



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

©

All rights reserved

*This work, in whole or in part, may not be copied (in any media), printed, translated, stored in a retrieval system, transmitted via the internet or other electronic means, except for "fair use" of brief quotations for academic instruction, criticism, or research purposes only.
Commercial use of this material is completely prohibited.*

©

כל הזכויות שמורות

אין להעתיק (במדיה כלשהי), להדפיס, לתרגם, לאחסן במאגר מידע, להפיץ באינטרנט, חיבור זה או כל חלק ממנו, למעט "שימוש הוגן" בקטעים קצרים מן החיבור למטרות לימוד, הוראה, ביקורת או מחקר. שימוש מסחרי בחומר הכלול בחיבור זה אסור בהחלט.

קידוד דיבור רועש באמצעות חיזוי לינארי
ושארית קפסטרלית.

חבור על מחקר

לשם מילוי חלקי של הדרישות לקבלת התואר

מגיסטר למדעים

בהנדסת חשמל

2033066



000001012934

5/2/88

מאת

קופולוביץ צבי

הוגש לסנט הטכניון - מכון טכנולוגי לישראל

תשרי - תשמ"ה ח'פה ספטמבר 1984

טכניון
פקולטה להנדסת חשמל
פקולטה למדעי מחשב
ספרייה מס' 493

המחקר נעשה בהנחיית פרופסור דוד מלאך
במעבדה לעבודות אותיות כפקולטה להנדסת חשמל

תודתי נתונה לפרופסור דוד מלאך על הנחיתו המועילה
סבלנותו ואורך הרוח שגילה כלפי ככל שלבי המחקר.

כמו כן מודה אני למהנדס המעבדה מר יורם אור חן על
עזרתו הרבה ויחסו החברי. ולגברת ציפי פורטנוי
על סיועה בשעות הקשות בעבודתי במחשב.

תוכן הענינים

עמוד

1 תקציר	
3 רשימת סמלים וקיצורים	
7 מבוא	פרק 1 :
11 חיצוי לינארי של דבור (LPC)	פרק 2 :
11 מבוא	2.1
13 מודל מערכת יצור הדבור	2.2
15 נתוח האלגוריתם וחשוב הפרמטרים	2.3
15 שערך מקדמי החזוי	2.3.1
24 בעית v/uv וחשוב תדר ה- Pitch	2.3.2
31 אות השארית בחיצוי לינארי	2.4
32 חסרונות שיטת החיצוי הלינארי בקודד דבור	2.5
33 חיצוי לינארי של דבור ברעש לבן	2.6
34 LPC עם קודד השארית הקפסטרלית	פרק 3 :
34 מבוא	3.1
36 הקפסטרום	3.2
36 הקפסטרום מהו ?	3.2.1
36 תכונות הקפסטרום וקפסטרום הקומפלקסי	3.2.2
39 הצגת מודל האנליזה	3.3
39 עבוד מקדים והגדרת קטע האנליזה (Windowing)	3.3.1
40 מציאת פרמטרי ה- LPC וחשוב השארית הקפסטרלית	3.3.2
42 חשוב הממוצע של השארית הקפסטרלית	3.3.3
43 חשוב האפסים	3.3.4
44 חשוב השארית הקפסטרלית הנותרת	3.3.5
45 החלטת v/uv וחשוב תדר ה- Pitch	3.3.6
45 המודל השלם	3.3.7
48 מערכת הסינטזה	3.4

תוכן העניינים (המשך)

