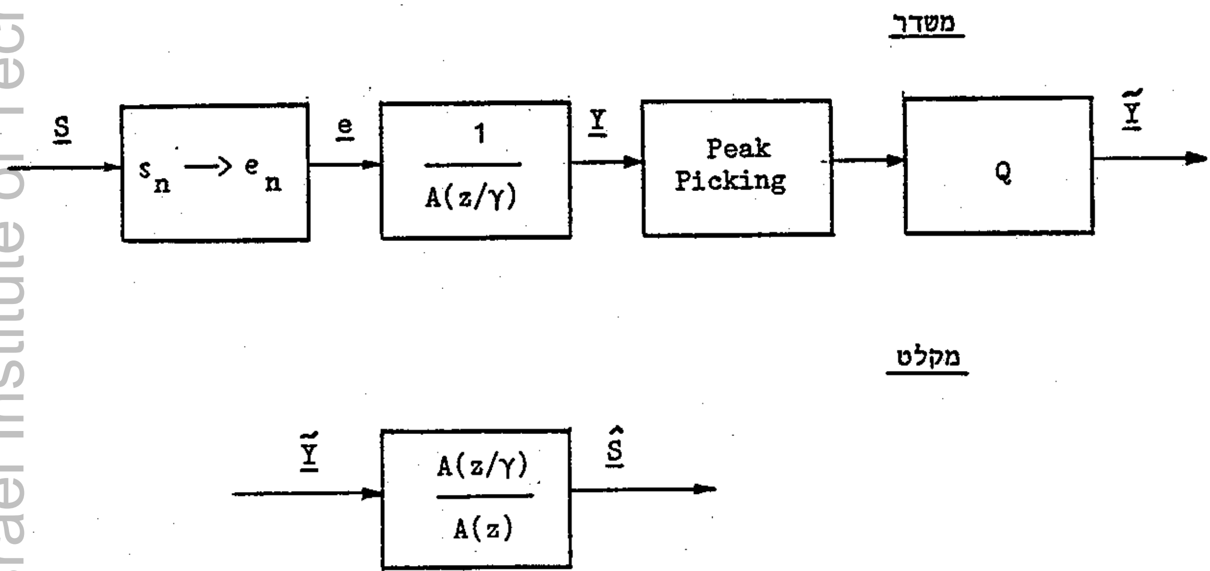
כאשר אברי הוקטור $\underline{\tilde{Y}}$ הם:

$$(4.4.9)\quad \tilde{y}_i = \begin{cases} y_i & \text{אם } i \text{ נבחר} \\ 0 & \text{אחרת} \end{cases}$$

$$i = 0,1,2\ldots N-1$$

בחינה מעמיקה יותר של הבחירה $V=F^{-1}$ תבהיר שבתלות ב-γ אנו נעים בין שני מקרי קצה - עם $\gamma = 0$ נקבל $F=I$ ובחירת השיאים נעשית על אות השארית עצמו. עבור $\gamma = 1$ נקבל שבחירת השיאים נעשית על אות המקור עצמו! בחירת ערך ביניים של γ בין 0 ל-1 מתאימה לבחירת שיאים מאות השארית צבוע לפי $\dfrac{1}{A(z/\gamma)}$.

<u>המערכת המוצעת:</u>



ציור 4.5 - סכמת משדר - מקלט GMRM

Fig. 4.5 GMRM Transmitter - Receiver

<u>הערות:</u>

- הפעולה $s_n \to e_n$ המסומנת בבלוק הראשון במשדר מתבצעת באופן הבא:
מפחיתים מוקטור הדיבור המקורי $\underline{S}$ את אות ה-Leftover מהמסגרת הקודמת:

$$\underline{S}' = \underline{S} - \underline{L} \qquad (4.4.10)$$

את האות המתקבל מעבירים דרך $A_{(z)}$ עם זכרון מאופס לקבלת e_n. ניתן לפשט ולהעביר את $\underline{S}$ ולא את $\underline{S}'$ דרך $A_{(z)}$ הפעם עם זכרון לא מאופס. אולם, במקרה זה אות השארית המתקבל איננו אות השארית הקיים במקלט.

- התהליך במקלט הוא:

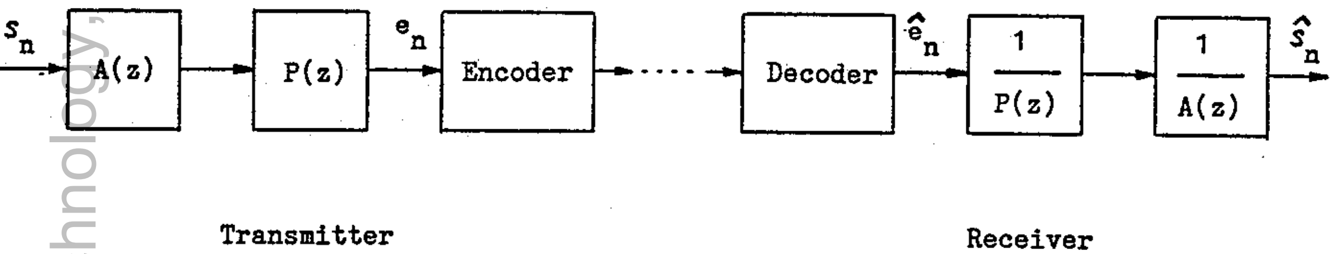
  * קבלת $\underline{U}$ על ידי העברת $\underline{\tilde{Y}}$ דרך $A_{(z/\gamma)}$ (לפי (4.4.8)).
  * העברת $\underline{U}$ דרך מסנן הסינטזה $\dfrac{1}{A_{(z)}}$.

לסכמה זו קראנו: GMRM - Generalized Maximum Residual Magnitude, מכיוון שהיא מקרה כללי של סכמה שהוצעה בספרות [16], [17] ונקראה MRM, בה נבחרו פולסים גדולים בערך מוחלט באות השארית עצמו. MRM מתקבל מ-GMRM כאשר $\gamma = 0$.

אלגוריתם ה-GMRM הוא דוגמא לאלגוריתם לא איטרטיבי, פשוט, אך נמצא בסימולציות כבעל ביצועים גרועים. בסעיף הבא נבחן אלגוריתם לא איטרטיבי נוסף הפועל בתחום התדר והוא כבר נותן תוצאות טובות יותר.

4.5 מציאת אוסף V אופטימאלי

לאחר שהצגנו אפשרויות שונות לבחירת V שהובילו אותנו לפתרונות שונים נשאל את עצמנו מה יהיה V אופטימאלי?
כדי לדבר על אופטימאליות של V נצטרך להניח הנחות סטטיסטיות על e. זאת נוכל לעשות באופן הבא:
נוסיף לסכמה הכללית חזאי Pitch - $P_{(z)}$:



ציור 4.6 הוספת חזאי pitch למערכת
Fig. 4.6 Incorporating a pitch predictor to the system

במקרה זה ניתן להניח שפונקצית צפיפות ההסתברות של דגמי e_n קרובה לגאוסית [11] ושהדגמים בלתי תלוים.
נניח לכך: $\underline{e} \sim N(0,G)$

כאשר G מטריצה אלכסונית בעלת אבר קבוע חיובי על האלכסון.
נשתמש באותם סימונים בהם השתמשנו בסעיפים הקודמים אלא שהפעם הם יסמנו דברים שונים:

$$(4.5.1) \qquad h'_n \overset{z}{\longleftrightarrow} \frac{1}{P_{(z)}\, A_{(z/\gamma)}}$$

$$(4.5.2) \qquad e_n \overset{z}{\longleftrightarrow} S_{(z)}\, A_{(z)}\, P_{(z)}$$

$$(4.5.3) \qquad u_n \overset{z}{\longleftrightarrow} U_{(z)}$$

נניח כעת שהמטריצה Q קבועה ואינה נבחרת בתלות ב-$\underline{e}$, נשאל מהו ה-Q האופטימאלי במקרה כזה.

השגיאה E_w תהיה כעת מ"א ונקבל:

$$(4.5.4) \qquad E(E_w) = G(\text{trace }(R)) - G(\text{trace }(RQ(Q^TRQ)^{-1}Q^TR))$$

כש-$E(\ )$ מסמן תוחלת.

נמצא איפוא Q כזה שעבורו הביטוי הבא יהיה מקסימאלי:

$$(4.5.5) \qquad \Delta E = \text{trace }(RQ(Q^TRQ)^{-1}Q^TR) =$$

$$= \text{trace }((Q^TRQ)^{-1}(Q^TR^2Q))$$

<u>טענה 4.1 (בנספח א'):</u>

Q אופטמאלי הוא כל מטריצה מהצורה

$$(4.5.6) \qquad Q = \hat{Q}\, A$$

כאשר $\hat{Q}$ היא מטריצת הוקטורים העצמיים של R המתאימים ל-k הערכים העצמיים הגדולים ביותר ו-A היא מטריצה kxk לא סינגולרית.

נחזור כעת לבעיה המקורית ונבחר את V להיות אוסף הוקטורים העצמיים של R - $\tilde{V}$.
אם ל-R יש ערכים עצמיים שונים זה מזה אזי:

$$\tilde{V}^T R \tilde{V} = \Lambda \tag{4.5.7}$$

כאשר Λ אלכסונית ומכילה את הערכים עצמיים המתאימים.
אם ל-R ערך עצמי מרובה אזי הוקטורים העצמיים השייכים לו בלתי תלויים וניתן להגיע על-ידי אורתוגונליזציה ל-ו"ע אורתוגונלים כך ש-(4.5.7) עדיין יתקיים [36].

לכל Q, תת מטריצה של $\tilde{V}$, נקבל:

$$Q^T R Q = \Lambda_Q \tag{4.5.8}$$

כאשר Λ_Q מטריצה אלכסונית המכילה על אלכסונה את הערכים העצמיים המתאימים לוקטורים העצמיים של R המרכיבים את Q.

נקבל כעת מתוך (4.2.21):

$$E_w = \underline{e}^T R \underline{e} - (\underline{e}^T R Q)\, \Lambda_Q^{-1}\, (Q^T R \underline{e}) \tag{4.5.9}$$

מכאן נקבל שהוקטורים שעלינו לבחור מ-$\tilde{V}$ יהיו אלו בעלי אינדקסים המתאימים ל-k הערכים הגדולים בערך מוחלט בוקטור:

$$\underline{\alpha} = \Lambda^{-1/2} \tilde{V}^T R \underline{e} = \Lambda^{1/2} \tilde{V}^T \underline{e} \tag{4.5.10}$$

השיוויון השני ב-(4.5.10) נובע מ- $\tilde{V}^T R = \Lambda \tilde{V}^T$ הנובע מ-(4.5.7).

כלומר, נוכל לבצע בחירה סימולטנית של הוקטורים להרכבת הערור על-ידי חישוב הוקטורים העצמיים והערכים העצמיים של המטריצה R, חישוב הוקטור $\underline{\alpha}$ ובחירת האינדקסים המתאימים בו לערכים הגדולים ביותר בערך מוחלט.

מתוך (4.5.10) נובע שמה שמתבצע בעצם הוא פרוק K-L לאות השארית צבוע על ידי:

$$\frac{1}{A_{(z/\gamma)}}$$

הערכים העצמיים הנמצאים על האלכסון של Λ שואפים אסימפטוטית לספקטרום ההספק של h'_n ולכן האלכסון של $\Lambda^{1/2}$ שואף אסימפטוטית לספקטרום של h'_n.

פירוק K-L הוא כמובן מסובך מאד מבחינה חישובית ולכן נקרב אותו על ידי טרנספורם אחר שאינו תלוי באות כמו DFT או DCT.

נקרב בעזרת DFT ונקבל:

$$\underline{\alpha} = \Lambda^{1/2}\tilde{F}\underline{e} \qquad (4.5.11)$$

כאשר $\tilde{F}$ היא מטריצה פוריה.

עלינו למצוא אברים בעלי אמפליטודה מקסימלית ב-$\underline{\alpha}$.

כלומר, למצוא אברים מקסימליים ב-

$$|\underline{\alpha}| = \Lambda^{1/2}|\tilde{F}\underline{e}| \qquad (4.5.12)$$

נראה כיצד ניתן להגיע ל-$|\underline{\alpha}|$:

$$|\tilde{F}\underline{e}| = |DFT(\underline{e})| \quad \Longleftarrow \quad |S_{(z)}|\,|A_{(z)}|\,|P_{(z)}| \qquad (4.5.13)$$

$$\Lambda^{1/2} \simeq |DFT(h'_n)| \quad \Longleftarrow \quad \left|\frac{1}{P_{(z)}}\right|\,\left|\frac{1}{A_{(z/\gamma)}}\right| \qquad (4.5.14)$$

ובסה"כ:

$$(4.5.15) \qquad |\underline{\alpha}| \quad \Longleftarrow \quad \frac{|S_{(z)}|\,|A_{(z)}|}{|A_{(z/\gamma)}|}$$

הקירוב בעזרת DFT הוא בעצם בחירת V להיות $\tilde{F}^*$ כאשר $\tilde{F}$ היא מטריצה פוריה ו-$\tilde{F}^*$ היא מטריצה Transpose Conjugate של $\tilde{F}$:

$$(4.5.16) \qquad \tilde{F} = \begin{bmatrix} W^0 & W^0 & W^0 & . & W^0 \\ W^0 & W^{-1} & W^{-2} & . & . \\ W^0 & W^{-2} & W^{-4} & . & . \\ . & . & . & . & . \\ W^0 & . & . & . & W^{-1} \end{bmatrix} \qquad W = e^{j2\pi/N}$$

באופן אסימפטוטי הוקטורים העצמיים של המטריצה R שואפים לעמודות מטריצת פוריה אך מכיון ש-R סימטרית הערכים העצמיים שלה ממשיים ולכן גם עמודות $\tilde{F}^*$ יהיו וקטורים עצמיים.

המטריצה Q תהיה איפוא תת מטריצה של $\tilde{F}^*$ והמקדמים הכופלים יהיו:

$$(4.5.17) \qquad \underline{x} = (Q^*RQ)^{-1}Q^*R\underline{e} \simeq$$

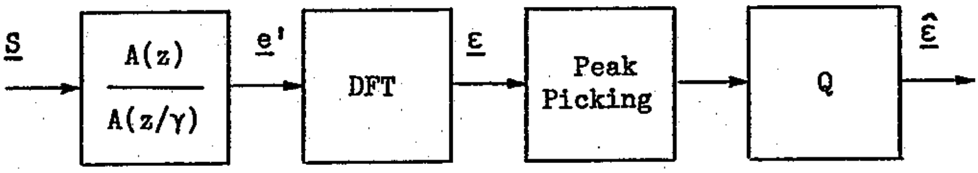
$$\simeq \Lambda^{-1}\Lambda Q^*\underline{e} = Q^*\underline{e}$$

פעולת * - Transpose Conjugate מחליפה כאן את פעולת ה-Transpose מכיון שהגדלים מרוכבים.

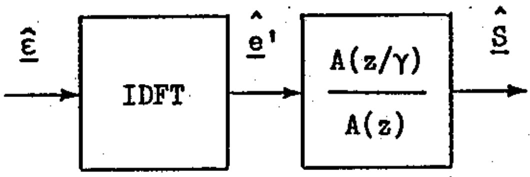
מכיון ש-Q תת מטריצה מ-$\tilde{F}^*$ נקבל שהמקדמים הכופלים הם האיברים הנבחרים מ-DFT($\underline{e}$).

המערכת המוצעת

משדר



מקלט



ציור 4.7 סכמת משדר-מקלט - Residual DFT

Fig. 4.7 Transmitter-Receiver Residual DFT

הערות:

- במשדר מחושב אות השארית הצבוע $\underline{e}'$ עליו מבצעים DFT לקבלת $\underline{\varepsilon}$:

$$|\varepsilon| = |\alpha| \qquad (4.5.18)$$

לכן ה-Peak Picking נעשה על $|\varepsilon|$.

- למקלט משודרים $k/2$ ערכי $R_e(\varepsilon)$ ו-$I_m(\varepsilon)$ סה"כ k ערכים ממשיים (ההתמרה סימטרית לכן דרושים $k/2$ מקדמים מרוכבים).
- במקלט מבצעים התמרה הפוכה לקבלת $\hat{\underline{e}}'$ - אות שארית צבוע משוחזר ואותו מעבירים דרך המסנן:

$$\frac{A_{(z/\gamma)}}{A_{(z)}}$$

(ההעברה דרך $A_{(z/\gamma)}$ דרושה מכיוון שהאמפליטודות של הוקטורים שוות לאברים הנבחרים מ-$DFT(\underline{e})$ ולא $DFT(\underline{e}')$).

בסכמה המוצעת ניתן להחליף את ה-DFT ב-DCT [37] ונצפה לקבל תוצאות טובות יותר - ה-DCT זהה אסימפטוטית ל-KLT עבור תהליך מרקובי מסדר כלשהו [53]. בפרט עבור תהליך מרקובי מסדר I ה-DCT מתכנס מהר יותר ל-KLT לעומת ה-DFT [52]. התהליך במקרה שלנו איננו מרקובי מסדר I אולם המטריצה R המתקבלת קרובה למטריצת אוטוקורלציה המתקבלת בתהליך מרקובי מסדר ראשון.
השימוש ב-DCT יהיה גם נוח יותר מכיוון שכל הגדלים איתם נעסוק יהיו ממשיים ולא יהיה צורך בעבודה עם גדלים מרוכבים.
ערכי ה-SNR המתקבלים כאשר משתמשים ב-DCT מבטיחים, אולם מבחינה סוביקטיבית מופיעה בעיה שהופכת את ביצועי האלגוריתם ללא קבילים:
עקב הבחירה של מספר סובים בודדים בתדר והשתנות הבחירה ממסגרת למסגרת מופיעה תופעה של "צלצולים" באות המשוחזר.
נסיון להתגבר על בעיה זו נעשה באמצעות הקצאה דינמית של סיביות ליותר רכיבי תדר בדומה לקידוד ATC [2]. דיון מפורט בסכמה הזו יובא בפרק 6. בשלב זה נסתפק בהצגה עקרונית של המערכת.

Predictive Transform Coder (PTC) 4.6

כדי למנוע את תופעת ה"צלצולים" נחלק את הסיביות הנתונות לנו לצורך הקידוד באופן שיותר רכיבי תדר יקבלו הקצאה [51]. בהתאם למספר הסיביות שיקבל כל רכיב תדר תתבצע קוונטיזציה ולמקלט ישודרו - האינפורמציה על ההתמרה ואינפורמצית צד בעזרתה יחשב המקלט את הקצאת הסיביות. כפי שנראה במקרה שלנו אינפורמצית הצד תהיה מקדמי מסנן הסינטזה עצמם וההקצאה תיעשה לפי מעטפת משוקללת של אות הדיבור. המוטיבציה לבצע את ההקצאה לפי מעטפת משוקללת של אות הדיבור נובעת מכך שבחירת השיאים בסכמה שהוצעה בסעיף הקודם נעשית מתוך שארית צבועה לפי $\frac{1}{A_{(z/\gamma)}}$ לכן גם את ההקצאה נבצע

לפי מעטפת מהצורה $\frac{1}{A_{(z/\gamma)}}$.

נשתמש כאן בשני ערכי γ:

γ_1 - עבור צביעת אות השארית.

γ_2 - עבור מציאת המעטפת לצורך הקצאת הסיביות.

הסכמה המוצעת תהיה לכך:

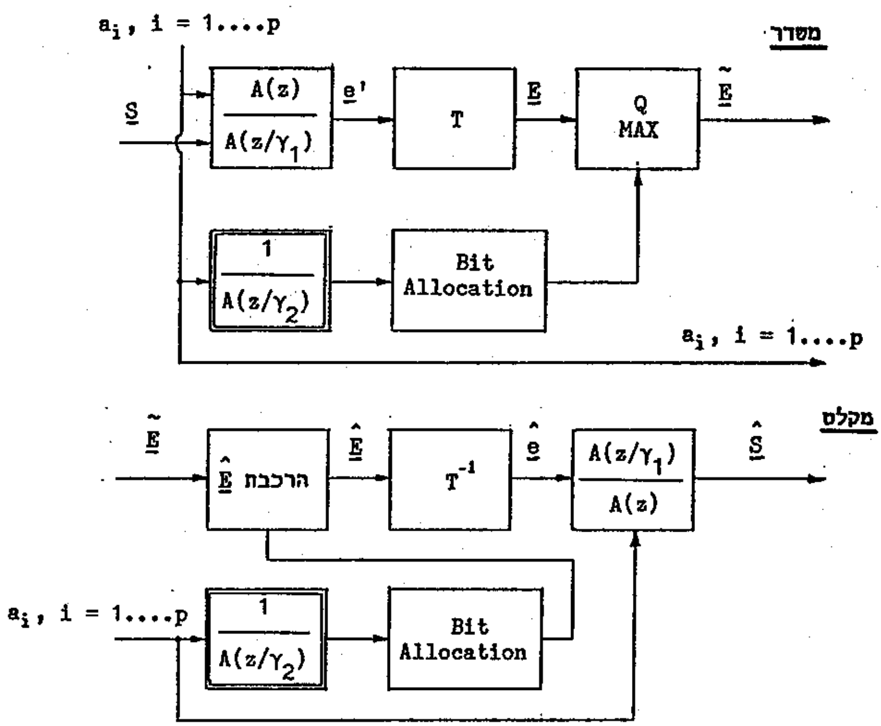
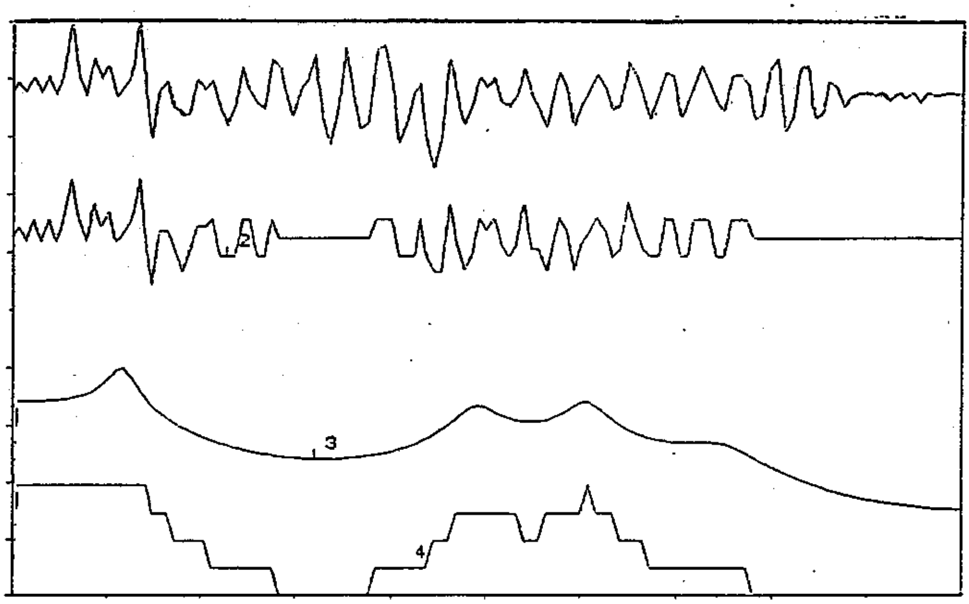


Fig. 4.8 PTC VOCODER - 4.8 ציור



PLOT #1 : ORIGINAL DCT TRANSFORM
PLOT #2 : RESTORED DCT TRANSFORM
PLOT #3 : SPECTRAL ENVELOPE
PLOT #4 : BIT ALLOCATION

ציור 4.9 - דוגמא להתמרת השארית והתמרה משוחזרת

Fig. 4.9 Example for Residual Transform and Reconstaucted Transform

הערות

- גם כאן אות השארית לא נמצא פשוט על-ידי העברה של $\underline{S}$ דרך $A_{(z)}$ אלא על-ידי חישוב ה-Leftover, הפחתתו מ-$\underline{S}$ והעברה דרך $A_{(z)}$ עם זכרון מאופס כפי שעשינו עבור האלגוריתם האיטרטיבי.

- T היא אחת הטרנספורמציות - DFT או DCT.

בעזרת מבחני האזנה נמצאו γ_1 ו-γ_2, האופטימליים - מתקבל שיש לבחור $\gamma_1 = 0$ ו-$\gamma_2=1$-, כלומר, אות השארית המותמר לא עובר כלל צביעה והקצאת הסיביות נעשית לפי המעטפת המקורית של אות הדיבור ללא שיקלול היינו לפי $1/A_{(z)}$.

בחירה זו של ערכי γ היא גם נוחה מכיון שהיא פשוטה - אין צורך לחשב את $\frac{1}{A(z/\gamma)}$ ולצבוע את אות השארית.

השימוש ב-DCT נותן תוצאות טובות יותר, אוביקטיביות וסוביקטיביות, לעומת השימוש ב-DFT.
בציור 4.9 מובאים טרנספורמצית DCT של אות השארית המקורי והמשוחזר וכמו כן מעטפת האות המקורי והקצאת הסיביות המתקבלת על פיו.
בפרק 6 נביא דיון מפורט יותר באלגוריתם ה-PTC נציג תוצאות סימולציה ונעריך את הסיבוכיות הכרוכה במימושו.