עמוד

50 מדד איכות E_c	3.5
52 סימולציה של מודל קדוד השארית	3.6
52 תוצאות הסימולציה	3.6.1
56 מסקנות	3.6.2
58 פרק 4 : חיזוי לינארי של אותות דבור הטבולים ברעש לבן	
58 מכוא	4.1
59 אות הדבור הרועש	4.2
59 הצגת מודל האות הרועש	4.2.1
61 השפעתו של הרעש על מודל ה-LPC	4.2.2
62 שיטות להקטנת השפעתו של הרעש על ביצועי מקדד LPC	4.3
62 הגדלת סדר המודל	4.3.1
63 הפחתה ספקטרלית (Spectral Subtraction)	4.3.2
65 הפחתת אוטוקורלציה	4.3.3
66 משוואות Yule Walker מסדר גבוה	4.3.4
67 שיערוך קטבים ואפסים	4.3.5
68 סכום	4.4
 פרק 5 : מקדד דבור עם הפרדה בין קטבים ואפסים במישור נגזרת	
69 הפאזה (DPS)	
69 מכוא	5.1
70 הפרדה בין קטבים ואפסים במשור DPS	5.2
70 הצגת הבעיה	5.2.1
70 הצגת ה-DPS ותכונותיו	5.2.2
72 הקשר בין DPS לבין הקפסטרום	5.2.3
74 אלגוריתם ההפרדה בין קטבים ואפסים	5.2.4
76 מקדד "YANG" של אות דבור נקי	5.3
76 הצגת מערכת האנליזה ("YANG")	5.3.1
77 מערכת הסינטזה ("YANG")	5.3.2
78 תוצאות הסימולציה	5.3.3

תוכן הענינים (המשך)

עמוד

83 מסקנות	5.3.4
84 מקדד "YANG" עבור אות דבור רועש	5.4
87 סכום	5.5
פרק 6 : מקדד דבור רועש עם הפחתה ספקטרלית וקדוד השארית		
88 הקפסטרלית	
88 מבוא	6.1
89 הצגת מערכת האנליזה	6.2
89 המבנה הכללי של המערכת	6.2.1
90 פרוט של מערכת האנליזה	6.2.2
93 בעיות אי היציבות של מסנן LPC	6.3
96 מקדד LPC ישיר על הספקטרום המנוקה	6.4
98 מבנה השארית הקפסטרלית	6.5
100 בחירת הסף באלגוריתם ההפחתה הספקטרלית	6.6
100 סף הנגזר מהעוטפת הספקטרלית של הקטע הקודם	6.6.1
103 סף המתחס לממוצע הספקטרלי בכל קטע	6.6.2
107 תוצאות הסימולציות	6.7
108 סכום ומסקנות	פרק 7 :
נספח א : חשוב נגזרת ספקטרום הפאזה (DPS) עבור מסנן מסדר		
110 ראשון ומסנן מסדר שני	
110 מבוא	A.1 :
110 DPS עבור מסנן מסדר ראשון	A.2 :
111 DPS עבור מסנן מסדר שני	A.3 :
112 רשימת מקורות	

תקציר.

קדוד באמצעות חיזוי לינארי (LPC) היא שיטה יעילה ומקובלת לקידוד דיבור בקצב סיביות נמוך (למשל 2400 סיביות לשניה). שיטה זו מתבססת על המודל הקלסי ליצירת הדיבור, הונה אומר: אות עירור המעורר מסנן לינארי המשתנה באופן איטי בזמן. אות העירור נבחר משתי אפשרויות: סדרה מחזורית של פולסים בעלת מחזור המשתנה בהתאם לשינוי המחזור היסודי של אות הדיבור או רעש לבן. מניחים כי המסנן הנו בעל קטבים בלבד. הנחה זו נמצאה קרובה למציאות ומתאימה במיוחד לשימוש באלגוריתם חיזוי לינארי.

הדיבור הנוצר מסינטזת LPC מאכד מטיבעיותו ונשמע מלאכותי. בכדי להחזיר לדיבור את טיבעיותו משתמשים באות השארית (אות העירור המבטיח אות משוחזר במדויק אך נדרשות סיביות רבות לקדודו), במקום או כתוספת לאות העירור הקלסי. שיטה אחת לשימוש יעיל באות השארית היא קידוד העוטפת הספקטרלית שלו במישור הקפסטרום ושימוש במקדמי קפסטרום אלו לעיצוב ספקטרום אות העירור לפני היכנסו למסנן הסינטזה.

כאשר מופעל אלגוריתם ה-LPC על אות דיבור הטכול ברעש לבן ניכרת ירידה דרסטית באיכות הדיבור המשוחזר. המקור לירידה באיכות הדיבור היא העובדה שהמודל המיצג את אות הדיבור הרועש שונה מזה המתאים לאות דבור נקי ואינו מתאים במכנהו לאלגוריתם ה-LPC. על מנת לאפשר שימוש במקדד LPC על אות רועש יש להקטין את השפעת הרעש.

בעבודה זו מוצעות ונבדקות שתי שיטות למימוש מקדד חיזוי לינארי הפועלות על אות דיבור רועש.

השיטה הראשונה מתבססת על כך שמודל אות הדיבור הרועש מתאים למודל קטבים ואפסים, כאשר הקטבים הם תרומתו של אות הדיבור וזהים לקטבים של הדיבור הנקי, בעוד שהאפסים הם תרומתו של הרעש. בשיטה זו מנוצל אלגוריתם ידוע להפרדה בין קטבים ואפסים של האות במישור נגזרת ספקטרום הפאה ושימוש בקטבים למימוש מסנן הסינטזה. בממוש מערכת זו הסתבר שאין אלגוריתם זה משערך בצורה נכונה את העוטפת הספקטרלית של הדיבור ובמיוחד גורם לעיוות בפורמנטים.

השיטה השניה מבוססת על מקדד החיזוי הלינארי עם קידוד השארית הקפסטרלית, כאשר מקדים אותו אלגוריתם ההפחתה הספקטרלית, שהיא שיטה מקובלת לניקוי דיבור רועש. במימוש מערכת זו נתגלו שתי בעיות עיקריות. הראשונה היא בעיית יציבותו של מסנן הסינטזה המתקבל, במסגרת עבודה זו מוצג פתרון אפשרי לבעיה. השניה חוסר התאמתו של אלגוריתם ההפחתה הספקטרלית למערכת המוצעת, עובדה שגררה מודיפיקציה של אלגוריתם זה.

לבחינת ביצועי המערכות שהוצגו לעיל נבחר מדד כמותי לאי ההתאמה הספקטרלית בין אות הדיבור הנקי לזה המשוחזר. מדד זה נמצא יעיל בייצוג איכותו של המקדד כאשר זה פועל על אות דיבור נקי, אולם נכשל כאשר דובר בקידוד דיבור רועש. לאור עובדה זו ביצעו המקדדים על דיבור רועש נבחנו בכחינות שמיעה סוביקטיביות. כאשר נבחנו המקדדים שהוצגו לעיל בכדיקות הנ"ל מסתבר שהמערכת הראשונה מיצרת דיבור מעוות מאוד (למרות שבהתכוננות במדד הכמותי ניתן היה להניח שמקדד זה יעיל). איכות הדיבור המשוחזר על ידי המערכת השניה נמצאה סבירה.

רשימת סמלים.