פרק 5: אלגוריתמים איטרטיביים לקידוד אות השארית

5.1 מבוא

כפי שראינו בפרק 4, הבחירה $V = I_N$ ושימוש באלגוריתם איטרטיבי מטיפוס (SFS) Sequential Forward Selection מובילים לפתרון איטרטיבי לא מסובך מדי למציאת קומבינצית הוקטורים. ראינו כמו כן כי בתלות בתחום הסכימה בחישוב המטריצה R ניתן להגיע לאנליזת "קווריאנס" או אנליזת "אוטוקורלציה" ומאנליזת ה"קווריאנס" הגענו גם לאלגוריתם ה-Multi-Pulse המקורי. בפרק זה נדון בצורות המימוש השונות לפתרון איטרטיבי, ונציג תוצאות והערכות סיבוכיות לצורות האנליזה השונות. לפני שנעשה זאת, נציג בסעיף הבא פישוט אפשרי למטריצה R, שיפור זה מקטין בהרבה את סיבוכיות הפתרון האיטרטיבי.

5.2 פישוט המטריצה R

לפי (4.2.13) האיבר הכללי ב-R מחושב על-ידי:

$$r_{ij} = \sum_n h'_{n-i}\, h'_{n-j} \qquad 0 \le i \le N-1,\quad 0 \le j \le N-1 \tag{5.2.1}$$

כאשר תחום הסכימה קובע את סוג האנליזה.

h'_n מתקבל על-ידי:

$$h'_n = h_n * w_n \tag{5.2.2}$$

כאשר:

h_n - תגובת מסנן הסינטזה.

w_n - תגובת המסנן המשקלל.

נקבל:

$$H'_{(z)} = H_{(z)}W_{(z)} = \frac{1}{A(z)}\,\frac{A_{(z)}}{A(z/\gamma)} = \frac{1}{A(z/\gamma)} \qquad (5.2.3)$$

ומכאן:

$$h_n' = h_n \cdot \gamma^n \qquad (5.2.4)$$

h_n היא תגובת מסנן יציב ולכן דועכת בזמן, $\gamma<1$ ולכן h_n' דועך מהר יותר מ-h_n.
בציור 5.1 מושווה h_n אופייני ל-h_n' מתקבל כאשר $\gamma=0.8$.

על סמך הדעיכה של h_n' הבעיה תהיה פשוטה יותר אם נניח:

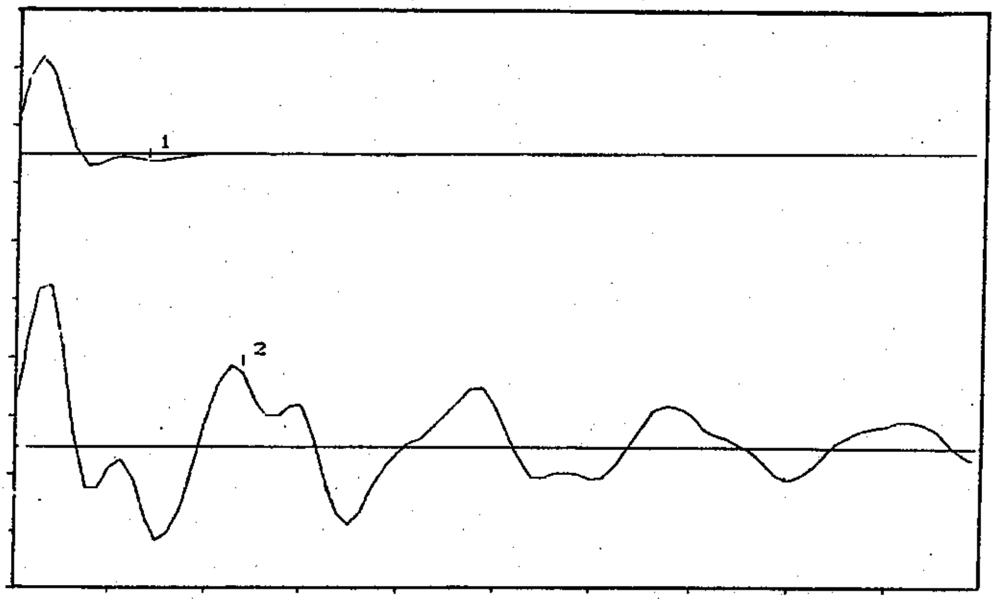
$$h_n' = 0 \qquad \forall\, n \geq L\ ,\ L < N \qquad (5.2.5)$$

כלומר, נניח שמנקודה מסוימת התגובה להלם h_n' מתאפסת.

מ-(5.2.1) ו-(5.2.5) נקבל:

$$r_{ij} = \begin{cases} \sum\limits_n h'_{n-i}h'_{n-j} & |i-j| \leq L-1 \\ 0 & \text{אחרת} \end{cases} \qquad (5.2.6)$$

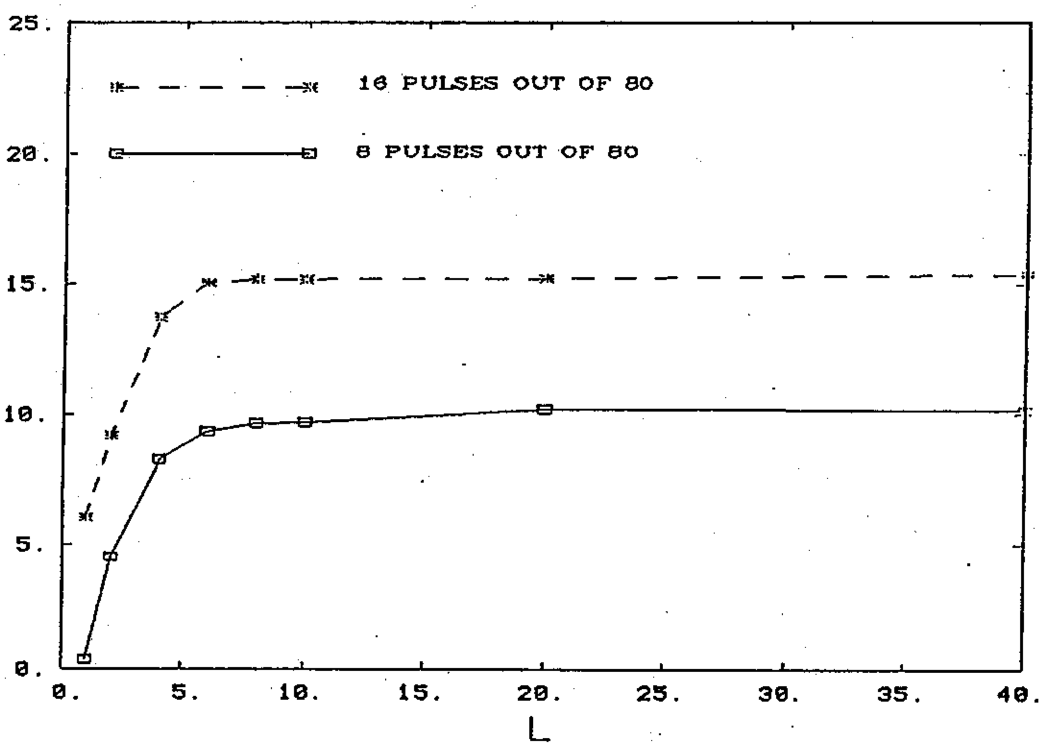
כלומר, R המתקבלת תהיה מטריצת פס ונקבל סיבוכיות חישוב ומקום קטנים יותר.
תוצאות סימולציות מראות כי מספיקים ערכי L נמוכים (L=10) עבור $\gamma=0.8$ כדי לקבל אותן תוצאות כפי שמתקבלות כש-R איננה מטריצת פס, בציור 5.2 מוצג ה-SNR המתקבל עבור ערכי L שונים בשני מקרים: כשנבחרים 8 וקטורים (פולסים) במסגרת של 80 וכשנבחרים 16 וקטורים (פולסים) באותו גודל מסגרת, כאמור, כל התוצאות הן עבור $\gamma=0.8$.



1-h'$_n$, 2-h$_n$

ציור 5.1 h_n אופיני לעומת h'_n אופיני כאשר $\gamma=0.8$

Fig. 5.1 typical h_n vs. typical h'_n when $\gamma=0.8$



ציור 5.2 השפעת L על ה-SNR עבור $\gamma=0.8$

Fig. 5.2 SNR vs. L for $\gamma=0.8$

5.3 אנליזת קווריאנס

צורת האנליזה בסעיף זה היא שתוביל אותנו לבסוף לאלגוריתם ה-Multi-Puse המקורי [13].

באנליזה קווריאנס הסכימה בחישוב R היא על פני N דגמים:

$$r_{ij} = \sum_{n=0}^{N-1} h'_{n-i}\, h'_{n-j} \qquad 0 \le i \le N-1, \quad 0 \le j \le N-1 \tag{5.3.1}$$

המטריצה R המתקבלת היא מטריצת אברי קווריאנס סימטרית (לא מטריצת Toeplitz).
במקרה זה:

$$r_{ii} \ne r_{jj} \quad \text{for } i \ne j \tag{5.3.2}$$

<u>האלגוריתם המתקבל:</u>

<u>נתונים:</u>

N - גודל מסגרת הדיבור

P - סדר מסנן ה-LPC

k - מספר הוקטורים (פולסים) למסגרת

<u>S</u> - וקטור הדבור המקורי

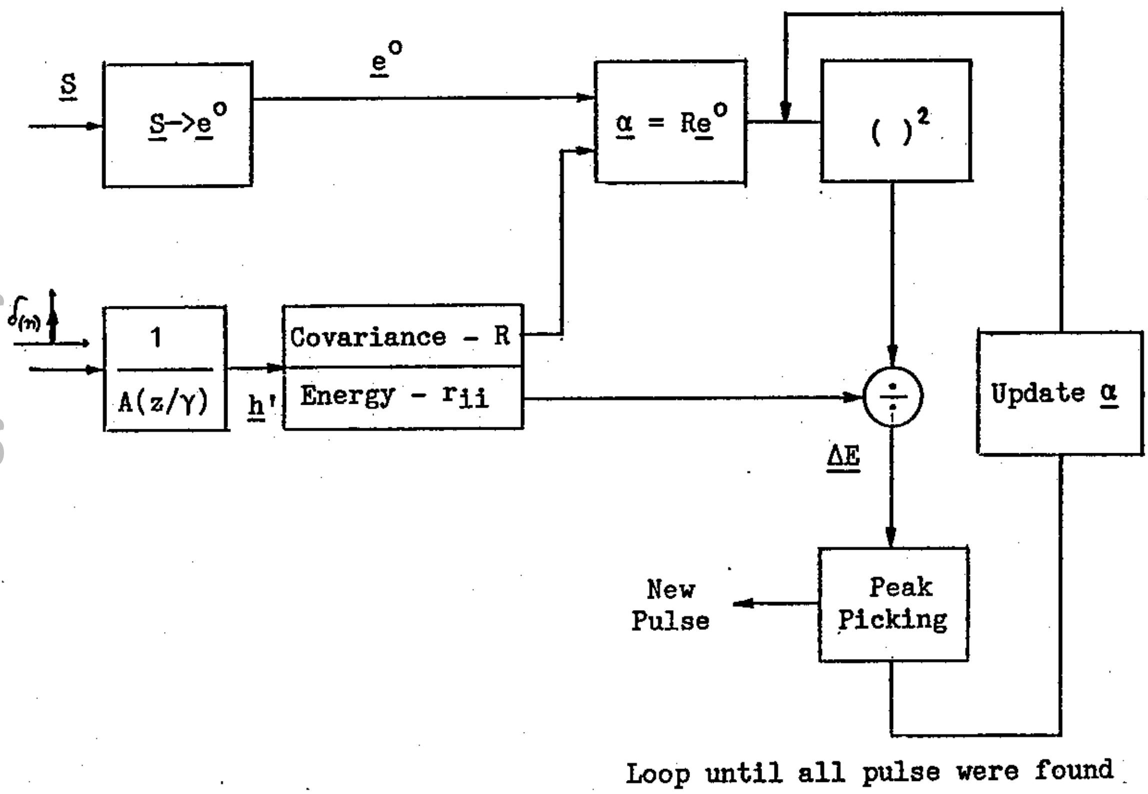
<u>a</u> - וקטור מקדמי מסנן ה-LPC, נתונים משלב של אנליזת LPC

L - מספר הדגמים השונים מ-0 ב-h'_n.

<u>אלגוריתם I</u>:

1. חשב את h'_n, תגובת המסנן המשוקלל $\frac{1}{A(z/\gamma)}$.
2. חשב את מטריצת הקווריאנס R.
3. חשב את הוקטור $\underline{\Psi}$: $\Psi_i = \frac{1}{r_{ii}}$ $\quad 0 \le i \le N-1$.
4. חשב את: $\underline{e}^o$, אות השארית ההתחלתי.
5. חשב את: $\underline{\alpha} = R\underline{e}^o$ ומתוכו גם את $\hat{\underline{\alpha}}$: $\quad 0 \le i \le N-1, \quad \hat{\alpha}_i = \alpha_i^2$
6. חשב את: $\underline{\Delta E} = \text{diag}\,(1/r_{ii})\hat{\underline{\alpha}}$.
7. מצא איבר מקסימלי ב-$\underline{\Delta E}$, יהי אינדקסו l.
8. חשב את האמפליטודה המתאימה: $x_l = \frac{\alpha_l}{r_{ll}}$
9. אם נמצאו k פולסים חשב את האות המשוחזר $\hat{\underline{S}}$ לצורך חישוב ה-Leftover במסגרת הבאה וסיים. אחרת:
10. עדכן את $\underline{\alpha}$: $\quad \underline{\alpha} = \underline{\alpha} - x_l R\underline{v}_l$, חשב את $\hat{\underline{\alpha}}$.
11. חזור ל-6.

בציור 5.3 מוצגת סכמת בלוקים של אלגוריתם I.



ציור 5.3 סכמת בלוקים - אנליזת "קווריאנס", אלגוריתם I

Fig. 5.3 Block scheme - Algorithm I, Covariance Analysis

סיבוכיות אלגוריתם I

בנספח ג' מובאות בצורה מפורטת המלצות למימוש האלגוריתם ומתוך המימוש מחושבת הסיבוכיות. בפרק זה נביא רק תוצאות סופיות לגבי כל האלגוריתמים שנציג.

הסיבוכיות המתקבלת באלגוריתם I היא:

פעולות למסגרת:
$$O\left(P(L+1) + \frac{L(L+1)}{2} + N(2P+1) + 2NL + k(2N+6L) + NP\right)$$

+

פעולות חילוק למסגרת: $O(L)$

לדוגמא:

עבור גודל מסגרת N=80, סדר מסנן P=10 ו-L=10 נקבל:

<u>k=8 (מספר הפולסים למסגרת)</u>:

6005 פעולות למסגרת שהם 76 פעולות לדגם

+

10 פעולות חילוק למסגרת.

<u>k=16</u>

7765 פעולות למסגרת שהם 98 פעולות לדגם

+

10 פעולות חילוק למסגרת

כאמור בנספח ג', הסיבוכיות הזאת היא רק עבור חישוב מקומות ואמפליטודות הפולסים ללא חישוב מקדמי ה-LPC.

סימולציה לאלגוריתם I

הסימולציות שיובאו להלן ובהמשך העבודה נערכו על 3 משפטים שונים באנגלית שנאמרו כל אחד על-ידי גבר וע"י אשה (סה"כ 6 משפטים נבדקים). סדר מודל ה-LPC נלקח תמיד להיות P=10 ושיטת האנליזה לחישוב מקדמי ה-LPC היא שיטת האוטוקורלציה. נערכו גם

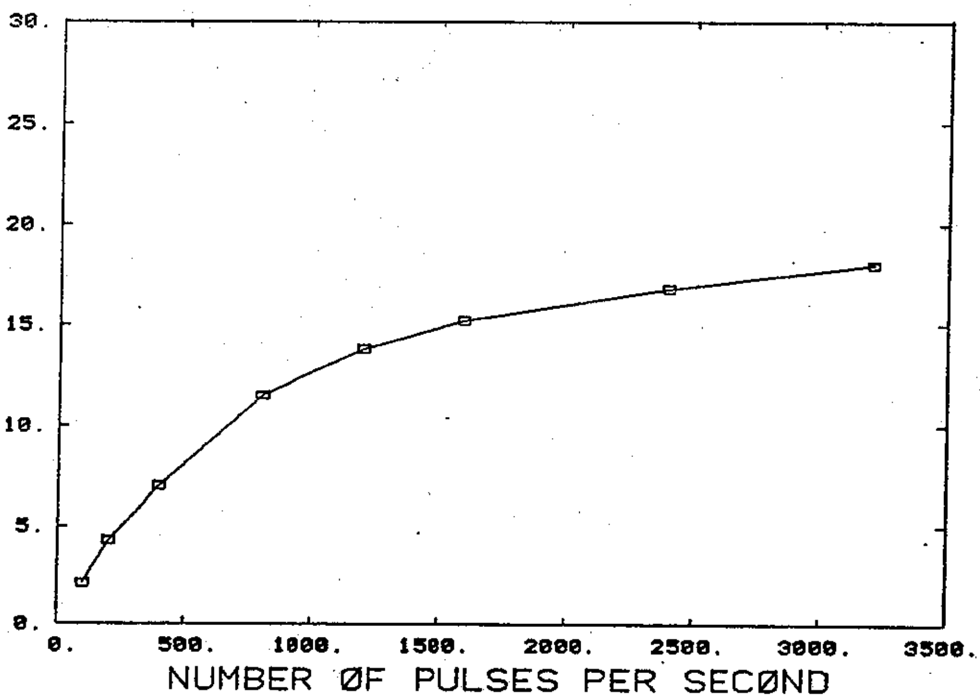
סימולציות כאשר האנליזה נעשתה בשיטת ה-Stabilized Covariance [12] והתוצאות המתקבלות דומות.

תדר הדגימה הוא 8 Khz, רוחב מלת דגימה - 12 bit, סינון מוקדם - 200-3200 hz.

בציור 5.4 מתואר ה-SNR המתקבל כפונקציה של מספר הפולסים בערור, גודל מסגרת האנליזה באלגוריתם היה 80 דגימות, גודל צעד עדכון מקדמי ה- LPC - 160 דגימות, L=10 ו-γ=0.8, קוונטיזצית מקדמי המסנן נעשית לפי הסכמה של LPC-10 (נספח ב'). בנוסף לבדיקות האוביקטיביות נערכו גם מבחני השמעה והתקבל שדי בכ-800 פולסים לשניה בכדי לקבל איכות דיבור טובה.

האלגוריתם האחרון סובל מכמה חסרונות:

1. נדרשות L פעולות חילוק למסגרת.
2. עבור L גדול נקבל סיבוכיות גבוהה במקום.
3. בגלל המבנה המורכב של R ו-$\underline{\psi}$ יש צורך במבנה נתונים וניהול מבנה נתונים לא פשוטים, אלא אם כן מוכנים להקצות הרבה מקום להחזקת R ו-$\underline{\psi}$ באופן מלא.



ציור 5.4 ה-SNR המתקבל באלגוריתם I כפונקציה של מספר הפולסים

Fig. 5.4 SNR vs Number of Pulses, algorithm I

5.3.1 אלגוריתם ה-Multi-Pulse המקורי [13] כמקרה פרטי

מתוך אנליזת הקווריאנס שהוצגה לעיל ניתן להגיע לאלגוריתם ה-Multi-Pulse המקורי שהוצע על-ידי Atal & Remde [13] כמקרה פרטי, נחזור בקצרה על הניתוח האנליטי שהוצג בפרק 4:

המטריצה R ניתנת לפרוק ל-

$$(5.3.3) \qquad R = F^T F$$

כאשר F היא מטריצה NxN משולשת תחתונה הכוללת בשורה ה-n-ית ובעמודה ה-i-ית את האבר h'_{n-i}:

$$(5.3.4) \qquad F = \begin{bmatrix} h'_0 & & & & & & \\ h'_1 & h'_0 & & & 0 & & \\ h'_2 & h'_1 & h'_0 & & & & \\ \cdot & \cdot & \cdot & \cdot & & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \\ h'_{N-1} & h'_{N-2} & \cdot & \cdot & \cdot & \cdot & h'_0 \end{bmatrix}$$

נראה כיצד ניתן להשתמש ב-(5.3.3) באלגוריתם:

הצבה של (5.3.3) ל-(4.3.2) תתן:

$$(5.3.5) \qquad \underline{\alpha} = R\underline{e}^j = F^T F \underline{e}^j = F^T \underline{\tilde{S}}^j$$

כאשר:

$$(5.3.6) \qquad \underline{\tilde{S}}^j = F\underline{e}^j$$

$\underline{\tilde{S}}^j$ הוא וקטור המתקבל על-ידי העברת וקטור השארית $\underline{e}^j$ דרך מסנן הסינטזה המשוקלל - בעל התגובה להלם h'_n.

לכן, עבור $j=0$, $\underline{\tilde{S}}^j$ יהיה אות הדיבור משוקלל ללא התחשבות בזכרון מסנן הסינטזה:

$$(5.3.7) \quad \underline{\tilde{S}} = F\underline{e}^o = (\underline{S} - \underline{L})_w$$

כאשר:

$\underline{S}$ - וקטור אות הדיבור המקורי

$\underline{L}$ - וקטור האות הנוצר כתוצאה מזכרון המסנן (leftover)

$(\ )_w$ - שקלול באמצעות המסנן $\frac{A_{(z)}}{A(z/\gamma)}$.

הוקטור $\underline{\alpha}$ מהווה איפוא וקטור של מקדמי Crosscorrelation בין התגובה המשוקללת h'_n ואות הדיבור המשוקלל לאחר הפחתת השפעת זכרון המסנן - $\underline{\tilde{S}}$.

<u>האלגוריתם המתקבל:</u>

<u>נתונים:</u>

N - גודל מסגרת דיבור,

P - סדר מסנן ה-LPC,

k - מספר הוקטורים (פולסים) למסגרת,

$\underline{S}$ - וקטור הדבור המקורי,

$\underline{a}$ - וקטור מקדמי ה-LPC, נתונים משלב אנליזת ה-LPC,

L - מספר האברים השונים מ-0 ב-h'_n.

אלגוריתם II

1. חשב את h'_n - תגובת המסנן המשוקלל $\frac{1}{A(z/\gamma)}$

2. חשב את $\underline{\tilde{S}}$.

3. חשב את $\underline{\alpha}$ לפי $\underline{\alpha} = F^T\underline{\tilde{S}}$, ומתוכו את $\underline{\hat{\alpha}}$: $\hat{\alpha}_i = \alpha_i^2 \quad ,0 \le i \le N-1$

4. חשב את $\underline{\Psi}$ - $\Psi_i = \frac{1}{r_{ii}}$, $i=0\ldots N-1$

5. חשב את $\underline{\Delta E} = \text{diag}\ (1/r_{ii})\ \underline{\hat{\alpha}}$

6. מצא את האבר המקסימלי ב-$\underline{\Delta E}$, יהי אינדקסו l.

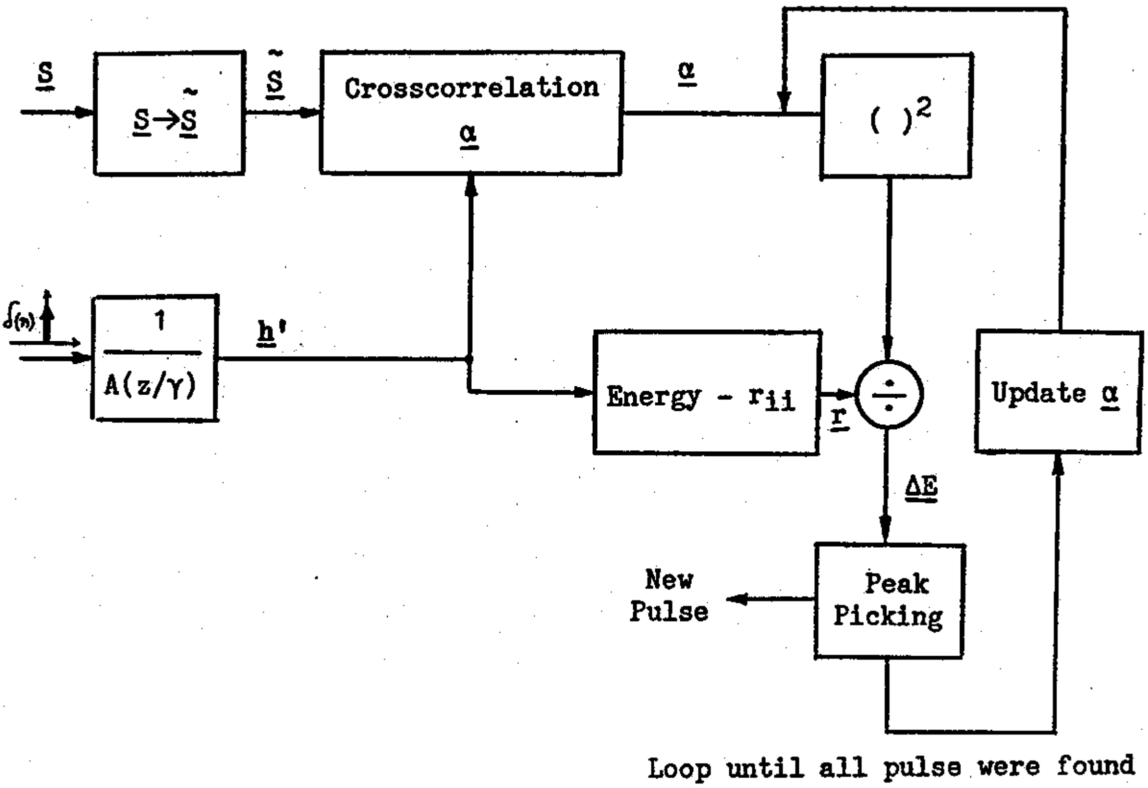
7. חשב את האמפליטודה $x_l = \frac{\alpha_l}{r_{ll}}$

8. אם נמצאו k וקטורים חשב את האות המשוחזר $\underline{\hat{S}}$ לצורך חישוב ה-leftover במסגרת הבאה וסיים, אחרת:

9. עדכן את $\underline{\alpha}$ לפי $\underline{\alpha} = \underline{\alpha} - x_l R \underline{v_l}$, חשב את $\underline{\hat{\alpha}}$

10. חזור ל-5.