- $A(z)$ - פונקציית התמסורת של המסנן ההפוך.
- $A(w)$ - תגובת החדר של המסנן ההפוך.
- a_k - מקדם של המסנן ההפוך.
- $B(z)$ - פונקציית התמסורת של מסנן האפסים במודל קטבים ואפסים.
- $B(w)$ - תגובת החדר של מסנן האפסים במודל אפסים וקטבים.
- b_k - מקדמי המסנן $B(z)$.
- $c(n)$ - הקפסטרום הממשי של אות הדיבור.
- $c_a(n)$ - הקפסטרום הממשי המיצג את המסנן ההפוך.
- $c_r(n)$ - אות השארית הקפסטרלית.
- $\langle c_r(n) \rangle$ - הממוצע של השארית הקפסטרלית, בזמן.
- $\tilde{c}_r(n)$ - השארית הקפסטרלית חסרת ההטיה.
- $\hat{c}_r(n)$ - השארית הקפסטרלית הנותרת במודל PCR.
- $c_t(n)$ - השארית הקפסטרלית המקודדת.
- $c_e(n)$ - אות השגיאה הקפסטרלית לאחר קדור חלק מהשארית הקפסטרלית.
- E - אנרגיית אות השגיאה על פני קטע האנליזה.
- E_p - אנרגיית השגיאה הנותרת ב-LPC עבור מודל מסדר p .
- $E(z)$ - התמרת z של אות השארית.
- E_c - מדד אי ההתאמה הספקטרלית, ב-dB.
- $F\{ \}$ - התמרת פוריה.
- $F^{-1}\{ \}$ - התמרת פוריה ההפוכה.
- F_c - תדר ההחלפה במודל ערוור מסוג Mixed source.
- F_p - התדר היסודי של הדבור.
- G - הגבר מסנן הסינטזה.
- $H(z)$ - פונקציית התמסורת של מסנן הסינטזה.
- $H(w)$ - תגובת החדר של מסנן הסינטזה.
- $h(n)$ - התגובה להלם יחידה של מסנן הסינטזה.
- $\hat{h}(n)$ - הקפסטרום הקומפלקסי של אות המקור.
- $\hat{h}^+(n)$ - הקפסטרום הקומפלקסי המיצג קטבים.
- $\hat{h}^-(n)$ - הקפסטרום הקומפלקסי המיצג אפסים.
- k_i - מקדם ההחזרה החלקית.
- K_0 - הסף באלגוריתם ההפחתה הספקטרלית.

רשימת סמלים - (המשך).

- $K_m(w)$ - סף אדפטיבי באלגוריתם ההפחתה הספקטרלית, משתנה בתדר.
- K_m - סף אדפטיבי באלגוריתם ההפחתה הספקטרלית, קבוע בתדר.
- N - אורך חלון האנליזה.
- N_0 - ספקטרום הספק של רעש לבן.
- $N(w)$ - ספקטרום הספק הרעש.
- $\hat{N}(w)$ - משעריך של ספקטרום הספק הרעש.
- p - סדר מודל ה-LPC.
- $P_x(z)$ - התמרת z של ספקטרום הספק האות הרועש.
- $P_s(z)$ - התמרת z של ספקטרום הספק האות הנקי.
- $P_x(w)$ - ספקטרום הספק האות הרועש.
- $P_s(w)$ - ספקטרום הספק האות הנקי.
- $\hat{P}_s(w)$ - משעריך של ספקטרום הספק האות הנקי.
- $P(\alpha > p)$ - פונקציית ההסתברות של p לקבלת מסנן יציב.
- $R(\)$ - מקדם אוטוקורלציה.
- \underline{R} - וקטור של מקדמי אוטוקורלציה.
- $\underline{\underline{R}}$ - מטריצת האוטוקורלציה.
- $R_s(n)$ - אוטוקורלציה של האות הנקי.
- $R_x(n)$ - אוטוקורלציה של האות הרועש.
- $R_n(n)$ - אוטוקורלציה של הרעש.
- $\hat{R}_s(n)$ - משעריך אוטוקורלציה של האות הנקי.
- $\hat{R}_n(n)$ - משעריך אוטוקורלציה של הרעש.
- $s(n)$ - אות הדבור הנקי.
- $\hat{s}(n)$ - החזאי הלינארי של אות הדבור הנקי.
- $S(z)$ - התמרת z של אות הדבור הנקי.
- $S(w)$ - ספקטרום של הדבור הנקי.
- $\hat{S}(w)$ - משעריך ספקטרום של הדבור הנקי.
- $u(n)$ - אות הערור.
- $U(z)$ - התמרת z של אות הערור.
- $U(w)$ - ספקטרום של אות הערור.
- w - תדר זוויתי ברדיאנים.

רשימת סמלים - (המשך).

- $x(n)$ - אות הדבור הרועש.
 α - תדר חצי הספק במסנן מסדר שני.
 β - תדר מרכזי של מסנן מסדר שני.
 δ - תדר חצי הספק של מסנן מסדר ראשון.
 $\theta_x(w)$ - הפאזה של האות הרועש.
 $\theta'(w)$ - ספקטרום הפאזה.
 $\theta''(w)$ - נגזרת ספקטרום הפאזה.
 $[\theta'(w)]^+$ - נגזרת ספקטרום הפאזה המיצגת אפסים.
 $[\theta'(w)]^-$ - נגזרת ספקטרום הפאזה המיצגת אפסים.
 ϵ_c^2 - מדד אי התאמה ספקטרלית.

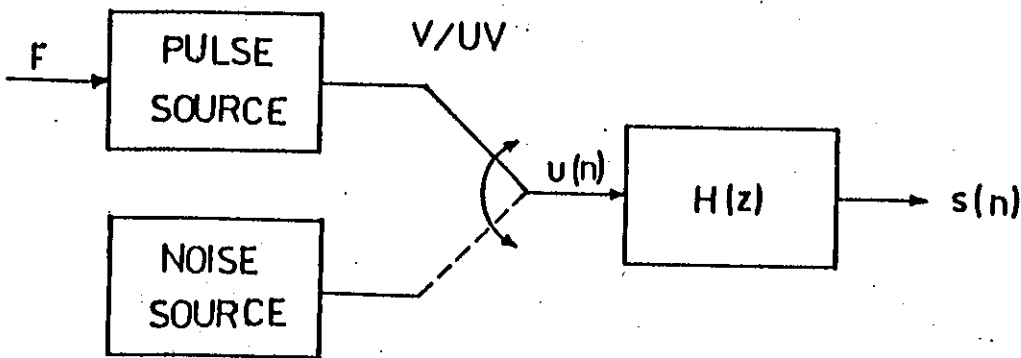
רשימת קיצורים.

- .Derivative of Phase Spectrum - DPS
- .לוגריתם על בסיס טבעי. - Log
- .Linear Predictive Coding - LPC

- .מקדד חיצוני לינארי עם קדוד שארית קפסטרלית. - PCR
- .מקדד חיצוני לינארי עם קדוד אפסים ושארית קפסטרלית. - PZCR
- .Simplified Inverse Filter Tracking - SIFT
- .Spectral Subtraction - SPS
- .Voice/UnVoice - V/UV
- .Yanganarayana - אלגוריתם ההפרדה בין קטבים ואפסים לפי "YANG"
- .Yule Walker - Y.W.

פרק ראשון: מכוא.

חיזוי לינארי (LPC) היא שיטה מקובלת ויעילה לקידוד דיבור בקצב סיביות נמוך (למשל 2400 סיביות לשניה) [2]. שיטה זו מבוססת על מודל יצירת הדיבור המתואר בציור (1.1), כאשר מניחים במודל כי המסנן $H(z)$ מכיל קטבים בלבד. החסרון של השיטה הוא בעיקר בחוסר טיבעיותו של הדיבור המשוחזר ובאי חסינותו לרעש ולטעויות ערוץ.



ציור (1.1): מודל יצירת הדיבור.

Fig. (1.1): Speech production model.