בציור 5.5 מובאת סכמת בלוקים של המערכת.



ציור 5.5 סכמת בלוקים - אלגוריתם II

Fig. 5.5 Algorithem II - Block scheme

<u>סיבוכיות אלגוריתם II</u>

הסיבוכיות המתקבלת היא:

פעולות למסגרת: $O(P(L+1) + N(4P+1) + N(L+1) + 2L + k(2N+6L+L^2)$

\+

פעולות חילוק למסגרת: $O(L)$

<u>לדוגמא:</u>

עבור: N=80, P=10, L=10 נקבל:

: <u>k = 8</u>

6850 פעולות למסגרת שהן 86 פעולות לדגם, לעומת 76 באלגוריתם I.

+

10 פעולות חילוק למסגרת.

: <u>k =16</u>

9410 פעולות למסגרת שהן 118 פעולות לדגם, לעומת 98 באלגוריתם I.

+

10 פעולות חילוק למסגרת.

<u>הערה:</u> גדלי המסגרת בדוגמאות - 160 דגימות במסגרת עדכון LPC ו-80 דגימות במסגרת עבודה באלגוריתם הם גדלים אופייניים למערכת שתהיה מבוססת על אלגוריתם איטרטיבי. בחירה של k=8 מתאים לתחום הקצבים - 9.6 Kbps ו-k=16 מתאים לתחום הקצבים - 16 Kbps.

כאמור, גם עבור בחירה של 8 פולסים במסגרת של 80, שנותנים בסך הכל 800 פולסים לשניה, מקבלים איכות דיבור טובה. ירידה מתחת למספר זה נותנת אותותיה באיכות. בחירת 16 פולסים במסגרת של 80 דגימות נותנת איכות מצוינת והאות המשוחזר כמעט "שקוף" ביחס לאות המקורי.

5.4 אנליזת אוטוקורלציה

במקרה זה המטריצה R המתקבלת תהיה מטריצת אוטוקורלציה, Toeplitz, סימטרית, חיובית מוגדרת ומספיק כמובן לחשב רק שורה אחת שלה. כלומר, את N מקדמי האוטוקורלציה r_k $0 \le k \le N-1$.

$$r_{|i-j|} = r_{ij} = \sum_{n=-\infty}^{\infty} h'_n h'_{n-|i-j|} \tag{5.4.1}$$

האלגוריתם המתקבל פשוט מכיון שבמטריצה R:

$$r_{ii} = r_o \qquad 0 \le i \le N-1 \tag{5.4.2}$$

כלומר, r_{ii} קבוע ולכן ניתן לבצע את בחירת האבר המקסימאלי על הערכים המוחלטים של הוקטור $\underline{a}$.

כמו-כן, לאור ההנחה ש-$h'_n=0$ עבור $N \ge L$ נקבל מ- (5.4.1):

$$r_k = \sum_{n=k}^{L-1} h_n' \, h'_{n-k} \tag{5.4.3}$$

נציג את האלגוריתם המתקבל, סכמת הבלוקים שלו מוצגת בציור 5.6 .

אלגוריתם III

נתונים:

N - גודל מסגרת הדיבור $\underline{S}$,

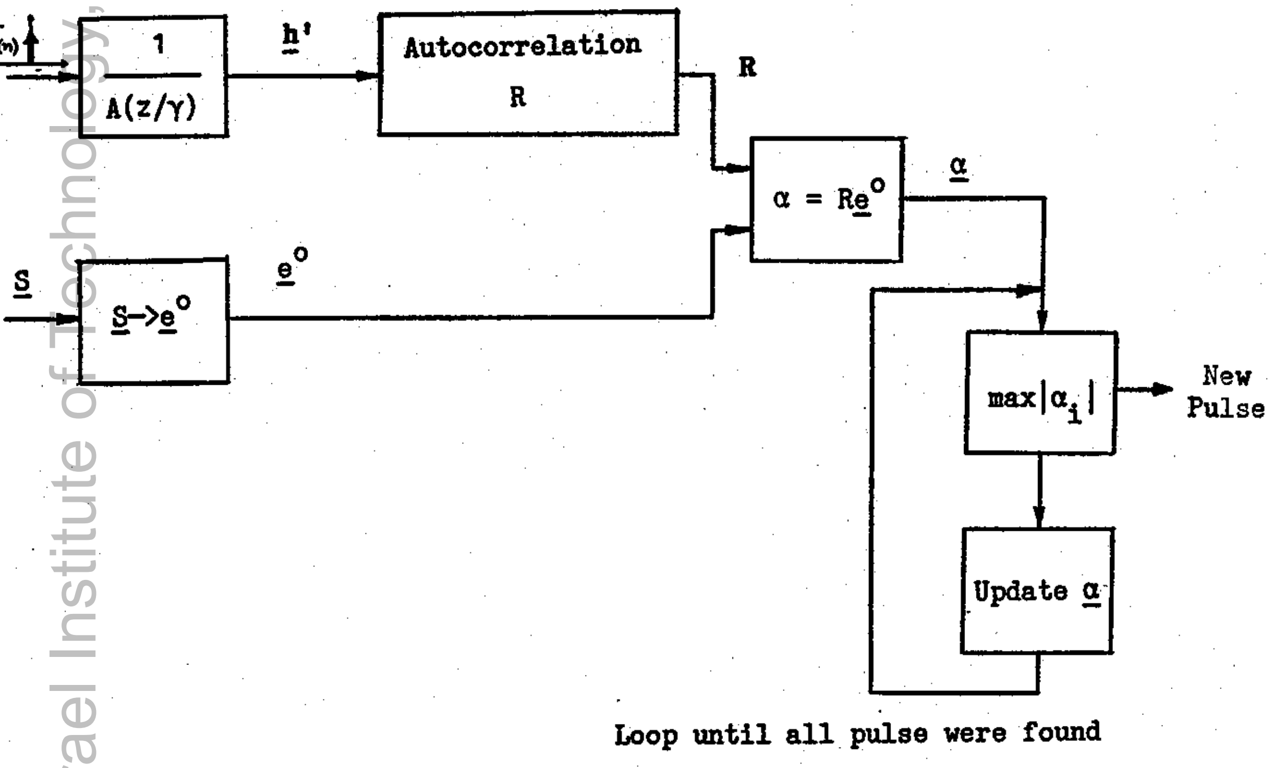
P - סדר המסנן,

k - מספר הוקטורים (פולסים) הרצויים למסגרת,

$\underline{S}$ - וקטור הדבור המקורי,

$\underline{a}$ - וקטור מקדמי ה-LPC, נתון משלב אנליזת ה-LPC,

L - מספר הדגמים השונים מ-0 ב-h'_n.



ציור 5.6 - סכמת בלוקים, אנליזת 'אוטוקורלציה' - אלגוריתם III

Fig. 5.6 - Block scheme - Algorithem III

האלגוריתם

1. חשב את h'_n - תגובת המסנן המשוקלל $\frac{1}{A(z/\gamma)}$

2. חשב את R - מטריצת האוטוקורלציה של h'_n.

3. חשב את $1/r_o$.

4. חשב את $\underline{e}^o$.

5. חשב את $\underline{\alpha} = R\underline{e}^o$

6. מצא איבר מקסימאלי בערך מוחלט ב-$\underline{\alpha}$, יהי אינדקסו l.

7. חשב את האמפליטודה $x_1 = \frac{\alpha_1}{r_o}$

8. אם נמצאו k פולסים חשב את האות המשוחזר $\hat{\underline{S}}$ לצורך חישוב ה-leftover במסגרת הבאה וסיים. אחרת:

9. עדכן את $\underline{\alpha}$: $\underline{\alpha} = \underline{\alpha} - x_1 R \underline{v}_1$

10. חזור ל-6.

סיבוכיות אלגוריתם III

פעולות למסגרת: $O(P(L+1) + \frac{(L+1)L}{2} + (2P+1)N + N(2L-1) + k(2N+3L) + NP)$

+

פעולות חילוק יחידה למסגרת.

לדוגמא:

עבור: N=80, P=10, L=10 נקבל:

k=8

5685 פעולות למסגרת שהן 71 פעולות לדגם לעומת 76 באלגוריתם I.

+

פעולת חילוק אחת למסגרת לעומת 10 באלגוריתם I.

k=16

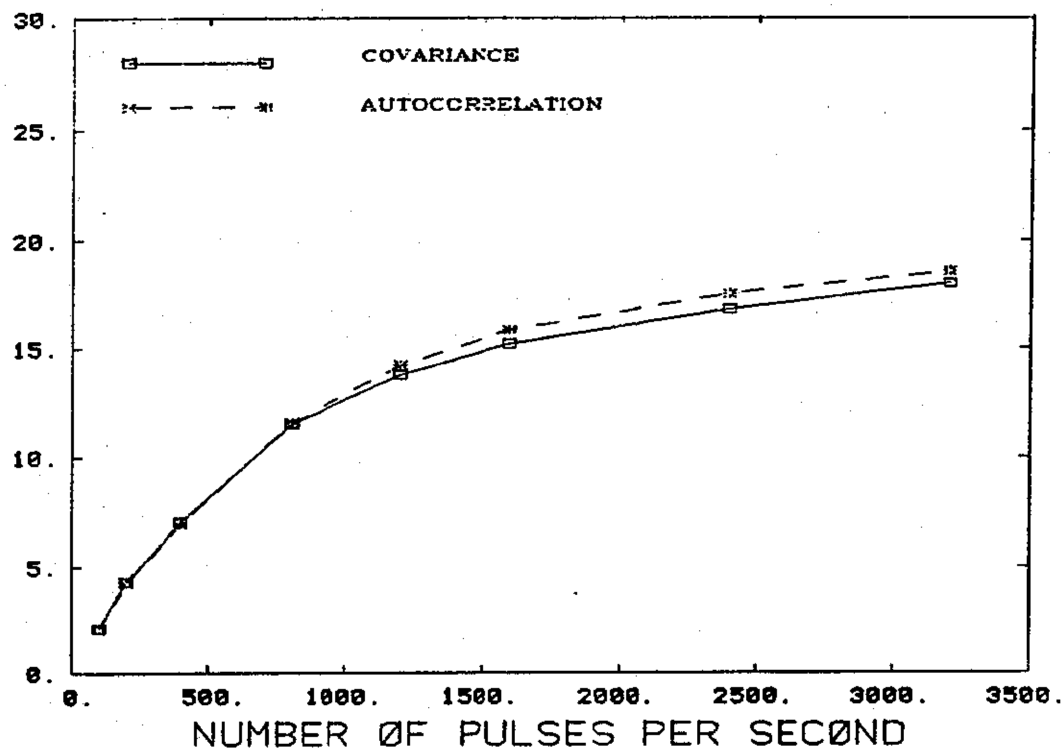
7205 פעולות למסגרת שהן 90 פעולות לדגם לעומת 98 באלגוריתם I.

+

פעולת חילוק אחת למסגרת לעומת 10 באלגוריתם I.

סימולציה לאלגוריתם III

ערכנו סימולציה באותם תנאים בהם נעשתה סימולציה לאנליזה ה"קווריאנס" והתוצאות כמעט זהות, גם במבחני השמעה כמעט ולא מורגש הבדל בין סוגי האנליזה השונים. בציור 5.7 מובאים ערכי ה-SNR המתקבלים בשתי השיטות בתנאי הסימולציה שצויינו כשדנו באנליזת קווריאנס.



ציור 5.7 SNR מתקבל באלגוריתם I ו-III כפונקציה של מספר הפולסים

Fig. 5.7 SNR in Algorithms I & III vs. Number of Pulses

5.5 המלצות למימוש המערכת

כפי שראינו בניתוח צורות האלגוריתם השונות האיכות המתקבלת זהה עבור אנליזת "קווריאנס" ואנליזת "אוטוקורלציה" בעוד שהסיבוכיות באנליזת "אוטוקורלציה" נמוכה יותר. בבואנו להציע איפוא מערכת המבוססת על אלגוריתם איטרטיבי נבחר באנליזת "אוטוקורלציה" עם $\gamma=0.8$ ו-$L=10$, זהו ערך מספיק כפי שניתן לראות מציור 5.2 .

כפי שראינו בפרק הקודם ניתן להשיג שיפור באיכות אם נבצע אופטימיזצית אמפליטודות. מבחני שמיעה הראו שדי להסתפק באופטימיזציה יחידה בסוף כל מסגרת כדי להשיג שיפור מורגש באיכות, מה עוד שהסיבוכיות הכרוכה באופטימיזציה לאחר כל איטרציה תהיה גבוהה יותר.

הסיבוכיות הכרוכה באופטימיזצית אמפליטודות

בסעיף (4.3.1) ראינו כי האופטימיזציה כרוכה בפתרון מערכת משוואות ליניארית

$$(5.5.1) \qquad R'\underline{x} = \underline{\alpha}'$$

כאשר המטריצה R' והוקטור $\underline{\alpha}'$ מתקבלים על-ידי בחירת אברים מתוך המטריצה R והוקטור $R\underline{e}$ בהתאמה. את פתרון המערכת עצמה ניתן לעשות בעזרת Cholesky Decomposition.
נקבל איפוא:

- עבור אופטימיזציה בודדת בסוף כל מסגרת - תוספת סיבוכיות מסדר של $O(k^3)$ פעולות למסגרת.
- עבור אופטימיזציה לאחר כל איטרציה נקבל תוספת סיבוכיות:

$$O\,(2^3+3^3+4^3+\ldots+k^3) = O\,(\frac{k^2(k+1)^2}{4})$$

כלומר, עבור 8 פולסים למסגרת (k=8)' נקבל תוספת סיבוכיות של כ-512 פעולות למסגרת אם נבצע אופטימיזציה יחידה בסוף המסגרת, ו-1300 פעולות למסגרת אם נבצע אופטימיזציה אחרי כל איטרציה.
עבור 16 פולסים למסגרת (k=16) נקבל תוספת סיבוכיות של כ-4096 פעולות למסגרת אם נבצע אופטימיזציה יחידה בסוף המסגרת, וכ-18,500 פעולות למסגרת אם נבצע אופטימיזציה אחרי כל איטרציה.

גדלי מסגרות

שאלה נוספת שעלינו לשאול בבואנו לבחור פרמטרים למערכת היא באיזה גודל מסגרת להשתמש?
בחירת מסגרת קטנה תקטין את הסיבוכיות אך נצפה לקבל פגיעה באיכות.
כדי לבחון את השפעת גודל המסגרת בוצעו סימולציות עם גדלי מסגרות שונים:
נתוני המערכת הם: סדר מסנן LPC - P=10, L=10, γ=0.8, נבחרים 800 או 1600 פולסים לשניה, מתבצעת אופטימיזציה יחידה בסוף כל מסגרת ומבוצעת קוונטיזציה למקדמי ה-LPC.

בטבלה 5.1 מוצגות תוצאות ה-SNRSEG (Segmented SNR) כאשר גודל אינטרוול עדכון מקדמי ה-LPC הוא 160 דגמים וגודל המסגרת עליו מופעל האלגוריתם הוא: 10, 20, 40, 80, 160 .

בטבלה 5.2 מוצגות תוצאות ה-SNRSEG כאשר גודל אינטרוול עדכון מקדמי ה-LPC הוא 240 דגמים וגודל המסגרת עליו מופעל האלגוריתם הוא: 10, 20, 40, 80, 120, 240.
בכל מקרה גודל המסגרת עליה מחושבים מקדמי האוטוקורלציה באנליזת ה-LPC הוא 256.
כפי שניתן לראות מהתוצאות מקבלים פגיעה אובייקטיבית בביצועים כשיורדים מתחת לגודל מסגרת של 40 דגמים, בפגיעה זו מבחינים היטב בבדיקות שמיעה.

SNRSEG [dB] 1600 פולסים לשניה	SNRSEG [dB] 800 פולסים לשניה	גודל מסגרת בסיסי
15.5	10.2	10
16.3	11.1	20
16.5	11.5	40
16.3	11.6	80
15.8	11.6	160

טבלה 5.1 - SNRSEG כפונקציה של גודל המסגרת באלגוריתם, M=160

Table 5.1 - SNRSEG obtained for different frame sizes, M=160

SNRSEG [dB] 1600 פולסים לשניה	SNRSEG [dB] 800 פולסים לשניה	גודל מסגרת בסיסי
15.3	10.2	10
16.1	11.1	20
16.4	11.5	40
16.3	11.6	80
16.2	11.6	120
16.2	11.6	240

טבלה 5.2 - SNRSEG כפונקציה של גודל המסגרת באלגוריתם, M=240

Table 5.2 - SNRSEG obtained for different frame sizes, M=240

סיבוכיות המקלט

המקלט במערכת זו פשוט מאד מכיון שלצורך קבלת האות המשוחזר יש להעביר את ערור ה-Multi-Pulse דרך מסנן הסינטזה: O(NP) פעולות. כלומר, עבור מסנן LPC מסדר 10 נקבל שיש צורך ב-O(10) פעולות לדגם לבצוע סינטזה אם נשתמש במימוש ישיר.

קידוד אות הערור

האינפורמציה שיש לשדר על אות הערור מורכבת מ-

(i) אינדקסי הוקטורים שנבחרו (מיקום הפולסים במסגרת).

(ii) המקדמים הכופלים את הוקטורים (אמפליטודות הפולסים).

נראה כיצד נקודד אינפורמציה זו:

(i) השיטה ה"קלאסית" לקידוד מיקום הפולסים היא לשדר את המקום הראשון ואח"כ לשדר הפרשים בין המקומות. וזאת השיטה בה נקטנו לצורך חישוב הקצבים. עבור צפיפות פולסים של כ-8 מתוך 80 מספיקות 5 סיביות לקידוד כל מקום, כדי לא לאפשר חריגה מעבר למרחק של 32 דגמים בין כל שני פולסים ניתן להכניס הגבלה תוך כדי שלב חיפוש הפולסים.
עבור צפיפות של כ-16 פולסים במסגרת של 80 דגמים ניתן להסתפק ב-4 סיביות לפולס לצורך קידוד המקום.
שיטה אחרת שהוצעה בספרות [18] - קידוד קומבינטורי: כל קומבינצית פולסים מיוצגת ע"י וקטור מאורך N, וקטורים אלה מסודרים בסדר לקסיקוגרפי ומשודר האינקדס המתאים לוקטור המיקומים שנבחר. ב-[18] מוצעת שיטה פשוטה למציאת האינדקס המשודר ללא צורך בהחזקת מילון גדול או חיפוש מייגע. החסרון של שיטת קידוד זאת הוא בחסינות נמוכה לרעש.

(ii) הצורה בה בחרנו לקודד את אמפליטודות הפולסים מבוססת על שיטה שהוצעה ב-[25]:

1) נרמל את כל הפולסים לפי הגדול ביותר, פקטור הנרמול נקרא - SF - Scale Factor.

2) SF עובר קוונטיזציה ל-6 סיביות לאחד הערכים המתקבלים מ-

$$SF = (ILINER + 8) \times \left({}_2 IRANGE\right)$$

כאשר: $3 \text{ bit} \longleftarrow 0 \le ILINER \le 7$

$\underline{3 \text{ bit}} \longleftarrow 0 \le IRANGE \le 7$

6 bit

בצורה זו SF עובר קוונטיזציה "לוגריתמית".

כלומר, SF מקודד לאחד מ-64 ערכים. הקידוד והפענוח נעשים באמצעות טבלה.

3) הפולסים המנורמלים עוברים כימוי ל-4 סיביות כל אחד: סימן + 8 רמות קוונטיזציה.

סיכום

בטבלה 5.3 מופיעים אינטרוולי עדכון מקדמי LPC, גדלי מסגרות, סיבוכיות ו-SNRSEG מתקבל עבור צורות שונות למימוש מערכת בתחום הקצבים 9.6 Kbps ÷ 8 Kbp .

אינטרוול עדכון LPC	מסגרת Multi-Pulse	פולסים למסגרת	סיבוכיות [פעולות]	קצב [bps]	SNRSEG [dB]
160	80	7	93	8950	10.8
160	40	3	83	8650	9.6
180	90	9	99	9556	11.2
180	45	4	84	9289	10.7
200	100	10	101	9320	11.5
200	50	4	81	8360	10
240	80	8	91	9167	11.6
240	40	3	77	7967	9.5

טבלה 5.3 סיכום תוצאות עבור תחום הקצבים 9.6 Kbps ÷ 8 Kbp .

Table 5.3 Results for 8 ÷ 9.6 Kbps

הערות:

- סיבוכיות חישוב מקדמי ה-LPC למסגרת באלגוריתם היא:

$O(\frac{N}{M}(m(P+2)+P^2))$ כאשר:

m - הוא אורך הבלוק עליו מחושבים מקדמי אוטוקורלציה באנליזת ה-LPC.

M - הוא אינטרוול העדכון של מקדמי ה-LPC.

N - הוא גודל מסגרת באלגוריתם, M=cN כאשר c מספר שלם.

- הסיבוכיות הנתונה בטבלה כוללת את חישוב מקדמי ה-LPC.

- חישוב הסיבוכיות הנתון בטבלה 5.3 הוא ב-"פעולות לדגם" והוא עבור מערכת בעלת מסנן LPC מסדר 10 (P=10) מקדמי המסנן מחושבים בשיטת האוטוקורלציה. אלגוריתם ה-Multi-Pulse הוא מהסוג של אנליזת "אוטוקורלציה", L=10, מקדם השקלול - γ=0.8. מתבצעת <u>אופטימיזציה יחידה</u> בסוף כל מסגרת.

יש לזכור שבנוסף לפעולות הרגילות הנתונות בסיבוכיות לעיל יש לבצע פעולת חילוק יחידה עבור כל מסגרת, אולם, פעולה זו זניחה יחסית למספר הפעולות הדרוש לדגם.

בטבלה 5.4 מופיעים גדלי מסגרות, סיבוכיות, ו-SNRSEG מתקבל עבור צורות שונות למימוש מערכת בתחום הקצבים 16 Kbps .

SNRSEG [db]	קצב [bps]	סיבוכיות [פעולות]	פולסים למסגרת	מסגרת Multi-Pulse	אינטרוול עדכון LPC
16.3	15450	161	16	80	160
15.5	14450	102	7	40	160
16.6	15867	190	19	90	180
16.4	15689	112	9	45	180
16.4	15560	209	21	100	200
16.5	15400	116	10	50	200
16.7	15567	167	17	80	240
16.4	15367	102	8	40	240

טבלה 5.4 סיכום תוצאות עבור תחום הקצבים 16 Kbps

Table 5.4 Results for the 16 Kbps range

5.5.1 מערכת מומלצת

כדי לשמור על סיבוכיות נמוכה, איכות טובה וכדי להתקרב לקצבי השידור 9.6 Kbps ו-16 Kbps מומלצת המערכת הבאה:

$m = 256$ - גודל הבלוק עליו מחושבים מקדמי אוטוקורלציה באנליזת ה-LPC.
$M = 180$ - אינטרוול עדכון מקדמי ה-LPC.
$N = 45$ - גודל המסגרת באלגוריתם, כל 4 מסגרות כאלה מעודכנים מקדמי ה-LPC.
$P = 10$ - סדר מסנן.
$\gamma = 0.8$ - מקדם השקלול.
$L = 10$ - מספר האברים השונים מ-0 ב-h'_n.
מספר הפולסים הנבחר: $k = 4$ לקצב שידור 9.6 Kbps
$k = 9$ לקצב שידור 16 Kbps
בטבלה 5.5 מסוכמים ביצועי המערכת.

קצב [bps]	סיביות שנותרות להגנה וסנכרון	SNR [dB]	SNRSEG [dB]	פעולות לדגם משדר	פעולות לדגם מקלט
9600	311	9.6	10.7	84	10
16000	311	15.6	16.4	112	10

טבלה 5.5 תוצאות מערכת מומלצת

Table 5.5 Results for recommended system

במערכת זו נשתמש גם בפרק 7 להערכת ביצוע האלגוריתם האיטרטיבי מבחינת Tandeming, חסינות לרעש רקע, רעש ערוץ, העברת אותות (VBD) Voice Band Data והשוואה ל-PCM.

פרק 6: אלגוריתמים לא איטרטיביים

6.1 מבוא

כפי שראינו בפרק 4 ניתן למצוא פתרונות לא איטרטיביים לבעיית הפרוק הוקטורי (פתרונות לא בהכרח אופטימליים).

בפרק זה נבחן ביצועי שניים מאלגוריתמים אלה: ה-GMRM שמבוסס על בחירת שיאים באות השארית וה-PTC (Predictive Transform Coder) שמבוסס על הקצאה דינאמית של סיביות בתחום התדר.

6.2 אלגוריתם ה-GMRM

הדיון באלגוריתם זה יהיה קצר מכיוון שביצועיו גרועים גם ללא קוונטיזציה של אינפורמצית הצד.
המערכת עצמה הוצגה בסעיף 4.4.1 והיא מבוססת על בחירת שיאים באות השארית המשוקלל בתדר (צבוע). האלגוריתם הוא הכללה של אלגוריתם ה-MRM [17] שבו נבחרו שיאים מאות השארית ללא "צביעה" כלומר - $\gamma = 0.0$, כאשר γ קובע את מידת הצביעה על-ידי המסנן $1/A(z/\gamma)$.

בטבלה 6.1 מסוכמים ערכי ה-SNR המתקבלים עבור ערכי γ שונים עבור הפרמטרים הבאים: אינטרוול עדכון מקדמי LPC - 180 דגמים, מסגרת של 45 דגמים ו-9 שיאים הנבחרים למסגרת.

SNRSEG [dB]	SNR [dB]	γ
7.3	7.7	0.0
8.3	8.5	0.2
7.6	7.7	0.4
7.3	7.7	0.6
7.2	7.6	0.8
4.3	4.6	1.0

טבלה 6.1 - השתנות ה-SNR בתלות ב-γ באלגוריתם GMRM

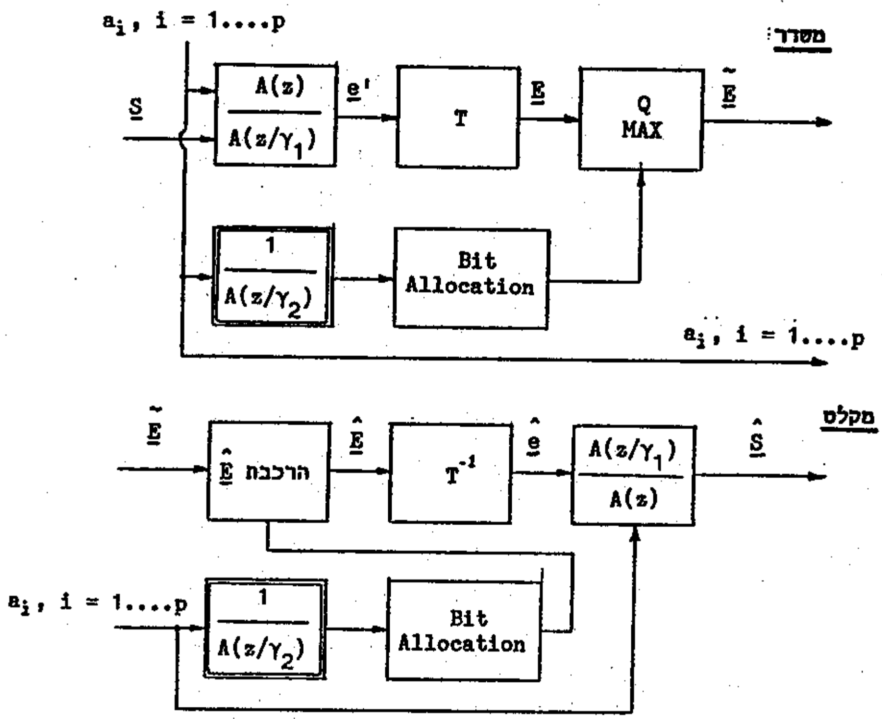
Table 6.1 - SNR vs. γ in GMRM algorithm

מתוך הטבלה ומתוך מבחני האזנה מתקבל שהתוצאות הטובות לא מתקבלות בהכרח עבור $\gamma = 0.0$ כפי שהוצע ב-MRM.

התוצאות בטבלה הן ללא קוונטיזציה של אינפורמצית הצד ומבחינת תחום קצב השידור הן מתאימות לקצב שידור של כ-16 Kbps, כמובן שהתוצאות שהתקבלו לא עומדות במבחן לעומת תוצאות שהתקבלו באלגוריתם האיטרטיבי: עבור תחום הקצבים של 16 Kbps - SNRSEG = 16.1 dB.

הסיבוכיות של האלגוריתם היא אמנם נמוכה מאד אולם התוצאות אינן טובות. שיפור שניתן להוסיף לאלגוריתם ושהוצע במקור ב-[16] הוא לחשב מחדש את מקדמי ה-LPC בהנתן הערור, תוספת זו משפרת את האיכות אולם לא במידה מספקת.

ניזכר במערכת שהוצגה בפרק 4:



ציור 6.1 מערכת משדר-מקלט PTC

Fig. 6.1 PTC Vocoder

במערכת זו איננו בוחרים שיאים בודדים בתחום התדר, אלא, מפזרים את הסיביות הנתונות לנו לצורך השידור בדומה לנעשה ב-(ATC) Adaptive Transform Coding על פני יותר רכיבי התמרה (להבא נשתמש במונח "רכיבי תדר" מכיון שההתמרות בהן אנו דנים - DCT ו-DFT הן ממשפחת ה-Frequency transformations).

במשדר מייצרים אות שארית משוקלל $\underline{e}$' ע"י העברת אות השארית המקורי דרך המסנן $1/A(z/\gamma_1)$.

האות $\underline{e}$' עובר התמרה ולמקדמי ההתמרה מוקצות סיביות לפי שערוך הווריאנס שלהם. הווריאנס משוערך על-ידי חישוב הספקטרום של המסנן $1/A(z/\gamma_2)$, שערוך זה נעשה בהתבסס על העובדה שאות השארית $\underline{e}$ הוא בקירוב לבן והוא נצבע ע"י $\frac{1}{A(z/\gamma_1)}$.

ערכו של γ_1 המשמש לצביעה איננו בהכרח זהה לערכו של γ_2 המשמש לצורך חישוב המעטפת הספקטרלית אם ברצוננו להשיג תוצאות מיטביות וערכים אלה נקבעו בעזרת מבחני האזנה.

למקלט משודרים מקדמי ההתמרה ואינפורמצית צד הכוללת מקדמי PARCOR של LPC ושערוך ווריאנס. המקלט משתמש באינפורמצית הצד לשחזר את ההתמרה של אות השארית, ומבצע התמרה הפוכה לקבלת אות שארית משוקלל משוחזר אותו הוא מעביר דרך מסנן סינטזה $\frac{A(z/\gamma_1)}{A(z)}$ לקבלת אות הדיבור המשוחזר.

היחידות העיקריות במערכת הן:

- התמרת אות השארית.
- קוונטיזצית מקדמי ההתמרה.

נייחד את שני הסעיפים הבאים לדיון תאורטי בשתי יחידות אלו ואת הסעיף השלישי לשיקולי מימוש המערכת וסיבוכיות.

6.3.1 DCT על אות השארית

כפי שראינו בפרק 4 התמרת ה-DCT נותנת תוצאות עדיפות על פני התמרת ה-DFT ולכן נשתמש בה למימוש ההתמרה T במערכת.
התמרת ה-DCT הוצעה לראשונה ב-[37] והיא מוגדרת כדלקמן:

$$X_k = \frac{2C(k)}{N} \sum_{m=0}^{N-1} x_m \cos\left[(2m+1)\pi k/2N\right] \qquad (6.3.1)$$

$$k=0,1,2\ldots N-1$$

כאשר:

$$C(k) = \begin{cases} 1/\sqrt{2} & k=0 \\ 1 & k=1,2\ldots N-1 \end{cases} \qquad (6.3.2)$$

באופן דומה ההתמרה ההפוכה מוגדרת כ-

$$x_n = \sum_{k=0}^{N-1} X_k \, C(k) \cos\left[(2n+1)\pi k/2N\right] \qquad (6.3.3)$$

$$n=0,1,2\ldots N-1$$

ניתן לראות מ-(6.3.1) כי מקדמי ה-DCT - X_k הם מספרים ממשיים עבור x_n ממשי.

בספרות הוצעו כמה שיטות לחישוב ה-DCT. השיטה הקלאסית שהוצעה ב-[37] היא כדלקמן:
(6.3.1) ניתן להצגה כ-

$$X_k = \frac{2C(k)}{N} \operatorname{Re}\left[e^{-jk\pi/2N} \sum_{m=0}^{2N-1} x_m W_{2N}^{mk}\right] \qquad (6.3.4)$$

$$k=0,1,2\ldots N-1$$

כאשר:

$$W_{2N} = \exp(-j2\pi/2N) \qquad (6.3.5)$$

ו-

$$x_m = 0 \qquad m = N, N+1, \ldots 2N-1 \qquad (6.3.6)$$

ומכאן, ניתן לחשב DCT של סדרה מאורך N על יד הוספת N אפסים לסדרה, שמוש ב-FFT על סדרה מאורך 2N, כפל ב-$\exp(-jk\pi/2N)$ ולקיחת החלק הממשי.
ל-IDCT (6.3.3) צורה דומה ל-DCT (6.3.1) וגם הוא ניתן לחישוב בעזרת FFT מאורך 2N.

חסרונה של שיטה זו הוא הסיבוכיות הגבוהה יחסית, בספרות הוצעו שיטות נוספות לחישוב ה-DCT, ביניהן שיטות המתבססות על FFT מאורך N [38] ושיטות המתבססות על פקטוריזציה של מטריצות התמרה ממשיות לתת מטריצות [39], [40], [41], [42].
השיטות האחרונות הובילו לאלגוריתמים מהירים לחישוב ה-DFT, DCT ו-DST (Discrete Sine Transform).
לצורך ההשוואה בין האלגוריתמים המהירים מובאת הטבלה הבאה המפרטת את מספר הכפלים והחיבורים הממשיים הנדרשים בכל אחת מהשיטות המהירות:

מקור	[39] CHEN	[40] WANG	[41] SUEHIRO [42] WANG
מספר כפלים	$N\log_2 N-3N/2+4$	$N((3/4)\log_2 N-1)+3$	$(N/2)\log_2 N+1$
מספר חיבורים	$(3N/2)(\log_2 N-1)+2$	$N((7/4)\log_2 N-2)+3$	$(3/2)N\log_2 N-N+1$

טבלה 6.2 - סיבוכיות נדרשת בחישוב DCT

Table 6.2 Arithmetic Operation needed for DCT

לדוגמא, עבור N=128 נקבל:

	[39] CHEN	[40] WANG	[41] SUEHIRO [42] WANG
מספר כפלים	708	547	449
מספר חיבורים	1154	1315	1217
סה"כ פעולות	1862	1862	1666

טבלה 6.3 - סיבוכיות נדרשת בחישוב DCT, N=128

Table 6.3 Arithmetic Operations needed for DCT, N=128

אנו נניח שהסיבוכיות הנדרשת לצורך חישוב DCT מאורך N היא סכום פעולות הכפל והחיבור למרות שב-DSP קיימים כיום ניתן לבצע פעולת MAD (Multiply & Add) המתבצעת במחזור מכונה אחד [43].

6.3.2 קוונטיזצית מקדמי ההתמרה

את קוונטיזצית מקדמי ההתמרה אנו מבצעים בעזרת קוונטיזר אחיד. כל מקדם עובר קוונטיזציה ומשודר למקלט. הקוונטיזר למקדם מסוים $E_{(k)}$ מאופין על-ידי צעד קוונטיזציה $\Delta_{(k)}$ ומספר רמות $2^{b(k)}$ כאשר $b_{(k)}$ הוא מספר הסיביות שהוקצו לצורך קוונטיזצית המקדם.

גודל צעד הקוונטיזציה $\Delta_{(k)}$ ומספר הסיביות $b_{(k)}$ תלויים בשערוך ווריאנס המקדם $E_{(k)}$. נניח כרגע שהקצאת הסיביות למקדמים כבר נמצאה ושנתון שערוך לווריאנס המקדמים, נעיר כי אם מבצעים התמרה לאות שארית משוקלל נקבל ווריאנס שונה עבור כל מקדם בעוד שאם מבצעים התמרה לאות שארית לא צבוע ("לבן") ניתן להניח שהווריאנס הוא אותו לכל המקדמים.

למקדמי ההתמרה ניתן להתייחס כאל משתנים אקראיים. כל מקדם הוא קומבינציה של N משתנים אקראיים כאשר N הוא אורך ההתמרה. ה-pdf של מקדם מסוים תלוי איפוא בסטטיסטיקת הכניסה, סוג ההתמרה והגודל N. עבור כניסות גאוסיות גם מקדמי ההתמרה יהיו גאוסיים ועבור כניסות בעלות pdf שונה, ה-pdf של מקדמי ההתמרה ישאף לגאוסי כאשר $N \longrightarrow \infty$.

עבור N מספיק גדול נניח כי בקירוב טוב מקדמי ההתמרה מפולגים גאוסית. במקרה כזה בחירת גודל צעד הקוונטיזציה (אחיד) $\Delta_{(k)}$ האופטימאלי נקבע מתוך הווריאנס המשוערך $\hat{\sigma}^2_{(k)}$ לפי MAX [44]. עבור מספר נתון של סיבית $b_{(k)}$ גודל הצעד האופטימאלי יהיה:

$$\Delta_{(k)} = \alpha(b_{(k)})\ \hat{\sigma}_{(k)} \qquad (6.3.7)$$

כאשר $\alpha(b_{(k)})$ הוא קבוע הנקבע לפי $b_{(k)}$ מתוך הטבלאות של MAX עבור קוונטיזר אחיד [44].

אלגוריתם ה-PTC דומה בקווי הכלליים ל-ATC וגם הקצאת הסיביות תיעשה בצורה דומה. בהנחה שהמקדמים מפולגים בקירוב גאוסית ידוע מ-ATC‏ [45], [46], [47], [48] שהקצאת הסיביות האופטימאלית ניתנת על-ידי:

$$(6.3.8) \qquad b(k) = \bar{B} + \frac{1}{2}\log_2 \frac{\hat{\sigma}_{(k)}^{\;2}}{\left[\prod_{j=0}^{N-1} \hat{\sigma}_{(j)}^{\;2}\right]^{1/N}}$$

כאשר $\bar{B}$ הוא הקצב הממוצע:

$$(6.3.9) \qquad \bar{B} = 1/N \sum_{k=0}^{N-1} b_{(k)} = \text{Const.}$$

על ההקצאה (6.3.8) פועלים כמה אילוצים:

(i) $b_{(k)}$ חייב להיות שלם.

(ii) $b_{(k)} \geq 0$, יש לתת הקצאה של 0 סיביות למקדמים שקיבלו הקצאה שלילית.

(iii) $b_{(k)} \leq b_{max}$ אנו נגביל את מספר הסיביות המקסימאלי שמקדם יכול לקבל.

הקצאת הסיביות נמצאת באופן מעשי בתהליך בן 3 פאזות [48]:
פיתוח של (6.3.8) יתן:

$$(6.3.10) \qquad b_{(k)} = \bar{B} + \log_2 \hat{\sigma}_{(k)} - 1/2 \log_2\left[\prod_{j=0}^{N-1} \hat{\sigma}_{(j)}^{\;2}\right]^{1/N} = \log_2 |\hat{\sigma}_{(k)}| - D$$

כאשר, D הוא קבוע.

$$b_{(k)} = \lfloor \log_2 |\hat{\sigma}_{(k)}| - D \rfloor^* \tag{6.3.11}$$

כאשר:

$$\lfloor u \rfloor^* = \begin{cases} 0 & \text{if } u < 0 \\ \lfloor u \rfloor & 0 \le u < b_{max} \\ b_{max} & \text{if } u \ge b_{max} \end{cases} \tag{6.3.12}$$

כאשר b_{max} הוא מקסימום הסיביות שאנו מיעדים למקדם.

הקבוע D צריך להבחר כך שמתקיים:

$$B = \sum_{k=0}^{N-1} b_{(k)} \tag{6.3.13}$$

כאשר B הוא סך כל הסיביות המוקצות למקדמי ההתמרה למסגרת.

ערכו של D משוערך בשני מעברים. תחילה משוערך D כ-

$$D_1 = 1/N \left[\sum_{k=0}^{N-1} \log_2 |\hat{\sigma}_{(k)}| - B \right] \tag{6.3.14}$$

תוצאה זו מתקבלת אם נסכם את $b_{(k)}$ מ-(6.3.10) על פני $k = 0 \ldots N-1$. בדרך כלל שערוך זה של D הוא גדול מדי מכיוון שקיימות הרבה נקודות עבורן:

$(\log_2 |\hat{\sigma}| - D)$ שלילי.

לכן, מתבצע מעבר נוסף לתיקון D. נגדיר את הקבוצה S^+:

$$S^+ = \{ k \text{ such that } (\log_2 |\hat{\sigma}_{(k)}| - D_1) > 0 \} \quad (6.3.15)$$

כלומר, זוהי קבוצת אינדקסי המקדמים שמקבלים הקצאה לאחר המעבר הראשון לשערוך D. יהיה N^+ מספר האברים ב-S^+ אזי שערוך שני ל-D יתקבל על-ידי:

$$D_2 = D_1 + 1/N^+ \left[\sum_{k \in S^+} [\log_2 |\hat{\sigma}_{(k)}| - D_1]^* - B \right] \quad (6.3.16)$$

כלומר, את עודף או חוסר הסיביות הנוצר כתוצאה משימוש ב-D_1 מקזזים על-ידי תיקון D_1. בשלב זה, ערכו של D_2 קרוב לערך האופטימאלי של D. ערכי k שאינם שייכים ל-S^+ מקבלים הקצאה של 0 סיביות, ומשתמשים בחוק היוריסטי כדי לקזז סיביות עודפות או חסרות הנגרמות כתוצאה משימוש ב-D_2 כשערוך ל-D. בסך הכל מגיעים ל-B סיביות למסגרת.

החוק בו השתמשנו אומר שכאשר יש הקצאת יתר של סיביות הן מופחתות מהמקדמים שקבלו הקצאה עשירה של סיביות, כאשר יש מקום להוסיף סיביות מוסיפים אותן למקדמים שקבלו הקצאה של סיבית אחת בלבד.

6.3.2.2 שערוך הווריאנס של המקדמים

ביצוע האדפטציה בהקצאת הסיביות ובגודל צעד הקוונטיזציה תלוי כפי שציינו בשערוך ווריאנס המקדמים - הגודל שסימנו ב-$\hat{\sigma}^2_{(k)}$.

שערוך ווריאנס זה נקבל מתוך חישוב הספקטרום של המסנן הצובע את אות השארית. למעשה, נתייחס לשני מסננים צובעים: $1/A(z/\gamma_1)$ שיצבע את אות השארית, מתוכו נשערך את ווריאנס המקדמים לצורך קביעת צעד הקוונטיזציה, ו-$1/A(z/\gamma_2)$ שעל-פי המעטפת שלו נקבע את הקצאת הסיביות $b_{(k)}$. המוטיבציה לא להשתמש באותו ערך γ בשני המקרים היא ששימוש באותו γ יתן מינימום שגיאה ריבועית ממוצעת על אות השארית הצבוע e', אולם, אנו מחפשים שגיאה ריבועית ממוצעת מינימאלית על אות הדיבור המשוחזר ומקסימום איכות במבחני האזנה. לכן, את ערכי γ_1 ו-γ_2 קבענו במבחני האזנה כפי שנראה בסעיף הבא.

6.3.3 המלצות למימוש מערכת ה-PTC

6.3.3.1 בחירת ערכי γ_1 ו-γ_2

γ_1 קובע את מידת הצביעה של אות השארית, אות השארית נצבע על ידי המסנן: $\frac{1}{A(z/\gamma_1)}$.

γ_2 קובע את הקצאת הסיביות $b_{(k)}$ מכיוון שהיא מבוצעת לפי הספקטרום של תגובת המסנן $\frac{1}{A(z/\gamma_2)}$.

טבלה 6.4 מתארת תוצאות אוביקטיביות - SNRSEG שהתקבלו עבור ערכי γ_1 ו-γ_2 שונים. הערכים המצויינים בטבלה התקבלו עבור גודל מסגרת: N=128 ללא קוונטיזציה של מקדמי ה-LPC ושערוך הווריאנס. כאשר לכל מסגרת מוקצות 100 סיביות (ערך אופייני לקבלת 9.6 Kbps) ומקסימום הסיביות המוקצות למקדם הוא 5 סיביות.

γ_2 \ γ_1	0.0	0.4	0.8	1.0
0.0	2.3	2.4	2.5	2.4
0.4	7.0	7.2	7.4	7.5
0.8	10.6	10.8	11.2	11.2
1.0	12.2	12.6	12.9	12.9
1.2	11.0	11.4	11.5	11.5

טבלה 6.4 SNRSEG ב-dB המתקבל במערכת PTC בתלות ב-γ_1 ו-γ_2

Table 6.4 SNRSEG in dB obtained in PTC Vocoder for different γ_1 & γ_2

מתוך הטבלה ניתן להבחין בתופעה שבה מבחינים גם בבחינות סוביקטיביות: ההשפעה המכריעה על האיכות ניתנת על-ידי הצביעה של מעטפת המסנן $\frac{1}{A(z/\gamma_2)}$ לצורך הקצאת הסיביות $b_{(k)}$.

לכן, נשתמש בערך $\gamma_2 = 1.0$ לצורך הקצאת הסיביות, כלומר, הקצאת הסיביות נעשית לפי מעטפת מסנן הסינטזה. עליה מעל ערך של $\gamma_2 = 1$ גורמת להופעת "צלצולים" ולתופעת Low-pass עקב ריכוז סיביות במעט מקומות בתדר.
למרות ערכי ה-SNR הדומים ל"רוחב" הטבלה, ניתן להרגיש ברעש הנגרם כתוצאה מהתפרים בין המסגרות החל מ-$\gamma_1 = 0.6$ ומעלה. רעש זה נובע מהעובדה שככל ש-γ_1 גדול יותר אנו מתקרבים לאות הדיבור המקורי ומופיעה הבעיה הידועה מ-ATC של רעש הנובע מהמסגרות.
רעש המסגרת איננו מופיע עבור ערכי γ_1 נמוכים בגלל אפקט ההחלקה שנותן המסנן למעבר בין המסגרות.
עבור $\gamma_1 = \gamma_2 = 1.0$ נקבל את אחת מצורות המימוש של ATC.

בהנתן $\gamma_2 = 1.0$ קשה להבחין בהבדל בדיבור המשוחזר המתקבל עבור ערכי γ_1 קטנים מ-0.6 וכדי לחסוך בסיבוכיות ניתן לבחור $\gamma_1 = 0.0$, כלומר, אות השארית לא עובר כלל צביעה לפני ההתמרה.

לסיכום, נשתמש ב-$\gamma_1 = 0.0$ ו- $\gamma_2 = 1.0$ - ההתמרה מבוצעת על אות שארית לא צבוע והקצאת הסיביות נעשית לפי מעטפת מסנן הסינטזה $1/A(z)$.

6.3.3.2 קוונטיזציה אינפורמצית צד

כאשר אין צובעים את אות השארית לפני ההתמרה ניתן להניח כי בקירוב טוב הוא לבן והווריאנס של כל המקדמים $\hat{\sigma}^2$ הוא אותו ומתקבל על-ידי ווריאנס הדגם של אות השארית.

צעד הקונטיזציה של הקוונטיזרים האחידים ייקבע לכן לפי $\hat{\sigma}$ וגודל זה

יש לשדר כאינפורמצית צד.

ממדידות מתקבל כי ניתן להניח בקירוב טוב כי $\hat{\sigma} < 256$ כאשר אות הדיבור נדגם

ב-12 bit לכן $\hat{\sigma}$ מוכפל ב-8 למלוא התחום $0 \div 2048$ ומקודד ב-6 סיביות כפי שמקודד

SF (ראה סעיף 5.5).

אינפורמצית צד נוספת הם מקדמי מסנן ה-LPC שמשודרים כ-Reflection Coefficients
לפי התקן של LPC-10 (נספח ב').

כאשר מבצעים את הקוונטיזציה לאינפורמצית הצד מקבלים הרעה מורגשת באיכות בקצב
9.6 Kbps . בקצב 16 Kbps לא ניתן להרגיש בהבדל.

ב-SNRSEG השינויים המתקבלים הם:

עבור N=128 ו-100 סיביות מוקצות למסגרת (~9.6Kbps) מתקבל SNRSEG = 11.3 dB במקום
12.2 dB ללא קוונטיזציה.

עבור N=128 ו-200 סיביות מוקצות למסגרת (~16Kbps) מתקבל SNRSEG = 16.9 dB במקום
18.1 dB ללא קוונטיזציה.

<u>הערה</u>: כשאנו מציינים שמוקצות n סיביות למסגרת הכוונה רק לסיביות הנתונות לצורך
קוונטיזציית מקדמי ההתמרה.

משלב זה והלאה כל התוצאות שניתן יכללו גם קוונטיזציה של אינפורמצית הצד.