לצרכים מסוימים איכותו של הדיבור המשוחזר אינה מספקת ויש צורך לשפרה. לשם כך יש להוסיף מידע על זה הקיים במקדד LPC קלאסי. מידע זה קיים באות השארית אשר מתקבל על ידי סינון אות הדיבור במסנן ההפוך. על ידי שמוש באות השארית כאות הערוך ניתן לשחזר במדויק את אות הדבור. קידוד ישיר של אות השארית בתחום הזמן אינו יעיל כיון שקצב הסיביות הנדרש הוא גדול, דבר המוריד את כדאיות השימוש במודל החיזוי הלינארי בכללותו [15]. גישה שונה לקידוד אות השארית במישור הזמן, היא לקודר תחום תדרים צר (בתחום התדרים הנמוך base-band), ובצידו השני של הערוץ, לשחזר את מלוא תחום התדר של

השארית, על ידי הרחבת ספקטרום האות לתדרים גבוהים [11,12].

חסרונות שיטה זו:

1. פעולת יצירת התדרים הגבוהים יוצרת עיוות מורגש בדיבור.
2. אי ההתאמה הספקטרלית מתקזזת רק בתחום התדרים שקווד.

שיטות אחרות לקידוד השארית התרכזו בקידוד העוטפת הספקטרלית שלו במישור התדר, ללא התחשבות בפאזה, ושימוש במודל העירור הקלאסי (v/uv) , כאשר העוטפת הספקטרלית של השארית משמשת לעצוב ספקטרום אות העירור לפני הכנסו למסנן $H(z)$ [10].

כל השיטות שנסקרו לעיל, מנצלות את אות השארית כדי לשפר את ביצועי מערכת קידוד הדיבור. אולם אף אחת מהן אינה נותנת תשובה לכל המקורות לאי ההתאמה הספקטרלית אשר גורמת לחוסר טיבעיותו של הדיבור המשוחזר. הקידוד במישור הזמן אינו מנצל את העובדה שניתן במידה רבה לזוותר על קידוד הפאזה, ולכן קצב הסיביות בשיטות אלו גבוה. במודל קטבים ואפסים לא ניתנת תשובה מספקת לאיברים הנובעים מתכונות אי-רציונליות של ספקטרום אות הדיבור. בעוד שקידוד המעטפת הספקטרלית אינו יעיל בייצוג האפסים.

אחת השיטות [15] לשימוש באות השארית היא קידוד העוטפת הספקטרלית במישור הקפסטרום ושימוש במקדמי קפסטרום אלו במערכת הסינטזה, לעצוב ספקטרום אות העירור לפני היכנסו למסנן החיזוי הלינארי. שיטה זו יעילה לאור היכולת להפריד במישור הקפסטרום בין איברים השייכים לתופעות סינון כניסה קבועות וזאת על ידי הפחתת ממוצע לזמן ארוך של השארית הקפסטרולית. תהליך זה מפריד בין האיברים הקבועים המתאימים לפעולות סינון המשתנות לאט לבין איברים המשתנים מקטע לקטע, ההחסרה במישור הקפסטרום מקבילה לדה-קונוולוציה במישור הזמן.

כאשר אות הדיבור טבול ברעש מורגשת ירידה דרסטית באיכות הדיבור המשוחזר ממקדד ה-LPC. הסיבה לירידה זו באיכות הדיבור היא העובדה שאות הדיבור הרועש אינו מקיים את הנחת המודל של מסנן עם קטבים כלבד המודל המתאים לו במקרה זה הוא מודל קטבים ואפסים [13]. עובדה זו מפחיתה את יעילותו של אלגוריתם ה-LPC בשיערוך העוטפת הספקטרלית של אות הדיבור. השיטות המקובלות לקידוד השארית אינם מצליחות להתגבר על בעיה זו. דרך אחת

להתמודד עם הבעיה היא ניקוי אות הדיבור הרועש לפני ביצוע אלגוריתם ה-LPC כעבור מקדים [23,24]. דרך אחרת היא לשערך את מקדמי ה-LPC של האות הנקו מתוך האות הרועש [25]. השיטות המקובלות להקטנת השפעת הרעש על מקדרי ה-LPC הם כדלקמן:

1. הגדלת סדר המודל [15].
2. הפחתה ספקטרלית (SPECTRAL SUBTRACTION) [20-23].
3. הפחתת אוטוקורולציה [24].
4. משוואות Yule Walker מסדר גבוה [25].
5. שיערוך של מקדמי ההחזרה החלקית [25].
6. שיערוך קטבים ואפסים [26].