.6.3.3.3 הקצאת סיביות

כפי שראינו בסעיף 6.3.2.1, לצורך הקצאת הסיביות יש לחשב את הספקטרום של המסנן
$1/A(z)$ ולהשתמש בו כשערוך ל- $\hat{\sigma}(k),\ k = 0 \ldots N-1$. מכיון שהמסנן $A(z)$ מורכב ממספר
קטן של מקדמים. נוח לחשב את הספקטרום של $1/A(z)$ בעזרת DFT מאורך 2N על הסדרה:

$$a'_i = \begin{cases} a_i & i = 0 \ldots p \\ 0 & p < i \leq 2N-1 \end{cases} \qquad (6.3.17)$$

כאשר a_i הם מקדמי המסנן $A(z)$.

מתוך (6.3.11) אנו מחפשים את $\log_2 |\hat{\sigma}(k)|, \quad k = 0 \ldots N-1$

לכן הפעולות שיש לבצע יהיו:

- מצא את הסדרה a'_i על-ידי השלמת a_i באפסים לאורך 2N (ההתמרה היא סימטרית ואותנו מעניינות N הנקודות הראשונות).
- מצא את הספקטרום של a'_i בעזרת DFT:

$$|A(k)| = |DFT(\underline{a}')| \qquad k = 0 \ldots N-1 \tag{6.3.18}$$

ו- $$-\log_2 |\underline{A}| \tag{6.3.19}$$

יתן את הוקטור המבוקש $\log_2 |\hat{\underline{\sigma}}|$.

מעשית, אין צורך לחשב את ה-DFT בכל נקודה של a'_i אלא די להסתפק ב-J נקודות $N >> J$ בהן יחושב $-\log_2 |A(k)|$ ולבצע אינטרפולציה ליניארית ב-log domain. דבר זה מתאפשר מכיון שהמעטפת המתקבלת משמשת לצורך הקצאת סיביות והאינפורמציה החשובה היא הצורה הכללית של המעטפת.

בטבלה 6.5 מפורטים ערכי ה-SNRSEG המתקבלים עבור מספר שונה של נקודות בהן מחושב ה-DFT:

נקודות חישוב	חישוב מדויק	65	33	17	9	5
SNRSEG [dB] N=128, M=128	11.3	11.3	11.3	11.2	10.2	8.7
SNRSEG [dB] N=64, M=256	13.2	-	13.2	12.9	12.3	10.3

טבלה 6.5 ערכי SNRSEG עבור אינטרפולצית log הספקטרום

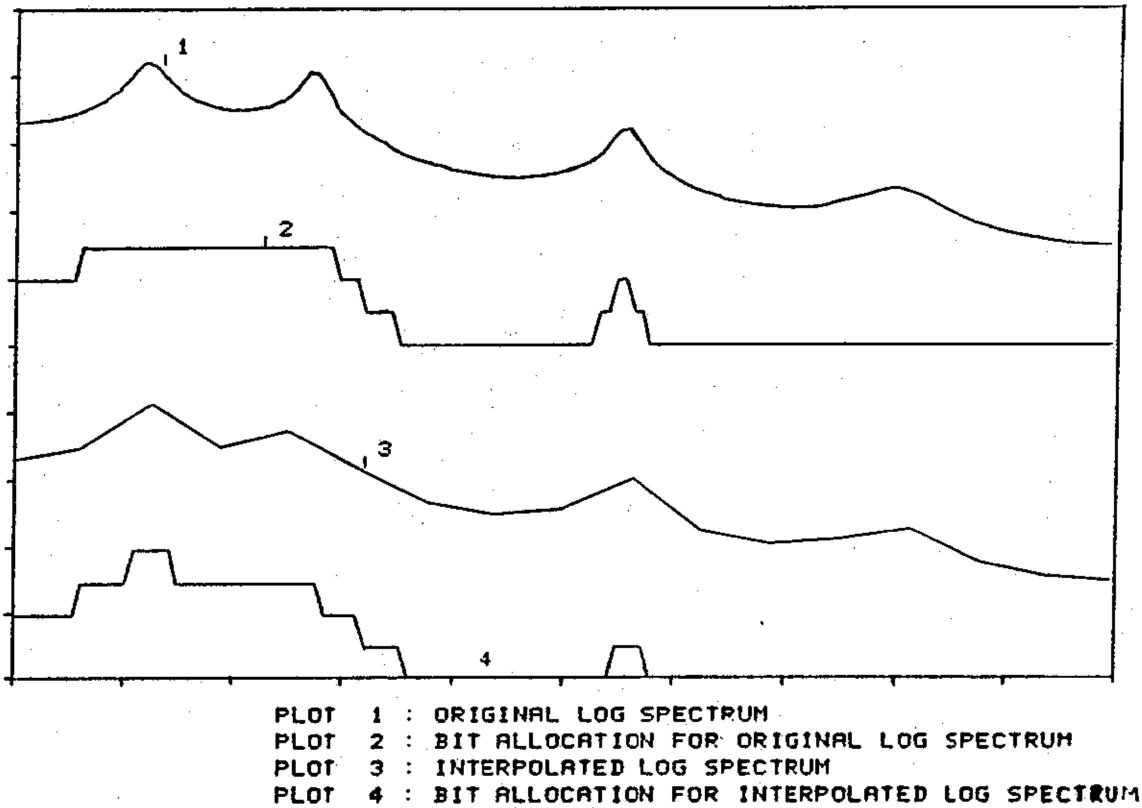
Table 6.5 SNRSEG obtained with interpolation of log Spectrum

הערות לטבלה:

- N - גודל המסגרת לה עושים התמרה.
  M - גודל אינטרוול העדכון בחישוב מקדמי ה-LPC.

- מספר הסיביות המוקצה בכל מקרה מתאים לקצב אופיני של 9.6Kbps.
  עבור N=128, M=128 מוקצות 100 סיביות למסגרת.
  עבור N=64, M=256 מוקצות 59 סיביות למסגרת.
- מקסימום סיביות מוקצות למקדם - 5.
- אינפורמצית הצד עוברת קוונטיזציה.

במבחני שמיעה נמצא שעבור התנאים הנ"ל ניתן לרדת עד ל-17 נקודות חישוב DFT (16 אינטרוולים לאורך הספקטרום) מבלי להרגיש הרעה בביצועים. בציור 6.2 מושווה הספקטרום המתקבל ללא אינטרפולציה עם זה המתקבל עם אינטרפולציה וניתנת הקצאת הסיביות בכל מקרה.



ציור 6.2 - log הספקטרום עם ובלי אינטרפולציה והקצאת סיביות

Fig. 6.2 log spectrum with and without linear interpolation & Bit allocation

מטרים נוספים הקשורים בהקצאת הסיביות הם:

- גודל המסגרת עליה מבוצעת התמרה

התמקדנו כאן בגדלי מסגרות שהן חזקות של 2 כדי להבטיח פשטות אלגוריתמי ההתמרות. גודל מסגרת העדכון באנליזת ה-LPC היא כפולה שלמה של גודל המסגרת במערכת. לדוגמא, אם מקדמי ההתמרה מחושבים כל 128 דגימות מקדמי ה-LPC יחושבו כל 256 או 128 דגמים.

ערכי ה-SNRSEG ב-dB המתקבלים עבור גדלי מסגרות שונים מסוכמים בטבלה 6.6:

גודל מסגרת / צעד עדכון מקדמי ה-LPC	32	64	128	256
128	9.9	11.1	11.1	-
192	11.2	12.5	-	-
256	11.7	12.9	12.4	11.3

טבלה 6.6 ה-SNRSEG המתקבל עבור גדלי מסגרות שונים

Table 6.6 SNRSEG obtained for different frame size

שימוש במסגרות קטנות יקטין את כמות הסיביות הנתונות "נטו" לצורך קידוד מקדמי ההתמרה, שימוש במסגרות גדולות יגרום לחוסר "עידון" במסגרות בהן אות הדיבור מורכב, לדוגמא, במעבר מקטע קולי לאל-קולי.

במבחני שמיעה נמצא שגודל אינטרוול עדכון של 192 או 256 דגמים באנליזת ה-LPC עדיף על אינטרוול עדכון של 128 דגמים. וגודל מסגרת באלגוריתם של 64 או 128 דגמים נותן את התוצאות המיטביות. נשתמש במערכת המומלצת באינטרוול עדכון מקדמי LPC - M=192 וגודל מסגרת באלגוריתם N=64. ערכי ה-SNR המתקבלים עבור M=192 אמנם נמוכים מאלה המתקבלים עבור M=256 אולם ההבדל באיכות איננו מורגש ושימוש באינטרוול עדכון (M=256) הזהה לאינטרוול האנליזה (m=256) כשאנו משתמשים בחלון Hanning, גורם לכך שחלק מאות הדבור איננו בא לידי ביטוי בקביעת מקדמי ה-LPC. שימוש באינטרוול גדול (M=256) עלול לגרום גם לשיחזור פחות מדויק של קטעים מסובכים, לא סטציונריים, כמו קטעי מעבר, דבר שיפגע במובנות האות המשוחזר.

2) <u>המספר המקסימאלי של סיביות המוקצות למקדם</u>

במבחני האזנה נמצא שמקסימום של 5 סיביות למקדם בהתמרה הוא מספיק לתחום הקצבים בו אנו פועלים (9.6 Kbps - 16 Kbps) ירידה מתחת לערך זה פוגעת באיכות המתקבלת מכיון שהמקדמים החשובים לא מקבלים מספיק סיביות, עליה מעל ערך זה לא משפרת בהרבה מכיון שאין הרבה מקדמים שמקבלים הקצאה עשירה יותר לפי חוק הקצאת הסיביות שהגדרנו ב-(6.3.11).

.3 <u>מספר הסיביות הכולל המוקצה למסגרת לצורך קידוד מקדמי ההתמרה</u>

גודל זה קובע את קצב השידור והוא בעל השפעה מכרעת על ביצועי המערכת.
בציור 6.3 ניתן לראות את ההשפעה של מספר הסיביות על ה-SNR המתקבל.
הפרמטרים ששמשו במערכת של ציור 6.3 היו: $N = 64$ ואינטרוול עדכון LPC - $M = 192$.

נראה כמה סיביות נצטרך להקצות עבור קצב של 9.6 Kbps ו-16 Kbps עבור הפרמטרים הנ"ל:

9.6 Kbps: בשניה של דבור יהיו 41.6667 עדכונים של מקדמי המסנן, לכן נזדקק ל-1708.3333 סיביות לקידוד מקדמי ה-LPC. כך יוותרו לנו 63.1333 סיביות לכל מסגרת, מתוכם החלוקה תהיה:

54 סיביות לקידוד מקדמי ההתמרה.

6 סיביות לקידוד $\hat{\sigma}$.

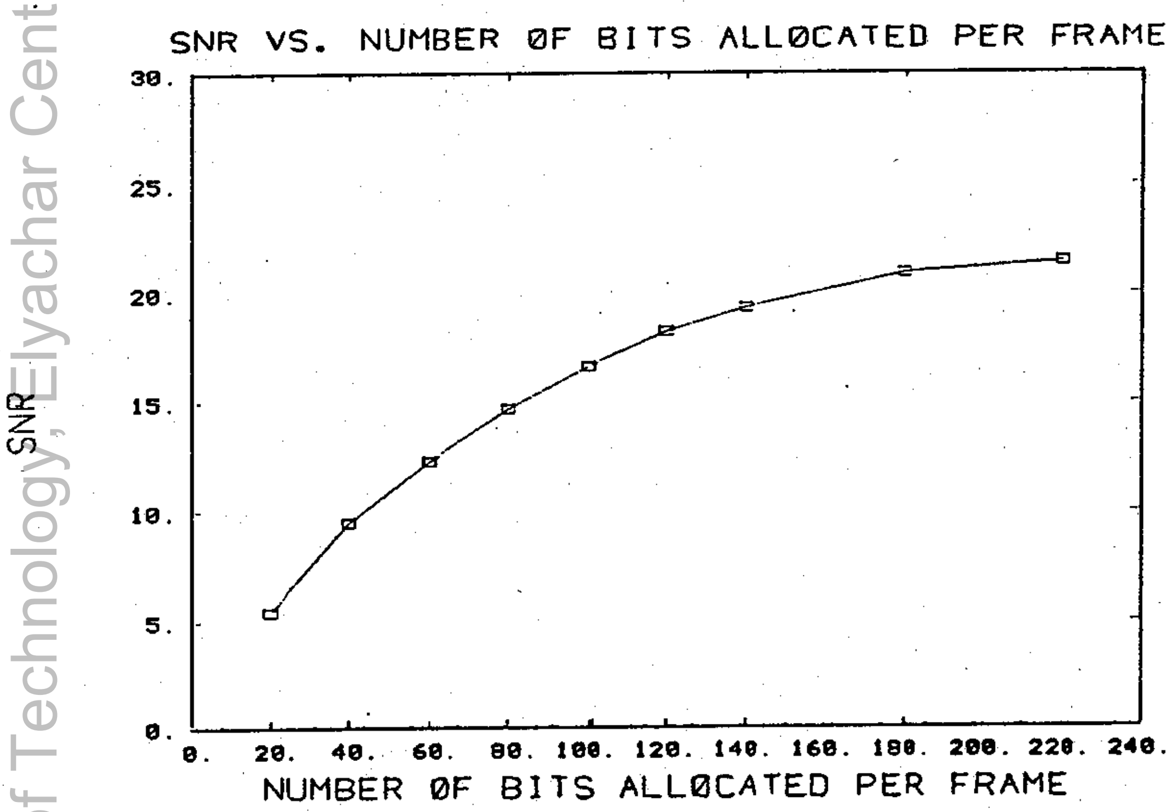
3.1333 סיביות להגנה וסנכרון כפי שנפרט בפרק הבא.

16 Kbps: גם כאן נזדקק ל-1708.3333 סיביות לקידוד מקדמי ה-LPC.
יוותרו לנו 114.3333 סיביות לכל מסגרת, מתוכם החלוקה תהיה:

106 - סיביות לקידוד מקדמי ההתמרה.

6 - סיביות לקידוד $\hat{\sigma}$.

2.3333 סיביות להגנה וסנכרון.



ציור 6.3 - SNR מתקבל ב-PTC כפונקציה של מספר הסיביות המוקצות למסגרת

Fig. 6.3 SNR in PTC vs. Number of bits allocated per Frame

6.3.4 סיבוכיות משדר - מקלט

בסעיף זה נחשב את הסיבוכיות הדרושה לצורך מימוש משדר ומקלט מטיפוס PTC, החישובים המפורטים של הסיבוכיות מובאים בנספח ד', כאן מובאים רק סיכומי התוצאות.

נתון:

m - גודל הבלוק עליו מחושבים מקדמי אוטוקורלציה באנליזת LPC.

M - אינטרוול עדכון מקדמי ה-LPC.

N - גודל מסגרת באלגוריתם (M כפולה שלמה של N).

B - מספר הסיביות הנתון לקידוד המקדמים במסגרת.

n_b - מספר הסיביות המקסימאלי הנתון לקידוד מקדם יחיד בהתמרה.

P - סדר המסנן.

J - מספר הנקודות בהן מחושב הספקטרום לצורך חישוב המעטפת להקצאת הסיביות, ראה סעיף 6.3.3.3 .

פעולות המתבצעות במשדר

1) חישוב Reflection Coefficients כל M דגמים - $O(m(P+2)+P^2)$
כאשר m הוא גודל הבלוק עליו מחושבים מקדמי אוטוקורלציה.

2) מציאת אות השארית e מתוך האות המקורי S - כרוך בסינון S - O(NP) פעולות.

3) חישוב המעטפת הספקטרלית: $O(2PJ + 7J + N + \frac{P(P-1)}{2})$ פעולות.

4) הקצאת סיביות: O(9N) פעולות + פעולת חילוק.

5) ביצוע DCT לפי האלגוריתם של WANG [42] ידרוש $O(2N\log_2 N-N)$ פעולות.

6) קוונטיזצית מקדמי ההתמרה לפי הקצאת הסיביות:
O(2N) פעולות + פעולת חילוק.

בסך הכל נזדקק במסגרת אחת ל:

פעולות: $O(N/M(m(p+2) + p^2) + NP + 11N + 2N\log_2 N + 2PJ + 7J + P(P-1)/2)$
+
2 פעולות חילוק.

<u>פעולות המתבצעות במקלט</u>

1. חישוב מעטפת ספקטרלית: זהה למשדר - $O(2PJ + 7J + N + \frac{P(P-1)}{2})$ פעולות.

2. הקצאת סיביות: זהה למשדר - $O(9N)$ פעולות + פעולת חילוק.

3. שיחזור ההתמרה $\underline{\hat{E}}$: יש לפרק את ה-bit stream המשודר לפי הקצאת הסיביות ולנרמל חזרה לפי $\hat{\sigma}$:

$O(N)$ פעולות לפרוק.

$O(N)$ פניות לטבלות קוונטיזציה.

$O(N)$ פעולות לנרמול.

סה"כ: $O(3N)$ פעולות.

4. ביצוע IDCT : סיבוכיות זהה ל-DCT - $O(2N\log_2 N-N)$ פעולות.

5. שחזור $\underline{\hat{S}}$: העברת $\underline{\hat{e}}$ דרך מסנן הסינטזה: $O(NP)$ פעולות.

בסה"כ נקבל:

פעולות למסגרת: $O(NP + 12N + 2N\log_2 N + 2PJ + 7J + \frac{P(P-1)}{2})$

+

פעולת חילוק אחת.

נסכם בטבלה את הסיבוכיות הדרושה עבור מערכת בעלת הפרמטרים הבאים:

אינטרוול עדכון לחישוב מקדמי LPC	$M = 192$
גודל הבלוק לחישוב מקדמי אוטוקורלציה	$m = 256$
סדר מסנן	$P = 10$
גדלי מסגרת באלגוריתם	$N=64$, $n_b = 5$, $J=17$

N	סיביות מוקצות למסגרת	פעולות משדר	פעולות מקלט
64	54	58	42

טבלה 6.7 – סיבוכיות משדר – מקלט PTC

Table 6.7 Complexity PTC Vocoder

<u>הערות:</u>

- הקצאת הסיביות הנתונה בדוגמאות בטבלה 6.7 מתאימה לקצב של 9.6 kbps .
- פעולות החילוק מוזנחות יחסית לשאר הפעולות הנדרשות.
- הסיבוכיות איננה תלויה בקצב השידור! לכן ערכי הסיבוכיות הנתונים מתאימים גם לקצב 16 Kbps .

6.3.5 מערכת PTC מומלצת

לאור כל הנאמר בסעיפים האחרונים המערכת המומלצת הן מבחינת ביצועים והן מבחי
סיבוכיות תהיה מערכת בעלת הפרמטרים הבאים:

- $m = 256$ - גודל הבלוק עליו מחושבים מקדמי האוטוקורלציה באנליזת ה-LPC.
- $M = 192$ - אינטרוול העדכון של מקדמי ה-LPC.
- $N = 64$ - גודל המסגרת עליה מבוצעת התמרה.
- B - מספר הסיביות המוקצה למסגרת:
  - $B = 54$ - עבור מערכת המיועדת לעבוד בקצב 9.6 Kbps .
  - $B = 106$ - עבור מערכת המיועדת לעבוד בקצב 16 Kbps .
- $n_b = 5$ - המספר המקסימאלי של סיביות המוקצה למקדם.
- $J = 17$ - מספר הנקודות בספקטרום $\frac{1}{A(z)}$ המחושבות לצורך הקצאת הסיביות.

- אות השארית איננו עובר צביעה והקצאת הסיביות נעשית לפי מעטפת $\frac{1}{A(z)}$.

- הסיבוכיות המתקבלת מובאת בטבלה 6.7 .
- ערכי ה-SNR המתקבלים:

(9.6 Kbps   SNR = 11.9dB   SNRSEG = 12.5dB .

(16 Kbps   SNR = 17.2dB   SNRSEG = 18.0dB .

- ב-16 Kbps האיכות המתקבלת שקופה.
- ב-9.6 Kbps מבחינים בתופעה של Low-pass .

פרק 7. ביצועי המערכות והשוואה

7.1 מבוא

לאחר שהצגנו בפרק 5 את המערכת האיטרטיבית (Multi-Pulse LPC) ובפרק 6 את מערכת ה-PTC נבחן בפרק זה את ביצועי שתי המערכות מהבחינות הבאות:

(i) חבור מערכות בטור (Tandeming).
(ii) ביצועים בהנתן רעש רקע.
(iii) חסינות לרעש ערוץ (Bit Error Rate - BER).
(iv) ביצועים בהעברת אותות נתונים (Voice Band Data Signal).
(v) השוואה ל-PCM.

בסוף הפרק נסכם את כל הנתונים יחד עם נתונים לגבי איכות הדיבור המתקבלת וסיבוכיות בתחום הקצבים בו עסקנו (9.6 - 16 Kbps).

7.2. המערכות המושוות

ההשוואה בין המערכות תעשה במקביל. כלומר, כל אחד מהסעיפים הבאים יעסוק באחד מהנושאים הנבדקים וביצועי המערכות יוצגו ביחד. בסעיף זה נציג את נתוני המערכות המושוות:

מערכת ה-Multi-Pulse LPC

המערכת בה נשתמש היא בעלת הנתונים הבאים:

- אינטרוול עדכון מקדמי ה-LPC : M=180
- גודל הבלוק לאנליזת ה-LPC : m=256
- סדר מסנן LPC : p=10
- גודל מסגרת באלגוריתם : N=45
- מספר פולסים נבחרים למסגרת :
  K=4 עבור קצב 9.6 Kbps
  K=9 עבור קצב 16 Kbps

- מספר הדגמים השונים מ-0 ב-h'_n: L=10
- $\gamma = 0.8$.
- קידוד מקדמי LPC, הגנה וסינכרון: בבדיקות נמצא שדי להגן על מקדמי ה-LPC כדי לקבל שיפור משמעותי בחסינות לרעשי ערוץ (סעיף 7.5). הגנה על ה-Scale (SF) Factor לא תורמת באופן משמעותי. מקדמי ה-LPC מקודדים לפי הסטנדרט של LPC-10 (נספח ב') לכן עלינו להגן על 41 סיביות. לפי הטבלה הנתונה ב-[49, pp. 675] נזדקק ל-6 סיביות אם נשתמש בקוד הנתון ב-[50].

אינטרוול עדכון מקדמי ה-LPC הוא 180 דגימות לכן המקדמים מתעדכנים 44.444 פעם בשניה. נניח שמספיקה סיבית אחת למסגרת עדכון מקדמים לצורך סינכרון ונקבל בסה"כ:

אינפורמצית מקדמים :	1822.2222	סיביות לשניה.
הגנה :	266.6666	סיביות לשניה.
סינכרון :	44.4444	סיביות לשניה.
סה"כ :	2133.3333	סיביות לשניה.

- קידוד אינפורמצית הערור:

בקצב 9.6 Kbps : K=4, נזדקק ל-5 סיביות לפולס לקידוד הפרשי מרחקים בין הפולסים. כמו כן נזדקק ל-6 סיביות למסגרת לקידוד SF ו-4 סיביות לפולס לקידוד האמפליטודה, סה"כ 7466.6666 סיביות לשניה. יחד עם מקדמי LPC: 9.6 Kbps.
בקצב 16 Kbps : K=9, נזדקק ל-4 סיביות לפולס לקידוד הפרשי מרחקים בין הפולסים. גם כאן נזדקק ל-6 סיביות למסגרת לקידוד SF ו-4 סיביות לפולס לקידוד האמפליטודה, סה"כ 13866.6666 סיביות לשניה. יחד עם מקדמי LPC: 16 Kbps.

<u>PTC</u>

המערכת בה נשתמש היא בעלת הנתונים הבאים:

- אינטרוול העדכון של מקדמי ה-LPC : M=192
- גודל הבלוק לאנליזת ה- LPC : m=256
- סדר מסנן LPC : P=10
- גודל המסגרת עליה מבוצעת התמרה : N=64
- מספר הסיביות המוקצה לקידוד מקדמי ההתמרה :

54 סיביות עבור קצב 9.6 Kbps

106 סיביות עבור קצב 16 Kbps

- מספר מקסימאלי של סיביות למקדם: $n_b=5$

- מספר הנקודות המחושבות בספקטרום $1/A(z)$ לצורך הקצאת הסיביות: $J=17$ (16 אינטרוולים).

- קידוד מקדמי LPC, הגנה וסינכרון:
אינפורמצית ה-LPC משודרת כפי שנעשה במקודד ה-Multi-Pulse המומלץ: 41 סיביות למקדמים + 6 סיביות הגנה, סיבית למסגרת עדכון LPC לסנכרון.