בעבודה זו מוצגות ונבחנות שתי שיטות למימוש מקדרי ה-LPC עבור אות דיבור רועש. הראשונה משתמשת בעובדה שהרעש הלכן המתווסף לדיבור משנה את המודל המיצג את הדיבור ממודל קטבים בלבד למודל קטבים ואפסים, כאשר מיקום הקטבים של אות-הדיבור נשמר, בעוד שהאפסים הם תרומתו העיקרית של הרעש. בשיטה זו משתמשים באלגוריתם הפרדה בין קטבים ואפסים במישור נגזרת ספקטרום הפאזה [27] ולאחר ההפרדה מוצע להשתמש בקטבים למימוש מסנן הסינטזה, בשיטה זו יש אפשרות להקטין את השפעת הרעש אולם מסתכנים באיבוד מידע על אפסים ששייכים לאות הדיבור עצמו במידה ואות זה חורג ממודל קטבים בלבד.

בשיטה השנייה מופעל אלגוריתם ההפחתה הספקטרלית לניקוי דיבור רועש [20-23], כמסנן כניסה למקדרי ה-LPC עם קידוד השארית הקפסטורלית. אלגוריתם ההפחתה הספקטרלית מסתפקים בשיערוך הספקטרום המנוקה וזה משמש ככניסה למקדרי. בשלב המימוש של מקדרי זה התעוררה בעית יציבות ובעבודה מוצעת שיטה לפיתרון בעיה זו. כן הסתבר שהספקטרום המנוקה אינו מתאים לשימוש מקדרי החיזוי הלינארי עם קידוד השארית הקפסטורלית בגלל היותו בעל תחום דינמי רחב כתוצאה מקפיצות גדולות בנקודות אי רציפות הנובעות מתכונות אלגוריתם ההפחתה הספקטרלית. לפיכך לא ניתן להשתמש באלגוריתם המקובל יחד עם קודד השארית הקפסטורלית. לכן נדרש היה לעשות שינוי באלגוריתם ההפחתה הספקטרלית על מנת להתאימו כמסנן כניסה למקדרי ה"ל".

מכנה החיבור:

פרק 2 - מציג את אלגוריתם ה-LPC ודן בתכונותיו.

פרק 3 - מציג את מקדד LPC עם קידוד השארית הקפסטרלית עבור דיבור נקי.

פרק 4 - דן בהשפעת הרעש על מקדד ה-LPC ומציג מספר שיטות לניקוי הרעש.

פרק 5 - מציג מקדד לאות דיבור רועש המבוסס על אלגוריתם הפרדה בין קטבים

לאפסים במישור נגזרת ספקטרום הפאזה (DPS).

פרק 6 - מציג את מקדד החיזוי הלינארי עם קידוד השארית הקפסטרלית עבור

אות דיבור רועש ודן כביצועי מקדד זה.

פרק 7 - דן במסקנות על המקודדים שהוצגו בפרקים 5 ו 6.

נספח א - מציג את נגזרת ספקטרום הפאזה (DPS) עבור מסננים מסדר ראשון

ושני.

פרק שני: חיזוי לינארי של דיבור (LPC).

2.1 מבוא [1].

קימות שתי שיטות עיקריות לקידוד אותות דיבור למטרות תקשורת ספרתית. הראשונה נקראת שיטת קידוד האות (Waveform coding), והשנייה נקראת קידוד פרמטרי (Parametric coding or Vocoding).