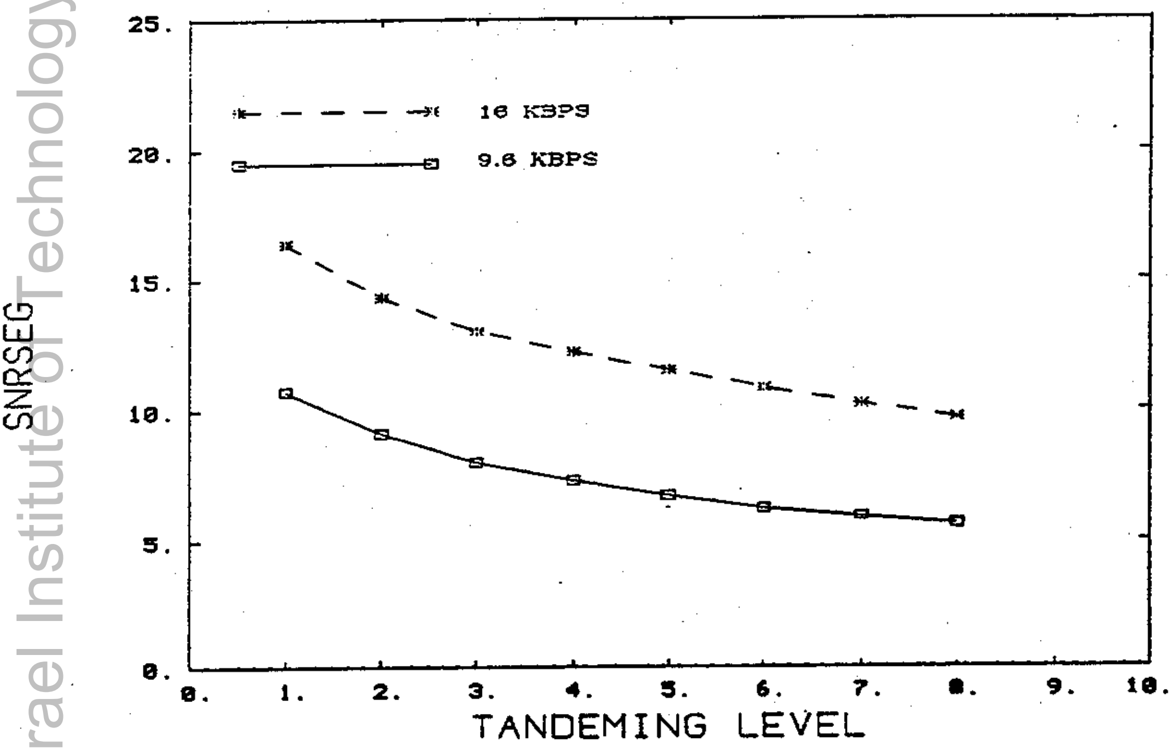
- קידוד אינפורמצית הערור: כפי שהוסבר בפרק 6.

7.3 חבור מערכות בטור - Tandeming

ברשתות תקשורת בהן מעבירים אותות דיבור מקודדים יתכן מצב בו האות המקורי יעבור דרך כמה צמתים ברשת בהן הוא יעבור שחזור וקידוד מחדש על-ידי מקודדים שונים. חשוב לכן לבדוק כיצד תיפגע איכות האות המתקבל לאחר מעבר דרך כמה שלבים של שחזור וקדוד על-ידי המקודדים המוצעים.

בציור 7.1 מובאים ערכי ה-SNRSEG המתקבלים עבור קצבים של 9.6 Kbps ו-16 Kbps כשהאות עובר כמה רמות Tandeming במערכת האיטרטיבית (Multi-Pulse).

בקצב 9.6 Kbps האות נפגע בצורה חריפה החל מ-5 רמות של Tandeming. ב-16 Kbps האיכות עדיין סבירה גם אחרי 8 רמות. אופי העיוות הוא של "חריקות" וצפצופים. כדאי לשים לב שה-SNRSEG המתקבל עבור 7 צמתות ב-16 Kbps משתווה ל-SNRSEG המתקבל עבור רמת Tandeming אחת ב-9.6 Kbps .

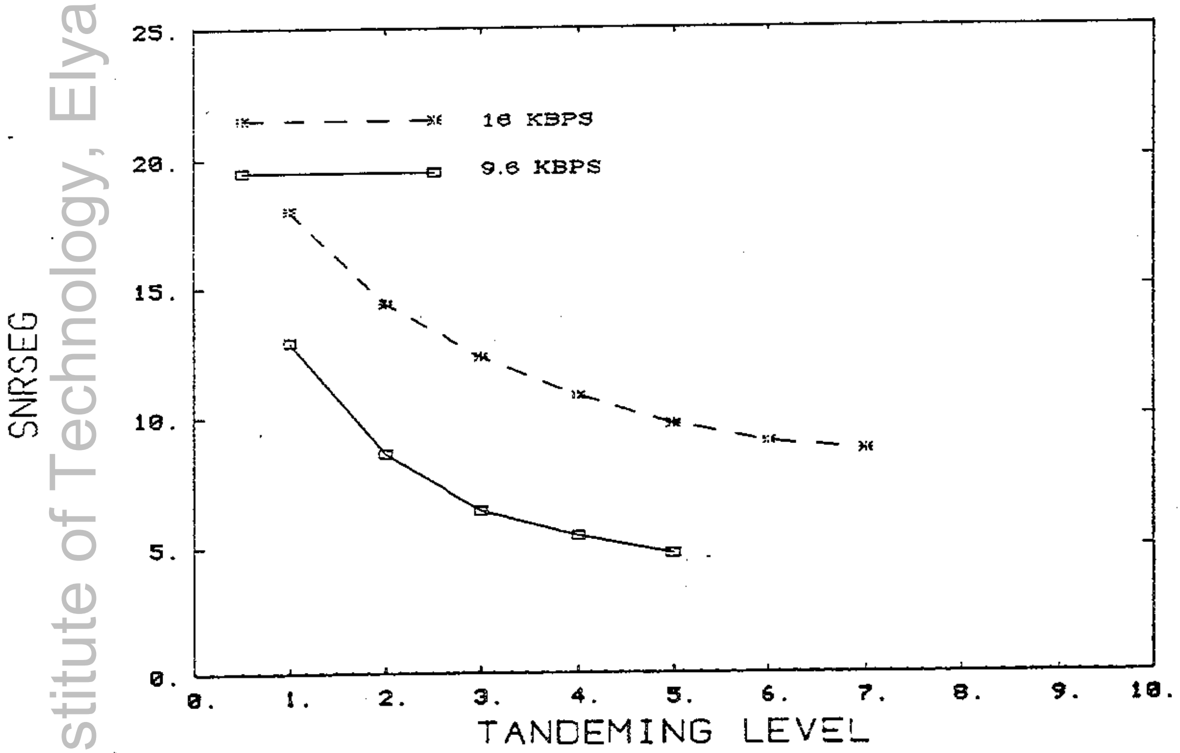


ציור 7.1 SNRSEG כפונקציה של Tandeming במערכת ה-Multi-Pulse

Fig. 7.1 SNRSEG vs. Tandeming in Multi-Pulse system

בציור 7.2 מובאים ערכי ה-SNRSEG המתקבלים עבור קצב של 9.6 Kbps ו-16 Kbps כשהאות עובר כמה רמות Tandeming במערכת ה-PTC.

בדיקות סוביקטיביות מראות כי בקצב 9.6 Kbps האיכות נפגמת בצורה חריפה כבר אחרי 3 רמות Tandeming, ב-16 Kbps האיכות יורדת במתינות ואופי ההפרעות הוא מקומי, האיכות סבירה עד 4 רמות Tandeming אך ניתן להבין את הנאמר גם אחרי 7 רמות.



ציור 7.2 SNRSEG כפונקציה של Tandeming ב-PTC

Fig. 7.2 SNRSEG vs. Tandeming in PTC Vocoder

7.4 ביצועים בתוספת רעש רקע

ביצועי מקודדי LPC נפגעים כאשר לאות המקורי מתווסף רעש רקע. נבחן לכן את ביצועי המערכת שהוצגו בתוספת רעש רקע. הרעש שהוסף היה רעש גאוסי לבן ב-4 דרגות: 5dB, 10dB, 15dB ו-25dB. התוצאות עבור מערכת ה-Multi-Pulse מסוכמת בטבלה 7.1 :

Input SNR [dB]	9.6 Kbps SNR [dB]	9.6 Kbps SNRSEG [dB]	16 Kbps SNR [dB]	16 Kbps SNRSEG [dB]
25	9.7	9.9	15.4	14.9
15	8.9	7.7	12.8	10.8
10	7.2	5.5	9.7	7.5
5	5.1	3	5.8	3.6

טבלה 7.1 ביצועי מערכת Multi-Pulse בנוכחות רעש רקע

Table 7.1 Multi-Pulse performance with background noise

ב-9.6 Kbps האיכות המתקבלת עבור רעש ברמה של 5 dB היא גרועה מאד. אופיו של הרעש באות המשוחזר אינו לבן ומופיע רעש דמוי "קליקים".
ב-16 Kbps המובנות נשמרת בכל רמות הרעש שנבחנו ואופיו של הרעש נשאר "לבן" כמו באות המקורי.

גם עבור מערכת ה-PTC הוסף רעש בדרגות הנ"ל והתוצאות שהתקבלו מסוכמות בטבלה 7.2. בקצב 16 Kbps האות המשוחזר נשמע בדיוק כמו אות המקור הרועש והרעש המשוחזר נשמע "לבן". בקצב 9.6 Kbps הרעש המשוחזר כבר אינו נשמע "לבן" וברמת רעש של 5 dB האיכות המתקבלת גרועה ביותר.

Input SNR [dB]	9.6 Kbps SNR [dB]	9.6 Kbps SNRSEG [dB]	16 Kbps SNR [dB]	16 Kbps SNRSEG [dB]
25	12.4	12.6	17.4	16.6
15	11.0	9.5	13.0	11.3
10	8.8	6.7	9.5	7.4
5	5.4	3.1	5.3	3.3

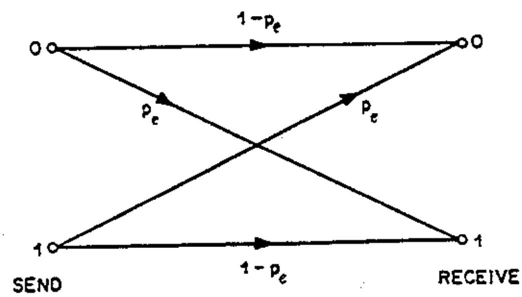
טבלה 7.2 ביצועי מערכת PTC בנוכחות רעש רקע

Table 7.2 PTC performance with background noise

7.5 חסינות לרעשי ערוץ

מוצא המשדר במקודדי דיבור מועבר אל ערוץ ספרתי. העברת האינפורמציה על-פני הערוץ עלולה להיות לא מושלמת, כלומר, כתוצאה מרעשים יתכנו שגיאות בהעברת האינפורמציה. ניתן להתיחס לשני סוגי שגיאות: שגיאות בודדות שהן מפולגות בינומיאלית עבור ערוץ בינארי סימטרי ושגיאות Burst המאופיינות בקטעים שגויים. כמות השגיאות בערוץ נמדדת ב-(BER) Bit Error Rate והיא יכולה להשתנות בין ערכים קטנים מ-10^{-6} במערכות בעלות איכות גבוהה לערכים גדולים מ-10^{-2} בסביבות רועשות כמו לדוגמא ב-Mobile Radio.

בסעיף זה נתייחס למקרה הפשוט של שגיאות בסיביות בודדות (לא Burst) כשהערוץ הוא ערוץ בינארי סימטרי חסר זיכרון בעל הסתברות שגיאה לסיבית - P_e. הסתברות השגיאה היא אותה לשתי הכניסות האפשריות ("0" ו-"1") לערוץ.



ציור 7.3 ערוץ בינארי סימטרי בעל הסתברות שגיאה לסיבית - P_e

Fig 7.3 Binary Symmetric channel with error probability P_e

שגיאות הערוץ פוגעות כמובן באיכות הדיבור המשוחזר המתקבל במקלט לכן במימוש של מקודדים מוסיפים לעיתים קרובות הגנת שגיאות (error protection). תוספת זו כרוכה בתוספת אינפורמציה משודרת, אולם, היא מאפשרת לגלות ו/או לתקן מספר מסוים של שגיאות. עבור ערוצים בעלי P_e גבוה תוספת כזו עשויה לגרום לשיפור משמעותי באיכות האות המתקבל.

לא נכנס במסגרת זו לדיון על שיטות לתיקון שגיאות, מקור טוב לדיון כזה הוא [49]. אנו נסתפק בקודים מתקני שגיאה אחת, כלומר, תהיה לנו האפשרות לתקן שגיאה אחת בתוך בלוק האינפורמציה עליו נגן. לא נטפל בתיקון של יותר שגיאות בגלל הסיבוכיות הגדולה יותר הכרוכה בקודים כאלה, הצורך להקצות יותר סיביות להגנה ובגלל ההסתברות הנמוכה יותר להפיכה של שתי סיביות או יותר בבלוק מוגן.

כדי לבחון את ההבדלים בין מערכת עם הגנה למערכת בלי הגנה נציג עבור כל אחד מהמקודדים מערכת שאינה כוללת הגנה בנוסף למערכות המוגנות שהוצגו בסעיף 7.2 .

Multi-Pulse LPC

מערכת ללא הגנה:

- אינטרוול עדכון מקדמי ה-LPC : M=180
- גודל מסגרת באלגוריתם : N=90
- מספר פולסים למסגרת :

	9.6 Kbps	16 Kbps
מספר פולסים	K = 9	K=19
סיביות אינפורמציה	9556	15867
סיביות סינכרון	44	133
סה"כ	9600 bit	16000 bit

מכיון שבמערכת ללא הגנה מוקצות יותר סיביות לשידור האינפורמציה נצפה לקבל תוצאות טובות יותר עבור BER=0. בטבלאות 7.3 ו-7.4 מסוכמות התוצאות המתקבלות עבור BER שונים בקצבי שידור 9.6 Kbps ו-16 Kbps בהתאמה.

עבור קצב 9.6 Kbps במערכת ללא הגנה האיכות ב-$BER=10^{-4}$ היא שקופה אך עבור BER גדול יותר כבר מרגישים בפגיעה באיכות וב-$BER=10^{-2}$ האיכות כבר גרועה ביותר.
במערכת עם ההגנה מקבלים איכות שקופה גם עבור $BER=10^{-3}$ וב-$BER=10^{-2}$ האיכות המתקבלת טובה באופן בולט לעומת המערכת ללא הגנה.
נסיון להגן גם על ה-Scale Factor (SF) שהוגדר בפרק 5 לא משפר בהרבה את התוצאות ולכן אנו מסתפקים בהגנה על ה-k-parameters בלבד.

עבור 16 Kbps ערכי ה-SNR המתקבלים הם כמובן טובים יותר אך התוצאות הסוביקטיביות זהות לאלו שהתקבלו עבור 9.6 Kbps כלומר: במערכת ללא הגנה לא מרגישים בפגיעה רק עבור $BER=10^{-4}$, במערכת עם הגנה מרגישים בשיפור עבור כל ערכי ה-BER כשאיכות שקופה נשמרת גם עבור $BER=10^{-3}$.

BER	With Protection		No Protection	
	SNR [dB]	SNRSEG [dB]	SNR [dB]	SNRSEG [dB]
0	9.6	10.6	10.0	11.2
10^{-4}	9.6	10.6	9.7	11.1
10^{-3}	8.5	10.1	8.8	9.9
10^{-2}	2.7	5.9	1.1	3.1

טבלה 7.3 ביצועי Multi-Pulse בערוץ רועש בקצב 9.6 Kbps

Table 7.3 Multi-Pulse performance with noisy channel at 9.6 Kbps

BER	With Protection		No Protection	
	SNR [dB]	SNRSEG [dB]	SNR [dB]	SNRSEG [dB]
0	15.2	16.1	15.5	16.1
10^{-4}	14.5	15.8	15.1	16.0
10^{-3}	11.8	14.6	8.8	13.5
10^{-2}	1.3	5.4	0.4	3.1

טבלה 7.4 ביצועי Multi-Pulse בערוץ רועש בקצב 16 Kbps

Table 7.4 Multi-Pulse performance with noisy channel at 16 Kbps

PTC

ההגנה במערכת ה-PTC נעשית כפי שהיא נעשתה במערכת האיטרטיבית: 6 סיביות מגינות על 41 הסיביות של אינפורמצית ה-k-parameters. במערכת ללא הגנה ניתן לנצל את הסיביות העודפות לקידוד אינפורמציה ממשית ולכן נקבל במערכת ללא הגנה - 57 סיביות לקידוד מקדמי ההתמרה למסגרת בקצב שידור 9.6 Kbps ו-108 סיביות בקצב 16 Kbps , לעומת 54 ו-106 סיביות בהתאמה במערכת עם הגנה. בטבלאות 7.5 ו-7.6 מסוכמות התוצאות המתקבלות עם ובלי הגנה בקצבים 9.6 Kbps ו-16 Kbps בהתאמה.

בקצב 9.6 Kbps ללא הגנה האיכות שקופה ב-$BER=10^{-4}$. עבור $BER=10^{-3}$ האיכות נפגמת ושומעים הפרעות באות. עבור $BER=10^{-2}$ האיכות גרועה ביותר. עם תוספת הגנה כמעט ולא מרגישים כלל בשגיאות גם עבור $BER=10^{-3}$ וב-$BER=10^{-2}$ ניתן להבין את הנאמר למרות שמרגישים בשגיאות.

בקצב 16 Kbps ללא הגנה האיכות שקופה ב-$BER=10^{-4}$, מורגשת הרעה ב-$BER=10^{-3}$ וב-$BER=10^{-2}$ האיכות כבר מאד ירודה למרות שניתן להבין את הנאמר. במערכת עם הגנה האיכות שקופה גם עבור $BER=10^{-3}$. עבור $BER=10^{-2}$ מקבלים איכות הרבה יותר טובה לעומת זו המתקבלת ללא הגנה למרות שמרגישים בשגיאות.

BER	With Protection		No Protection	
	SNR [dB]	SNRSEG [dB]	SNR [dB]	SNRSEG [dB]
0	11.9	12.5	12.2	12.9
10^{-4}	11.9	12.5	12.2	12.9
10^{-3}	10.5	11.9	10.1	11.8
10^{-2}	5.0	8.4	0.03	3.7

טבלה 7.5 ביצועי PTC בערוץ רועש בקצב 9.6 Kbps

Table 7.5 PTC perform nce with noisy channel at 9.6 Kbps

BER	With Protection		No Protection	
	SNR [dB]	SNRSEG [dB]	SNR [dB]	SNRSEG [dB]
0	17.2	18.0	17.3	18.1
10^{-4}	16.1	17.7	17.3	18.0
10^{-3}	13.2	16.9	8.6	15.3
10^{-2}	5.9	10.2	0.7	5.7

טבלה 7.6 ביצועי PTC בערוץ רועש בקצב 16 Kbps

Table 7.6 PTC performance with noisy channel at 16 Kbps

7.6 ביצועים בהעברת אותות נתונים (VBD)

בנוסף לתפקידם המקורי - העברת אותות דיבור, הולך ונפוץ כיום השימוש במקודדי
דיבור גם להעברת אותות MODEM, כלומר: מעבירים אותות נתונים (Data) בעלי אותו
רוחב פס כמו אותות דיבור ((Voice Band Data (VBD)) בעזרת מקודדי דיבור. דבר זה
קורה בדרך כלל ברשתות תקשורת המעבירות הן אינפורמצית דיבור והן אינפורמצית
Data.

ניסינו לכך את המערכות המוצעות גם בהעברת אותות נתונים. אות ה-MODEM בו השתמשנו
היה בקצב של 2400 bps , גל נושא של 1800 Hz ושיטת אפנון QPSK, כלומר, כל זוג
סיביות מקודד על-ידי סיבוס באחת מ-4 פאזות אפשריות.

כיון שמקודד הדיבור מותאם לאותות דיבור אין הבטחה שהוא יתפקד טוב גם עבור אותות
VBD.

Multi-Pulse LPC

מסתבר שכדי לקבל את התוצאות הטובות ביותר יש להשתמש בחלון מלבני בזמן אנליזת
ה-LPC. כמו כן אין צורך לשקלל (לצבוע) את אות השארית. הדבר האחרון ברור מכיון
ששקלול השגיאה פגע למעשה ב-SNR אך התאים לתכונות האוזן.

בשני הקצבים הנדונים התוצאות אינן משביעות רצון ($BER = 10^{-6}$ ייחשב כמשביע
רצון).

בטבלה 7.7 מוצגות התוצאות המתקבלות:

BER ביציאה	קצב
1.53×10^{-3}	9.6 Kbps
4.57×10^{-4}	16 Kbps

טבלה 7.7 ביצועי Multi-Pulse LPC עבור VBD

Table 7.7 Multi-Pulse LPC performance for VBD

<u>הערה</u>: ה-BER המוצג בטבלה הוא ביציאת המקודד <u>ללא</u> שגיאות ערוץ.

התוצאות שהתקבלו עבור מערכת ה-PTC בהעברת אותות VBD מסוכמות בטבלה 7.8 .

קצב	BER ביציאה
9.6 Kbps	8.0×10^{-4}
16 Kbps	1.1×10^{-4}

טבלה 7.8 ביצועי PTC עבור VBD

Table 7.8 PTC performance for VBD

ביטול האינטרפולציה בשערוך הספקטרום תורם אף הוא לשיפור הביצועים אם כי לא בצורה משמעותית: ב-9.6 Kbps התקבל BER של 7.6×10^{-4} במקום 8.0×10^{-4}.

השפעה קריטית על הביצועים יש לגדלים γ_1 ו-γ_2 שהגדרנו בהגדרה הבסיסית של מערכת ה-PTC (פרק 6). במערכת המומלצת לדיבור $\gamma_1=0.0$ ו-$\gamma_2=1.0$. כלומר, אות השארית לא עובר צביעה והקצאת הסיביות נעשית לפי הספקטרום של $1/A(z)$. עבור אותות VBD התקבל שרצוי $\gamma_1=1.0$ ו-$\gamma_2=1.0$, כלומר, לבצע בעצם ATC - האות <u>המקורי</u> עובר טרנספורמציה. בתנאים אלה התקבל בקצב 9.6 Kbps $BER=7.25\times10^{-5}$ ובקצב 16 Kbps התקבל BER=0. מתוצאות אלה נובע שאם קיים אינדיקטור להמצאות אות דיבור או נתונים כדאי לשנות את γ_1 לערך 0.0 או 1.0 בהתאם לאופי האות: $\gamma_1=0.0$ לאות דיבור ו-$\gamma_1=1.0$ לאותות VBD.

7.7 השוואה ל-PCM µ-law

מדד מקובל להשוואת ביצועי מקודדים הוא השוואה לאיכות האות המתקבל על-ידי קידוד µ-law PCM במספר שונה של סיביות.

אופי הפגיעה באיכות הדיבור בשני המקרים הוא כמובן שונה וניתן לשמוע את ההבדל במבחני האזנה.

מבחינה סובייקטיבית מתקבל שעבור שני המקודדים, Multi-Pulse LPC ו-PTC, האיכות המתקבלת בקצב 9.6 Kbps ניתנת להשוואה ל-PCM ב-6 סיביות והאיכות המתקבלת ב-16 Kbps ניתנת להשוואה ל-PCM ב-7 סיביות.

7.8 השוואה בין המערכות המומלצות

התוצאות שהתקבלו בסעיפים הקודמים והתוצאות האובייקטיביות לאיכות האות המשוחזר מסוכמות בטבלה 7.9 .

<u>הערות לטבלה:</u>

- הנתונים בטבלה הם עבור המערכות עם הפרמטרים המומלצים מסעיף 7.2 .

- הערך בעמודת ה-Tandeming הוא רמת ה-Tandeming שמעבר אליה יורד ה-SNRSEG מתחת ל-5 dB עבור קצב 9.6 Kbps ו-10 dB עבור קצב 16 Kbps.

- העמודה של SNR עבור $BER=10^{-3}$ מתיחס למערכות עם הגנת שגיאות.

מבחינת איכות הדבור המתקבלות (בדיקה סובייקטיבית) אזי ב-16 Kbps האיכות המתקבלת ב-PTC טובה מזו המתקבלת ב-MPLPC. ב-9.6 Kbps האיכות של PTC מעט נמוכה מזו של MPLPC בגלל אפקט של low-passing בו ניתן לחוש באות המשוחזר.

PCM μ-law [bit]	BER עבור VBD	SNRSEG[dB] ברעש ערוץ $BER=10^{-3}$	SNRSEG [dB] ברעש רקע 15 dB	Tandeming	סיבוכיות מקלט [פעולות לדגם]	סיבוכיות משדר [פעולות לדגם]	SNRSEG [dB]	SNR [dB]	מקודד
6	1.53×10^{-3}	10.1	7.7	8	10	84	10.6	9.6	MPLPC
6	8.0×10^{-4}	11.9	9.5	4	42	58	12.5	11.9	PTC
7	4.57×10^{-4}	14.6	10.8	8	10	112	16.1	15.2	MPLPC
7	1.1×10^{-4}	16.9	11.3	4	42	58	18.0	17.2	PTC

טבלה 7.9 - השוואת המקודד האיטרטיבי (MPLPC) והמקודד PTC בקצבים 9.6 Kbps ו-16 Kbps

Table 7.9 Comparison of MPLPC and PTC coders at 9.6 Kbps & 16 Kbps

פרק 8. סיכום ומסקנות

במסגרת חבור זה הוצעה שיטה חדשה לייצוג אות השארית מחזוי לינארי לשימושים של קידוד אותות דיבור. השיטה מבוססת על הצגת אות השארית כקומבינציה לינארית של מספר קטן של וקטורים הלקוחים מתוך אוסף נתון. אוסף הוקטורים נמצא הן במקלט והן במשדר והאינפורמציה המשודרת על אות השארית הם אינדקסי הוקטורים שמרכיבים את הקומבינציה וערכי המקדמים הכופלים אותם. ההצגה הזאת מהווה הכללה של מספר מקרים פרטיים שהוצעו בספרות בשנים האחרונות.

מתוך השיטה החדשה נבעו מספר מקודדים אשר מהם התרכזנו בשניים: הראשון מקודד שבו אוסף הוקטורים הוא וקטורי יחידה - וקטורים שרק איבר יחיד מתוכם שונה מ-0. אוסף כזה של וקטורים מוליך לפתרון איטרטיבי פשוט למציאת קומבינצית הוקטורים הדרושה עבור אות השארית. בעית מציאת הקומבינציה היא בעיה קשה וגם פתרון איטרטיבי זה הוא פתרון תת-אופטימאלי. הצגת המקודד הנ"ל מובילה למקודד ה-Multipulse LPC שהוצע בשנים האחרונות.

למקודד השני הגענו בתהליך חיפוש אוסף הוקטורים האופטמאלי הדרוש. למקודד זה קראנו Predictive Transform Coder (PTC), מכיון שהוא בעצם פועל על אות השארית בתחום התדר.

תחום קצבי השידור בו עסקנו היה 16 Kbps - 9.6 Kbps ושני המקודדים נתנו בו תוצאות יחסית טובות כשב-9.6 Kbps המקודד האיטרטיבי נותן איכות מעט טובה יותר (במקודד PTC בקצב זה מורגש אפקט של low-passing) וב-16 Kbps מקודד ה-PTC נותן תוצאות טובות יותר. במסגרת החבור מובאות השוואות אוביקטיביות וסוביקטיביות בין המקודדים ומובאות הערכות סיבוכיות לכל אחד מהם. כמו כן נערכו בדיקות לבחינת ביצועי המקודדים ב-Tandeming, רעש רקע, שגיאות ערוץ והעברת אותות VBD - Voice Band Data.

מקודד ה-PTC שהוצג בחבור נותן סיבוכיות נמוכה מזו של מקודד ה-Multipulse אולם, כאמור, איכותו בקצבים הנמוכים (9.6 Kbps) נופלת מזו של מקודד ה-Multipulse. שיפור איכותו של מקודד זה בקצבים הנמוכים ראוי למחקר נוסף.

כדי לשפר את האיכות ניתן להתמקד בשני כיוונים:

הראשון, לשדר ביתר יעילות את מקדמי מסנן ה-LPC כדי שיותר סיביות ישתחררו לטובת קידוד ההתמרה. טכניקות שונות הוצעו בשנים האחרונות הדורשות פחות סיביות לקידוד המקדמים לעומת השיטה הקלאסית בה השתמשנו בעבודה זו. לדוגמא [54].

כיוון שני יהיה לנסות לשנות את צורת הקצאת הסיביות הקלאסית: לבצע Preemphasis למעטפת לפיה מבצעים את ההקצאה או לשנות באופן מלאכותי אחר את קביעת הקצאת הסיביות כדי לגרום לכך שרכיבי תדר גבוהים יקבלו יותר משקל בהקצאת הסיביות.

כיוון נוסף למחקר יכול להיות חפוש אוספי וקטורים - V שונים מאלה שהוצעו בעבודה זו שיובילו לאלגוריתמים פשוטים מחד ובעלי איכות גבוהה מאידך.

נספח א' - הוכחת טענה 4.1

הצגת הבעיה

נתונים: R מטריצה ממשית, סימטרית, P.D ממימד NxN.
Q מטריצה ממשית NxK (לא ריבועית) בעלת עמודות בלתי תלויות.

מהו Q כדי ש-

$$\Delta E = \text{trace}\,[(Q^TRQ)^{-1}\ (Q^TR^2Q)]$$

יהיה מקסימאלי?

פתרון

- נראה תחילה שעבור כל $\hat{Q}=QA$ כאשר A מטריצה לא סינגולרית נקבל אותו ΔE:

$$\text{trace}\,[(\hat{Q}^TR\hat{Q})^{-1}\ (\hat{Q}^TR^2\hat{Q})] = \tag{A.1}$$

$$= \text{trace}\,[(A^TQ^TRQA)^{-1}\ (A^TQ^TR^2QA)] =$$

$$= \text{trace}\,[A^{-1}(Q^TRQ)^{-1}(A^T)^{-1}\ A^T(Q^TR^2Q)A] =$$

$$= \text{trace}\,[A^{-1}(Q^TRQ)^{-1}\ (Q^TR^2Q)A] =$$

$$= \text{trace}\,[(Q^TRQ)^{-1}\ (Q^TR^2Q)] = \Delta E$$

כאן נעזרנו בעובדה שה-trace אינוריאנטי לפעולת דמיון.

- כעת נראה שניתן לעבור מן הבעיה המקורית לבעיה פשוטה יותר:

המטריצה Q^TRQ היא מטריצה ממשית סימטרית .P.D,
המטריצה Q^TR^2Q היא מטריצה ממשית סימטרית .P.D,
לכן, קיימת מטריצה B לא סינגולרית (לא בהכרח אורתוגנלית) [10, pp. 231]
כך שמתקיים:

(A.2) $$B^T(Q^TRQ)B = I$$

(A.3) $$B^T(Q^TR^2Q)B = D$$

כאשר D מטריצה אלכסונית ו-I היא מטריצת היחידה.

לפי (A.1) לא נקבל שינוי ב-ΔE אם נעבור ל-

(A.4) $$\hat{Q} = QB$$

נקבל:

(A.5) $$\begin{aligned}&(\hat{Q}^TR\hat{Q})^{-1}(\hat{Q}^TR^2\hat{Q}) =\\ &= (B^TQ^TRQB)^{-1}(B^TQ^TR^2QB) =\\ &= I^{-1}D = D\end{aligned}$$

כלומר, בעזרת (A.4) אנו עוברים לבעיה פשוטה יותר שבה עלינו לבצע trace על מטריצה אלכסונית:

(A.6) $$\begin{aligned}&\text{trace}\ ((Q^TRQ)^{-1}(Q^TR^2Q)) =\\ &= \text{trace}\ ((\hat{Q}^TR\hat{Q})^{-1}(\hat{Q}^TR^2\hat{Q})) =\\ &= \text{trace}\ (D)\end{aligned}$$

- הבעיה המתקבלת כעת היא:

מה יהיה $\hat{Q}$ כך שיתקיימים:

(A.7)

(i) $\hat{Q}^TR\hat{Q} = I$: מטריצת יחידה

(ii) $\hat{Q}^TR^2\hat{Q} = D$: מטריצה אלכסונית

(iii) $\text{trace}\ [(\hat{Q}^TR\hat{Q})^{-1}\ (\hat{Q}^TR^2\hat{Q})] \longrightarrow \max$

המטריצות R ו-R^2 הן ממשיות סימטריות ומתלכסנות ללכסון יוניטרי על-ידי הוקטורים העצמיים שלהן.

לכן, כדי לענות על אילוצים (i) ו-(ii) נצטרך לבחור את וקטורי $\hat{Q}$ להיות קומבינציות אורתוגונליות מנורמלות של וקטורים עצמיים.
נזכיר כי אם ל-R ערכים עצמיים שונים אזי הוקטורים העצמיים השייכים אליהם אורתוגונלים. אם ל-R ערך עצמי כלשהו מרובה אזי הוקטורים העצמיים השייכים אליו בלתי-תלויים וניתן להגיע מהם לבסיס אורתוגונלי, מוקטורים עצמיים אורתוגונליים ניתן לעבור לוקטורים עצמיים אורתונורמאליים [36].

נסמן ב-$\underline{v}_i$ וקטור עצמי של R ונניח שהוקטורים העצמיים מנורמלים כך שהם אורתונורמאליים.

הכוונה בקומבינציות אורתוגונליות מנורמלות היא שנבחר כל וקטור ב-$\hat{Q}$ להיות:

$$\hat{\underline{q}}_i = \frac{1}{\beta_i} \sum_{j=1}^{m_i} \alpha_{i,j} \; \underline{v}_{i,j} \qquad i=1....k \tag{A.8}$$

הסימון $\underline{v}_{i,j}$ פרושו הוקטור ה-j-י שמשתתף בקומבינציה ה-i-ית.

כאשר

$$\sum_{i=1}^{k} m_i = N \tag{A.9}$$

$$\underline{v}_{i,j}^T \underline{v}_{k,l} = 0 \quad , i \neq k, \; \forall_{j=1...m_i}, \; \forall_{l=1...m_k} \tag{A.10}$$

ו-

$$\hat{\underline{q}}_i^T \hat{\underline{q}}_i = 1 \tag{A.11}$$

במלים פשוטות, את N הוקטורים הנתונים לנו נחלק ל-k קבוצות, כל קבוצה תכיל וקטורים שונים (A.10), בכל קבוצה ניצור קומבינציה לינארית של הוקטורים וננרמל את הוקטור המתקבל (A.8) כך שיתקיים (A.11). כתוצאה מחלוקה זו כל ה-$\hat{\underline{q}}_i$ המתקבלים יהיו אורתונורמליים, ונקבל ליכסון של R ו-R^2 (ל-R ו-R^2 אותם וקטורים עצמיים.

נמצא כעת את הגודל המנרמל β_i ב-(A.8):

(A.12) $$R\hat{\underline{q}}_i = 1/\beta_i \sum_{j=1}^{m_i} \alpha_{i,j}\, R\underline{v}_{i,j}$$

$$= 1/\beta_i \sum_{j=1}^{m_i} \alpha_{i,j}\ \lambda_{i,j}\underline{v}_{i,j}$$

(A.13) $$\hat{\underline{q}}_i^T R\hat{\underline{q}}_i = \left(1/\beta_i \sum_{j=1}^{m_i} \alpha_{i,j}\, \underline{v}_{i,j}\right)^T 1/\beta_i \sum_{j=1}^{m_i} \alpha_{i,j}\, \lambda_{i,j}\, \underline{v}_{i,j}$$

$$= 1/\beta_i^2 \sum_{j=1}^{m_i} \alpha_{i,j}^2\ \lambda_{i,j} = 1$$

לכן:

(A.14) $$\beta_i = \left(\sum_{j=1}^{m_i} \alpha_{i,j}^2\ \lambda_{i,j}\right)^{1/2}$$

כלומר, עבור הקומבינציה ה-i-ית הוקטור $\hat{\underline{q}}_i$ שנרכיב יהיה:

(A.15) $$\hat{q}_i = \left(\sum_{j=1}^{m_i} \alpha_{i,j}^2\ \lambda_{i,j}\right)^{-1/2} \sum_{j=1}^{m_i} \alpha_{i,j}\underline{v}_{i,j}$$

כך ייתקיימו האילוצים (i) ו-(ii).

כעת עלינו לדאוג גם למילוי תנאי (iii):

$$\text{trace}\ (D) \longrightarrow \max$$

אברי האלכסון של D יהיו מהצורה:

$$d_{ii} = \hat{\underline{q}}_i^T R^2 \hat{\underline{q}}_i \tag{A.16}$$

אם λ_i ערך עצמי של R המתאים לוקטור עצמי $\underline{v}_i$ אזי λ^2_i ערך עצמי של R^2 המתאים לוקטור עצמי $\underline{v}_i$ נקבל לכן:

$$\hat{\underline{q}}_i^T R^2 \hat{\underline{q}}_i = \left(\sum_{j=1}^{m_i} \alpha^2_{i,j} \lambda_{i,j} \right)^{-1} \sum_{j=1}^{m_i} \alpha^2_{i,j} \lambda^2_{i,j} \tag{A.17}$$

מכיון ש-R חיובית מוגדרת, ערכיה עצמיים חיוביים ולכן:

$$\left(\sum_{j=1}^{m_i} \alpha^2_{i,j} \lambda_{i,j} \right)^{-1} \sum_{j=1}^{m_i} \alpha^2_{i,j} \lambda^2_{i,j} \le \lambda_i^{max} \tag{A.18}$$

כאשר λ_i^{max} הוא הערך העצמי הגדול ביותר בקבוצה ה-i-ית.
כדי לקיים את השיוויון ב-(A.18) נבחר את הקובינציה בכל קבוצה להיות מנורנת ולהיות בדיוק הוקטור העצמי השייך לערך העצמי הגדול ביותר באותה קבוצה:

$$\hat{\underline{q}}_i = (\lambda_i^{max})^{-1/2} \underline{v}_i^{max} \tag{A.19}$$

כאשר $\underline{v}_i^{max}$ הוא הו"ע המתאים ל-λ_i^{max}.
מכיון שעלינו לבצע מקסימיזציה על כל הקומבינציות האפשריות לחלוקת N הוקטורים ל-k קבוצות נקבל ש-trace(D) יקבל מקסימום כאשר $\hat{Q}$ יכיל את הוקטורים העצמיים המתאימים לערכים העצמיים הגדולים ביותר ב-R כשהם מנורמלים לפיהם.

<u>מסקנה:</u>

$\hat{Q}$ המבוקש יהיה מורכב מהוקטורים העצמיים של R השייכים לערכים העצמיים הגדולים ביותר ומנורמלים לפיהם.

$$(A.20) \qquad \underline{\hat{q}_i} = \lambda_i^{-1/2}\underline{v_i}$$

כאשר λ_i הם הערכים העצמיים הגדולים ו-$\underline{v}_i$ הם הוקטורים העצמיים השייכים אליהם.

מכיון שראינו כבר שלא נקבל שינוי ב-ΔE אם נעבור מ-Q אחד ל-$\tilde{Q}$ אחר על ידי:

$$\tilde{Q} = QA$$

כאשר A אינה סינגולרית, אזי נוכל לעבור מ-$\hat{Q}$ ל-$\tilde{Q}$ שיכיל פשוט את הוקטורים העצמיים המתאימים לערכים עצמיים גדולים של R. בצורה כזאת נשאר עם אותו ΔE אבל עם Q יותר פשוט.

נספח ב' - חישוב מקדמי LPC וקוונטיזציה

ברוב חלקי העבודה חישוב מקדמי ה-LPC נעשה בשיטת האוטוקורלציה [5].
הסיבוכיות הדרושה לשם כך היא:

N פעולות - כפל בחלון,
N(P+1) פעולות - חישוב מקדמי אוטוקורלציה. $r_0 \ldots r_p$
P^2 פעולות - פתרון בעזרת Levinson - Durbin.

בסה"כ $O(N(P+2)+P^2)$ פעולות למסגרת עדכון LPC.

המקדמים שנשדר לא יהיו מקדמי מסנן ה-LPC אלא מקדמי החזרה Reflection Coefficients (או PARCOR) שהם בעלי תכונות טובות לקידוד [7]:

- קל לשמור באמצעותם על יציבות מסנן הסינטזה ($k_i < 1$, $1 \le i \le p$ כאשר k_i הוא מקדם PARCOR מבטיח את יציבות המסנן).
- מקדמי ה-PARCOR הם בעלי סדר טבעי בינהם דבר שמאפשר ללמוד את הסטטיסטיקה של התפלגות ערכיהם ומכאן ניתן לקבל סכמאות קידוד יעילות יותר [7].

צורת הקידוד היא כדלקמן:

- השתמשנו בסימולציות במסנן LPC מסדר 10.
- k_1, k_2 שני מקדמי ההחזרה הראשונים מקודדים בעזרת Log Area Ratio:

$$LAR_i = \ln \frac{1 + k_i}{1 - k_i} \tag{B.1}$$

- $k_3 - k_{10}$ מקודדים בצורה ישירה.
- סכמת הקוונטיזציה היא לפי הסכמה של LPC-10 [8]:

1) הקצאת הסיביות נתונה בטבלה B.1:

טבלה B.1 - הקצאת סיביות

מקדם	מספר סיביות
k_1	5
k_2	5
k_3	5
k_4	5
k_5	4
k_6	4
k_7	4
k_8	4
k_9	3
k_{10}	2
סה"כ	41 סיביות

2) k_1, k_2 מובאים ל-6 bit וסימן ומקודדים לסימן וגודל בעזרת טבלה B.2:

טבלה B.2 - קידוד k_1, k_2

ערך אמיתי	ערך מקודד	ערך מפוענח
0 - 5	0	2
6 - 12	1	9
13 - 19	2	16
20 - 26	3	23
27 - 33	4	30
34 - 38	5	36
39 - 43	6	41
44 - 48	7	46
49 - 52	8	50
53 - 55	9	54
56 ,57	10	57
58 ,59	11	59
60	12	60
61	13	61
62	14	62
63	15	63

3) k_3-k_{10} מובאים ל-14 סיביות וסימן, מופחת מהם bias לפי טבלה B.3 והם עוברים Scaling ל-6 bit וסימן לפי טבלה B.4. לאחר מכן הפרמטרים עוברים Scaling נוסף למספר הביטים שהוקצה להם לפי טבלה B.1 . הפענוח נעשה על-ידי תהליך הפוך.

טבלה B.3 - bias

המקדם	bias
k_3	-1152
k_4	+2816
k_5	+1536
k_6	+3584
k_7	+1280
k_8	+2432
k_9	- 768
k_{10}	+1920

טבלה B.4 - Scaling

המקדם	Scaling
k_3	0.0056
k_4	0.0003
k_5	0.0068
k_6	0.0072
k_7	0.0074
k_8	0.0073
k_9	0.0084
k_{10}	0.0102

נספח ג' - חישובי סיבוכיות לאלגוריתמים האיטרטיביים

מימוש אלגוריתם I וסיבוכיות

נפרט את שלבי האלגוריתם השונים מבחינת צורת המימוש והסיבוכיות. המספר יתאים לשלב באלגוריתם.

1. חישוב h'_n ייעשה על-ידי העברת דגם יחידה דרך $\frac{1}{A(z/\gamma)}$.

לכך יידרשו:

- P כפלים למציאת מקדמים משוקללים (בהנחה שבזכרון נתונים ערכי: $(\gamma, \gamma^2, \ldots\ldots, \gamma^P)$
- LP פעולות Multiply & Add (MAD) לצורך ההעברה במסנן.

* בסה"כ נזדקק לשלב זה - (L+1)P פעולות.

2. המטריצה R מחושבת באופן רקורסיבי:

$$r_{ij} = \begin{cases} h'_{N-i-1}\, h'_o & \text{for} \quad j = N-1, \quad 0 \le i \le N-1 \\ r_{i+1,j+1} + h'_{N-i} h'_{N-j} & \text{for} \quad 0 \le j < N-1, \quad 0 \le i \le j \end{cases} \tag{C.1}$$

$$r_{ji} = r_{ij} \tag{C.2}$$

למעשה, מכיוון שהנחנו:

$$h'_n = 0 \qquad \forall \quad n \ge L \tag{C.3}$$

נקבל:

$$r_{ij} = r_{|i-j|} \quad \forall \quad i \le N-L \ \text{ or } \ j \le N-L \tag{C.4}$$

כש-$r_{|i-j|}$ הוא מקדם אוטוקורלציה.

לדוגמא, עבור $N = 10$, $L = 4$ נקבל:

$$(C.5)\qquad R = \left[\begin{array}{ccccccccccc}
r_0 & r_1 & r_2 & r_3 & & & & & & 0 & \\
r_1 & r_0 & r_1 & r_2 & r_3 & & & & & & \\
r_2 & r_1 & r_0 & r_1 & r_2 & \cdot & & & & & \\
r_3 & r_2 & r_1 & r_0 & r_1 & \cdot & \cdot & & & & \\
 & r_3 & r_2 & r_1 & r_0 & \cdot & \cdot & \cdot & & & \\
 & & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & & \\
 & & & \cdot & \cdot & \cdot & r_0 & r_1 & r_2 & r_3 & \\
 & & & & \cdot & \cdot & r_1 & r_0 & r_1 & r_2 & r_3 \\
 & & & & & \cdot & r_2 & r_1 & r_{77} & r_{78} & r_{79} \\
 & 0 & & & & & r_3 & r_2 & r_{87} & r_{88} & r_{89} \\
 & & & & & & & r_3 & r_{97} & r_{98} & r_{99}
\end{array}\right]$$

Covariance part

מספר החישובים הנדרש:

- L מכפלות לחישובים $r_{i,N-1}$ $\qquad N - L \le i \le N - 1$

- $\dfrac{L(L-1)}{2}$ MADS (Multiply and Add) לחישוב המקדמים הנוספים ב-Covariance part של המטריצה ומקדמי האוטוקורלציה r_k $\quad 0 \le k \le L - 1$,

בסה"כ נזדקק ל- $\dfrac{L(L+1)}{2}$ פעולות בשלב זה.

3. חישוב הוקטור $\underline{\Psi}$ $\quad (\Psi_i = 1/r_{ii},\ \ 0 \le i \le N-1)$ מתבצע לפני התחלת האיטרציות כדי לחסוך פעולות חילוק בזמן האיטרציות עצמן. ב-DSP קיימים כיום פעולת חילוק היא בדרך-כלל יקרה ולכן נשתדל לחסוך בפעולות אלה.

מספר פעולות דרוש:

עבור $i \le N-L$ חילוק יחיד $1/r_o$.

עבור $i > N-L$ L-1 חילוקים נוספים.

בסה"כ נזדקק ל-L פעולות חילוק.

4. חישוב $\underline{e}^o$ נעשה באופן הבא:

(i) חישוב ה-leftover - l_n מהמסגרת הקודמת על-ידי הזנת אפסים למסנן הסינטזה - NP MADS.

(ii) מציאת $s'_n = s_n - l_n$, N חיסורים.

(iii) העברת s'_n דרך $A(z)$ - NP MADS.

בסה"כ $N(2P+1)$ פעולות.

אנו משתמשים באות שארית שנמצא בתהליך המתואר ולא באות שארית רגיל (המתקבל על-ידי העברת s_n דרך $A(z)$) מכיוון שזהו אות השארית שיווצר במקלט.

5. חישוב $\underline{\alpha}$, $\hat{\underline{\alpha}}$ - 2LN MADS.

6. חישוב $\underline{\Delta E}$ - (k איטרציות) x (N כפלים).

7. השוואות - (k איטרציות) x (N השוואות).

8. חישוב אמפליטודה - (k איטרציות) x (כפל יחיד).

9. חישוב $\hat{\underline{S}}$ בעזרת מסנן הסינטזה - NP פעולות.

10. עדכון $\underline{\alpha}$, וחשוב $\hat{\underline{\alpha}}$:

(2L-1) כפלים (במקרה הגרוע) - $x_1 R\underline{v}_1$

(2L-1) סיכומים - $\underline{\alpha}-x_1 R\underline{V}_1$ x (k איטרציות)

(2L-1) כפלים - חשוב $\hat{\underline{\alpha}}$

בסך הכל נקבל למסגרת:

פעולות: $O(P(L+1) + \frac{L(L+1)}{2} + N(2P+1) + 2NL + k(2N+6L) + NP)$

+

פעולות חילוק: $O(L)$

הערות:

- חישוב הסיבוכיות נעשה ללא תוספת סיבוכיות חישוב מקדמי ה-LPC או ביצוע קוונטיזציה, בנספח ב' מובא הסבר על אופן חישוב המקדמים והקוונטיזציה.

- הכוונה בחישוב הסיבוכיות היא לתת סדר גודל לכמות החישובים הנדרשת ולא נלקחות בחשבון פעולות העברה בין המעבד והזכרון או פעולות דרושות לצורך חישובי אינדקסים.
במקרה שלפנינו יש לשים לב למבנה המורכב של המטריצה R והוקטור $\underline{\Psi}$, דבר שיסבך את מניפולצית האינדקסים.
ניתן לחסוך בסיבוכיות זו אם נחזיק את המטריצה R והוקטור $\underline{\Psi}$ באופן מלא.

מימוש אלגוריתם II וסיבוכיות

1. חישוב h'_n : $P(L+1)$ פעולות.

2. חישוב $\tilde{\underline{S}}$:

NP MADS לקבלת ה- leftover - $\underline{L}$.

N חיסורים לקבלת $\underline{S}' = \underline{S} - \underline{L}$

2NP MADS להעברת $\underline{S}'$ דרך - $\dfrac{A(z)}{A(z/\gamma)}$ לקבלת $\tilde{\underline{S}}$

סה"כ - $N(3P+1)$ פעולות.

3. חישוב $\underline{\alpha}$ - NL MADS, חישוב $\hat{\alpha}$ - N כפלים, סה"כ $N(L+1)$ פעולות

4. חישוב $\underline{r}$ ומתוכו $\underline{\Psi}$:

L + L MADS כתיבות ביניים (חישוב באופן רקוסיבי)

L פעולות חילוק

סה"כ 2L פעולות + L פעולות חילוק.

5. חישוב $\underline{\Delta E}$ - (N כפלים) x (k איטרציות).

6. מציאת מקסימום ב-$\underline{\Delta E}$ - (N השוואות) x (k איטרציות).

7. חישוב אמפליטודה - (כפל יחיד) x (k איטרציות).

8. חישוב $\hat{\underline{S}}$ בעזרת מסנך הסינטזה NP פעולות.

9. עדכון $\underline{\alpha}$:

$$\left.\begin{array}{ll} L^2 \text{ MADS} & \text{לחישוב } R\underline{v}_1 \\ (2L-1) & \text{כפלים ב-}x_1 \\ (2L-1) & \text{סיכומים} \\ (2L-1) & \text{כפלים לחישוב } \hat{\underline{\alpha}} \end{array}\right\} \quad x \quad (\text{k איטרציות})$$

האלגוריתם המתקבל זהה למעשה לאלגוריתם I אלא שכאן לא מחושבת המטריצה R אלא הוקטור $\underline{\alpha}$ (Crosscorrelation) מחושב ישירות. כתוצאה מכך נחסך מקום מכיון שאין צורך להחזיק את אברי המטריצה, אולם, מפסידים בסיבוכיות חישוב מכיון שצריך לחשב איברים מהמטריצה בשלב העדכון של $\underline{\alpha}$.

בסה"כ נקבל סיבוכיות חישוב:

פעולות למסגרת: $O\,(P(L+1) + N(4P+1) + N(L+1) + 2L + k(2N+6L+L^2))$

\+ פעולות חילוק למסגרת $O\,(L)$

<u>מימוש אלגוריתם III וסיבוכיות חישוב</u>

1. חישוב h'_n : - $P(L+1)$ פעולות.

2. חישוב R : - MADS $\frac{L(L+1)}{2}$

3. חישוב $1/r_o$: חילוק אחד.

4. חישוב $\underline{e}^o$:
- חישוב $\underline{L}$ (leftover) - NP MADS
- חישוב $\underline{S}' = \underline{S} - \underline{L}$ - N חיסורים
- חישוב $\underline{e}^o$: $\underline{S}'$ מועבר דרך $A(z)$ - NP MADS

בסה"כ $(2P+1)N$ פעולות.

5. חישוב $\underline{\alpha}$: $N(2L-1)$ MADS

6. מציאת מקסימום בערך מוחלט:

N פעולות למציאת ערך מוחלט.

N השוואות למציאת מקסימום.

בסה"כ (2N פעולות) x k איטרציות.

7. חישוב אמפליטודה: כפל יחיד.

8. חישוב $\hat{\underline{S}}$ בעזרת מסנן הסינטזה: NP פעולות.

9. עדכון $\underline{\alpha}$:
- L כפלים לחישוב $x_1\underline{R}v_1$ (סימטריה).
- 2L-1 חיסורים לעדכון $\underline{\alpha}$.

בסך הכל: (3L-1 פעולות) x k איטרציות.

בסך הכל נקבל למסגרת:

פעולות: $O\left(P(L+1) + \frac{L(L+1)}{2} + (2P+1)N + N(2L-1) + k(2N+3L) + NP\right)$

פעולת חילוק אחת.

ערות:

בשלב זה אנו מתייחסים בחישוב הסיבוכיות רק לאלגוריתם מציאת הערור, בגוף העבודה נתייחס גם לסיבוכיות הנוספת כתוצאה מחישוב מקדמי המסנן והקוונטיזציה.

יתכן שבמימושים מסוימים יהיה נוח להחזיק את וקטור מקדמי האוטוקורלציה כוקטור משוקף:

$$r_{L-1} \cdots r_2\ r_1\ r_0\ r_1\ r_2\ \cdots\ r_{L-2}\ r_{L-1}$$

ואז באמצעות הזזות אינדקסים קל לחשב את המכפלות בהן מעורבת המטריצה R.

בחישוב $R\underline{c} = \underline{\alpha}$ יש יתירות בחישובים.
יהיו כפלים שיתבצעו פעמיים בגלל מבנה ה-Toeplitz של R, ניצול יתירות זאת כנראה לא יקטין סיבוכיות מכיון שאז ייפגע המבנה של Multiply & Accumulate שניתן להשיג על-ידי חישוב איבר ב-$\underline{\alpha}$.

חישוב הסיבוכיות שניתן נכון עבור $L \le N/2$.
עבור $L > N/2$ תתקבל סיבוכיות שונה.

נספח ד' - חישובי סיבוכיות במערכת PTC

פעולות המתבצעות במשדר

1) חישוב Reflection Coefficints כל M דגמים: מקדמי האוטוקורלציה מחושבים על פני בלוק באורך m דגמים ($m \geq M$) הסיבוכיות הדרושה היא $O(m(P+2) + P^2)$ לחישוב סט מקדמים (נספח ב').

2) מציאת אות השארית e מתוך S - אות הדיבור: כאן היינו צריכים לנהוג למעשה כפי שנהגנו באלגוריתם האיטרטיבי ולחשב את e על-ידי הפחתת ה-leftover מ-S וסינון הפוך עם זכרון מסנן מאופס. אולם, כדי לחשב את ה-leftover היה עלינו לבצע שחזור.
באלגוריתם האיטרטיבי פעולת השחזור פשוטה ולכן בצענו אותה. אולם, באלגוריתם ה-PTC ביצוע שחזור גם במשדר מעמיד בספק את כדאיות האלגוריתם. לכן e הוא פשוט ה-residual לפי הגדרה, כלומר, מעבירים את S דרך A(z) ללא איפוס הזכרון. במקרה זה ההפסד באיכות איננו מורגש.
נזדקק איפוא ל-O(NP) פעולות בשלב זה.

3) מציאת המעטפת הספקטרלית: כפי שראינו בסעיף 6.3.2.1 שלב זה יעשה על-ידי חישוב $\log_2 \left|\frac{1}{A_{(z)}}\right|$ ב-J נקודות וביצוע אינטרפולציה לינארית. ב- $\left|\frac{1}{A(z)}\right|$ אנו מסמנים את הספקטרום של המסנן $\frac{1}{A(z)}$.

המימוש יהיה כדלקמן:

(i) עבור מ-k פרמטרים למקדמי מסנן a_i, $i=1...k$ לפי (2.4.2), $O(\frac{P(P-1)}{2})$ פעולות.

(ii) חשב את ה-DFT מאורך 2N של הסדרה a'_i ב-J נקודות המפוזרות באופן אחיד בתחום [0, N-1].

$$(D.1) \qquad a'_i = \begin{cases} a_i & i = 0 \ldots P \\ 0 & P < i \le 2N-1 \end{cases}$$

a_i הם מקדמי המסנן $A(z)$.

ה-DFT מחושב בצורה ישירה.

סיבוכיות: $O\,(2(P+1)J)$

(iii) מצא את ה-Magnitude ב-J הנקודות.

יש לבצע J פעמים פעולת $Re(A(j))^2 + Im(A(j))^2$

ב-J הנקודות בהם חושבה ההתמרה.

כאשר $A(j)$ הוא הדגם ה-j-י של התמרת $\underline{a}'$.

סיבוכיות - $O(2J)$ פעולות.

(iv) כדי להגיע ל- $\log \left|\frac{1}{A(z)}\right|$ ב-J הנקודות יש לחשב log על הגדלים

שנתקבלו בסעיף הקודם לחלק ב-2 ולהפוך סימן. כל הפעולות הנ"ל יכולות

להעשות ע"י - פניות לטבלה - $O(J)$ פעולות.

(v) אינטרפולציה: J-1 חיסורים לחישוב ההפרשים בין J הנקודות.

J-1 כפלים לחישוב צעד האינטפולציה בכל אינטרוול.

(בהנחה שהגודל $(J-1)/N$ חושב מראש).

N חיבורים לחישוב נקודות המעטפת בין הנקודות

המחושבות.

בסה"כ $O(2J-2+N)$

סך הכל השלבים לחישוב המעטפת : $O(2PJ + 7J + N + \frac{P(P-1)}{2})$

(4 הקצאת סיביות לפי המעטפת הספקטרלית מסעיף 3.

הקצאת הסיביות נעשית בעזרת התהליך שתואר בסעיף 6.3.2.1 כזכור, ההקצאה

נעשית לפי:

$$(D.2) \quad b(k) = \lfloor \log_2 |X(k)| - D \rfloor^* \qquad 0 \le k \le N-1$$

כאשר הוקטור $0 \le k \le N-1$ $\log_2 |X(k)|$ חושב בשלב הקודם.
D מחושב בשתי פאזות:

$$(D.3) \quad D_1 = 1/N \, [\sum_{k=0}^{N} \log_2 |X_{(k)}| - B]$$

כאשר -
$$B = \sum_{k=0}^{N-1} b(k) = const.$$

נזדקק ל-N פעולות חיבור ופעולת כפל בהנחה ש-1/N חושב מראש. בסה"כ O(N) פעולות לחישוב D_1.
כעת מגדירים קבוצה S^+ שמכילה את אינדקסי המקדמים עבורם צריך הקצאה אם משתמשים ב-D_1 בשערוך ל-D.

$$(D.4) \quad S^+ = \{ k \mid (\log_2|X(k)| - D_1) > 0 , \quad 0 \le k \le N-1 \}$$

מציאת הקבוצה הזאת כרוכה ב-N השוואות - O(N).
יהי N^+ מספר האיברים ב-S^+ אזי בפאזה השניה מחושב D_2:

$$(D.5) \quad D_2 = D_1 + 1/N^+ \, [\sum_{k \in S^+} \lfloor \log_2|X(k)| - D_1 \rfloor^* - B]$$

במקרה הגרוע נזדקק להקצאה לכל המקדמים $N^+ = N$ ונזדקק ל-
N פעולות חיסור.
N פניות לטבלה לחישוב $\lfloor \quad \rfloor^*$.
N סיכומים.
פעולת חילוק $(1/N^+)$.
בסה"כ - O(3N) פעולות + פעולת חילוק לפאזה השניה.

D_2 מהווה כעת שערוך ל-D, נחשב כעת את b(k) מתוך (6.3.11). נזדקק ל-N חיסורים ו-N פניות לטבלה לחישוב $\lfloor \cdot \rfloor^*$, סה"כ O(2N) פעולות. N חיבורים ידרשו לחישוב סה"כ הסיביות.

לבסוף נעבור פעם נוספת על הוקטור להוסיף סיביות חסרות או להוריד סיביות עודפות - $O(N)$ בסה"כ נקבל:

פאזה ראשונה - $O(N)$

פאזה שניה — $O(4N)$ + פעולת חילוק.

חישוב סופי - $\underline{O(4N)}$

סה"כ: $O(9N)$ + פעולת חילוק.

5) ביצוע DCT - שימוש בשיטה שהוצעה ב-[42] תדרוש $N/2\log_2 N+1$ כפלים ו- $3/2N\log_2 N - N + 1$ חיבורים.

נניח שאיננו מנצלים פעולות Multiply & Add במחזור מכונה אחד ונקבל:

$2N\log_2 N - N + 2$ פעולות.

כלומר, $O(2N\log_2 N - N)$ פעולות לביצוע DCT.

6) קוונטיזצית מקדמי ההתמרה:

עבור כל מקדם שמקבל הקצאה דרושה פעולת כפל כדי לנרמל אותו לפי $1/\hat{\sigma}$ כדי שנוכל להשתמש בטבלאות קוונטיזציה אחידות. במקרה הגרוע נזדקק איפוא ל-N כפלים (כאשר כל המקדמים מקבלים הקצאה, עבור קצבים גבוהים רוב המקדמים מקבלים הקצאה) ו-N פניות לטבלאות קוונטיזציה.

$\hat{\sigma}$ עצמו מתקבל מתוך חישוב מקדמי ה-LPC:

אלגוריתם Levinson - Dorbin [5] נותן גם את אנרגית אות השארית ומתוכו משוערך ווריאנס הדגם - $\hat{\sigma}^2$.

$\hat{\sigma}$ מקודד ל-6 סיביות בעזרת טבלה.

בסה"כ נזדקק ל-$O(2N)$ פעולות + פעולת חילוק.

R E F E R E N C E S

[1] J.L. Flanagan et. al., "Speech Coding", IEEE Trans. on Commun, vol COM-27, No. 4, pp. 710-737, April 1979.

[2] N.S. Jayant and P. Noll, "Digital Coding of Wave Forms", Prentice - Hall, Signal Processing Series, 1984.

[3] B.S. Atal and S.L. Hanauer, "Speech analysis and synthesis by linea prediction of the speech wave", J. Acoust. Soc. Amer., vol. 50 no. 2, pp. 637-655, 1971.

[4] J. Makhoul, "Linear Prediction: A Tutorial Review", Proc. of the IEEE vol. 63, no. 4, pp. 561-580, April 1975.

[5] J.D. Markel and A.H. Gray Jr., "Linear Prediction of Speech", Springe - Verlag, Berlin - New York, 1976.

[6] J.D. Markel and A.H. Gray Jr., "A Linear Prediction Vocoder Simulatio Based upon the Autocorrelation Method", IEEE Trans. Acoust., Speech Signal Processing, vol. ASSP-22, pp. 124-134, 1974.

[7] A.H. Gray Jr. and J.D. Markel, "Quantization and Bit Allocation i Speech Processing", IEEE Trans. Acoust., Speech, Signal Processing vol. ASSP-24, pp. 459-473, Dec. 1976.

[8] T.E. Tremain, "The Government Standard Linear Predictive Codin Algorithm: LPC-10", Speech Technology, pp. 40-49, April 1982.

[9] D.Y. Wong, "On Understanding the quality problems of LPC Speech" Proc. IEEE Int. Conf. ASSP 1980, pp. 725-728.

[10] W. Ledermann and S. Vajda, "Handbook of Applicable Mathematics, Volum I: Algebra", John Wiley & Sons.

[11] B.S. Atal "Predictive Coding of Speech at Low Bit Rates", IEEE Trans Commun., vol. COM-30, pp. 600-614 April 1982.

[12] B.S. Atal and M.R. Schroeder, "Predictive Coding of Speech Signals an Subjective Error Criteria", IEEE Trans. on Acoust., Speech, Signa Processing, Vol. ASSP-27, No. 3, June 1979.

[13] B.S. Atal and J.R. Remde, "A new Model of LPC Excitation for Producin Natural - Sounding Speech at Low Bit Rates", Proc. Int. Conf. on ASSP Paris, France, 1982, pp. 614-617.

[14] S. Singhal and B.S. Atal, "Optimizing LPC Filter Parameters for Multi Pulse Excitation", Proc. Int. Conf. on ASSP, Boston, U.S.A., 1983, pp 781-784.

[15] S. Singhal and B.S. Atal, "Improving Performance of Multi-Pulse LP Coders at Low Bit Rates", Proc. Int. Conf. on ASSP, 1984, pp 1.3.1-1.3.4 .

[16] A. Parker, S.T. Alexander, H.J. Trussell, "Low Bit Rate Speech Enhancement Using a New Method of Multiple Impulse Excitation", Proc Int. Conf. on ASSP, 1984, pp. 1.5.1-1.5.4 .

[17] S.T. Alexander, "A Simple Noniterative Speech Excitation Algorith Using the LPC Residual", IEEE Trans. on Acoust., Speech, Signal Processing., vol. ASSP-33, No. 2, April 1985.

[18] M. Berouti et. al. "Efficient Computation and Encoding of the Multipulse Excitation for LPC", Proc. Int. Conf. ASSP, 1984, pp 10.1.1-10.1.4 .

[19] P. Kroon and E.F. Deprettere, "Experimental Evaluation of Differen Approaches to the Multi-Pulse Coder", Proc. Int. Conf. on ASSP, 1984 pp. 10.4.1-10.4.4 .

[20] K. Ozawa, S. Ono and T. Araseki, "A Study on Pulse Search Algorithms for Multi-Pulse Excited Speech Coder Realization", IEEE Journal on Selected Areas in Commun., vol. SAC-4, No. 1, Jan. 1986, pp. 134-141.

[21] A. Dembo and D. Malah, "A New Approach to Multi-Puse LPC Coder Design", Proc. Int. Conf. on ASSP, 1985, pp. 949-952.

[22] J.P. Lefevre and O. Passien, "Efficient Algorithms for Obtaining Multipulse Excitation for LPC coders", Proc. Int. Conf. on ASSP, 1985, pp. 957-960.

[23] E.F. Deprettere and P. Kroon, "Regular Excitation Reduction for Effective and Efficient LP-Coding of Speech", Proc. Int. Conf. on ASSP, 1985, pp. 965-968.

[24] Z.A. Putnins, "A Multi-Pulse LPC Synthesizer for Telecommunications Use", Proc. Int. Conf. on ASSP, 1985, pp. 981-991.

[25] R. Sharma, "Architecture Design of a High - Quality Speech Synthesizer Based on the Multi-Pulse LPC Technique", IEEE Journal on Selected Areas in Commun., vol. SAC-3, No. 2, March 1985, pp. 377-383.

[26] Y. Wake et. al., "A Multi-Pulse LPC Speech Coder Using Digital Signal Processors", Proc. Int. Conf. on ASSP, 1985, pp. 1429-1432.

[27] M.R. Schroeder and B.S. Atal, "Code - Excited Linear Prediction (CELP): High-Quality Speech At Very Low Bit Rates", IEEE Proc. Int. Conf. on ASSP, 1985,pp.937-940.

[28] C.K. Un and D.T. Magill, "The Residual Excited Linear Predictive Vocoder with Transmission Rate Below 9.6 K bits/s", IEEE Trans. Commun., Vol. COM-23, pp. 1466-1474,Dec. 1975.

[29] L.L. Burge and R. Yarlagadda, "An efficient Coding of the Prediction Residual", Proc. Int. Conf. on ASSP, 1979, pp. 538-541.

[30] I.M. Trancoso and B.S. Atul, "Efficient procedurs for finding optimum innovation in stochastic coders", IEEE proc. Int. Conf. on ASSP, 1986.

[31] J. Makhoul and M. Berouti, "High Frequency Regeneration in Speech Coding Systems", IEEE Proc. Int. Conf. on ASSP, 1979, pp. 428-431.

[32] V.R. Viswanathan, A.L. Higgins and W.H. Russell, "Design of a Robust Baseband LPC Coder for Speech Transmission over 9.6 K bit/s Noisy Channels", IEEE Trans. on Commun., vol. COM-30, NO. 4, April 1982.

[33] R.C. Rose and T.P. Barnwell III, "The Self Excited Vocoder - An Alternate Approach to Toll Quality At 4800 bps", IEEE Proc. Int. Conf. on ASSP, 1986, pp. 453-456.

[34] P.M. Narendra and K. Fukunaya, "A Branch and Bound Algorithm for Feature Subset Selection", IEEE Trans. on Computers, vol. C-26, No. 9, September 1977, pp. 917-922.

[35] A. Dembo, Private Communication.

[36] P. Lancaster, "Theory of Matrices", Academic Press, 1969.

[37] N. Ahmed, T. Naturajan and K.R. Rao, "Discrete Cosine Transform", IEEE Trans. on Computers, January 1974, pp. 90-93.

[38] H.J. Narasimha and A.M. Peterson, "On the Computation of the Discrete Cosine Transform", IEEE Trans. on Communications, vol. COM-26, No. 6, June 1978, pp. 934-936.

[39] W.H. Chen, H. Smith and S.C. Fralick, "A Fast Computational Algorithm for the Discrete Cosine Transform", IEEE Trans. on Communications, Vol. COM-25, No. 9, September 1977, pp. 1004-1009.

[40] Z. Wang, "Fast Algorithms for the Discrete W Transform and for the Discrete Fourier Transform", IEEE Trans. on ASSP Vol. ASSP-32, No. 4, August 1984, pp. 803-816.

[41] N. Suehiro and M. Hatori, "Fast Algorithms for the DFT and Other Sinusoidual Transforms", IEEE Trans. on ASSP, vol. ASSP-34, No. 3, June 1986, pp. 642-644.

[42] Z. Wang, "On Computing the Discrete Fourier and Cosine Transforms", IEEE Trans. on ASSP, Vol. ASSP-33, No. 4, October 1985, pp. 1341-1344.

[43] "TMS32020 User's Guide", Texas Instruments.

[44] J. Max, "Quantizing for Minimum Distortion", IRE Tran. on Information Theory, Vol. IT-6, pp. 7-12, Mar. 1960.

[45] R. Zelinski and P. Noll, "Adaptive Transform Coding of Speech Signals", IEEE Trans. on ASSP, Vol. ASSP-25, No. 4, Aug. 1977, pp. 299-309.

[46] R. Zelinski and P. Noll, "Approaches to Adaptive Transform Speech Coding at Low Bit Rates", IEEE Trans. on ASSP, Vol. ASSP-27, No. 1, Feb. 1979, pp. 89-95.

[47] J.M. Triholet and R.E. Crochiere, "Frequency Domain Coding of Speech", IEEE Trans. on ASSP, Vol. ASSP-27, No. 5, October 1979, pp. 512-530.

[48] R.V. Cox and R.E. Crochiere, "Real-Time Simulation of Adaptive Transform Coding", IEEE Trans. on ASSP, Vol. ASSP-29, No. 2, April 1981, pp. 147-154.

[49] F.J. Macwilliams, N.J.A. Slone, "The Theory of Error-Correcting Codes", North-Holland Mathematical Library, 1981.

[50] N.J.A. Slone and D.S. Whitehead, "A New Family of Single - Error Correcting Codes", IEEE Trans. on Info. Theory Vol. IT-16, 1970, pp 717-719.

[51] D. Malah, Private Communication.

[52] M. Hamidi and J. Pearl, "Comparison of the Cosine and Fourie Transforms of Markov-1 Signals", IEEE Trans. on Acoustics, Speech an Signal Processing Oct. 1976, pp. 428-429.

[53] Y. Yemini and J. Pearl, "Asymptotic Properties of Discrete Unitar Transforms", IEEE Trans. on PAMI, vol. PAMI-1, No. 4, October 1979.

[54] P.E. Papamichalis, "Markov - Huffman Coding of LPC Parameters", IEE Trans. on ASSP, vol. ASSP-33, No. 2, April 1985.

[55] P.A. Devijver and J. Kittler, "Pattern Recognition - A Statistica Approach", Prentice-Hall International, 1982.

SPEECH CODING BY VECTOR EXPANSION OF THE LPC RESIDUAL SIGNAL

RESEARCH THESIS

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

IN

ELECTRICAL ENGINEERING

BY

EREZ OFER

SUBMITTED TO THE SENATE OF THE TECHNION - ISRAEL INSTITUTE OF TECHNOLOGY

TISHRAY 5747 **HAIFA** **SEPTEMBER 1986**

This research was carried out in the Signal Processing Laboratory of the Faculty of Eelectrical Engineering under the supervision of Prof. David Malah.

I wish to express my sincere gratitude to Prof. David Malah for his dedicated supervision and patience throughout this work.

I would like to thank TADIRAN - Communication division which supported this research.

I am also grateful to Dr. Amir Dembo for many fruitful suggestions and useful discussions.
To Mrs. Ziva Avni for the assistance with the computer work.
To Mr. Yoram Or-Chen for his assistance.
To all the Signal Processing Laboratory staff and my friends who helped me complete this work.

Contents

ABSTRACT

In this work we propose a new method for coding the residual signal in residual coders : The residual signal is represented by a linear combination of a small number of vectors taken from a given set. The vector set is known both at the transmitter and at the receiver . The parameters transmitted are the indices of the chosen vectors from the set and the coefficients needed to linearly recombine the vectors.

Existing digital speech coding techniques require binary data rates varying from 64 Kbps down to about 2.4 Kbps. At the upper end of the scale stand waveform coders which provide high speech quality and are robust to acoustic and transmission noise. The performance of the waveform coders typically falls rapidly when the data rate is reduced below 16 Kbps . At the lower end of the data rate scale stand LPC pitch-excited vocoders which are based on a model of the human speech production system . LPC vocoders have a synthetic quality and do not perform well under poor acoustic conditions or when a number of speakers are speaking together. Increasing the data rate of a LPC vocoder above the typical rate of 2.4 Kbps does not significantly improve the quality or acoustic robustness of the coder .

It is evident therefore that it is difficult to achieve good performance in the data rate range of 4.8 - 16 Kbps by using either waveform coders or LPC vocoders. Various techniques were attempted for speech coding in this range of data rates . A class of coders designed to operate in these rates are residual coders . These coders have in common the feature of coding both the LPC residual and the LPC filter coefficients. At the top, both in quality and bit rate , are the full-band coders in which each residual sample is quantized individually, The APC (Adaptive Predictive Coding) system is an example to such a coder. However, in order to achieve data rates of 9.6 Kbps and below, the residual coders can not afford quantizing each residual sample individually. One approach is Multi-pulse LPC in which the residual is represented by only few significant samples , which are found by an iterative procedure in a closed loop. CELP (Code Excited Linear Prediction) coders use blocks of white Gaussian noise to represent the residual signal.

With the new method proposed in this work the residual signal is coded as a linear combination of a small number of vectors. This method is a generalization of the above mentioned schemes - the Multipulse LPC with its different variations and CELP.

With this coding scheme two problems are encountered : Firstly, given a vector set how do we find the optimal linear combination for representing the residual signal ? Secondly , What is the optimal vector set ?

Solving the first problem is a difficult task because even with a small set of vectors , the number of combinations is large. Hence, suboptimal solutions are examined. One of these solutions, is a simple iterative one which leads to the conventional Multipulse LPC algorithm. The solution of the second problem leads to a non-iterative solution operating in the frequency domain , Denoted as PTC (Predictive Transform Coding) . The PTC system is of lower complexity then iterative coders used in the data rate range 9.6 - 16 Kbps and the speech quality achieved at 16 Kbps is even better . At 9.6 Kbps the quality is good but a 'low passing' effect is noticed .

In this work we present a formal mathematical description of the new vector expansion coding scheme . An analytical solution is given to the problem of finding the optimal vector combination , a suboptimal iterative solution is also presented, and the proposed PTC scheme is derived. Objective and Subjective performance tests are presented for the PTC coder as compared with the conventional iterative (Multipulse LPC) coder - in terms of speech quality , noise immunity , tandeming and Voice Band Data (VBD) signal transmission .