בשיטה הראשונה מתרכז המאמץ בהעברת האות הזמני בצורה החסכונית ביותר (בסיביות) אל הצד שכנגד, וזה משחזר את אותו אות זמני. בשיטה זו, המתחלקת למספר תת שיטות, הנחות היסוד לגבי מבנה אות הדיבור הן מינימליסטיות. במקרה הישיר ביותר (PCM) ההנחה היחידה היא לגבי רוחב הסרט של הדיבור, השיטות הדיפרנציאליות (DM, DPCM) מניחות גם תלות בין הדגימות (קורולציה).

יתרונות השיטה:

1. שיטת קידוד זו מהירה וזולה למימוש בזמן אמיתי.
2. שיטה זו חסינה באופן יחסי, לטעויות בערוץ התיקשורת ולרעש המתווסף לאות הדיבור.

חסרונה העיקרי של שיטת קידוד האות הוא שלקבלת איכות טובה של אות הדיבור המשוחזר יש צורך בקצב סיביות גדול (יחסית) דרך ערוץ התיקשורת: בתחום 16,000-64,000 סיביות לשניה.

בקידוד פרמטרי המאמץ מתרכז במצוי והעברת פרמטרים המיצגים ומשמרים תכונות מסוימות של הדיבור, האחראיות לאיכות ומובנות הדיבור. אין הכוונת בשיטה זו לעקוב בצורה מדויקת אחרי האות, ומסתפקים כד"כ בקירוב הערך המוחלט של הספקטרום, כאשר מסתמכים על תוצאות ניסויים המראות שזהו החלק החשוב ביותר לאיכותו ומובנותו של הדיבור.

הנחות היסוד למבנהו של אות הדיבור, בשיטת קידוד זו, רחבות מאוד. מגדירים מודל למבנהו של האות, ומסתפקים בהעברת המידע על מבנה זה. עצם הנחת המודל היא יתרונה הגדול של השיטה, וכפי שנראה בהמשך החיבור, גם חסרונה העיקרי. הנחת מודל מתאים, בשני צידי ערוץ התיקשורת, מקטינה את

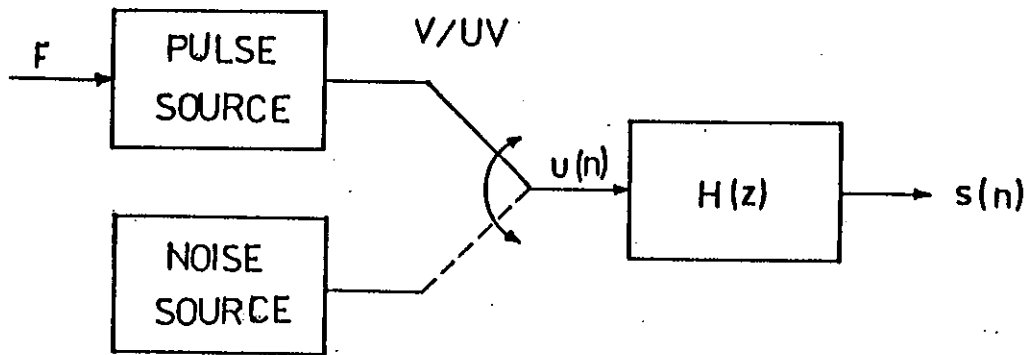
כמות המידע שנותר להעביר בערוץ ומקטינה לכך את קצב הסיביות שיש להעביר בערוץ. מידת ההקטנה היא לעיתים עד כדי סדר גודל, יחסית למקובל בשיטת קידוד האות, וניתן להגיע לקצב של 2400 סיביות לשניה, ואפילו לקצבים נמוכים יותר. צריך בשלב זה להדגיש שיש לשלם בעד ירידה זו בקצב המידע, התשלום הוא בחלקו בירידת איכות הדיבור המתקבלת, ובעיקרו ירידה מודגשת בעמידות הקידוד ברעשים הנלווים לאות ובטעויות ערוץ.

אחת הדרכים המקובלות ביותר למימוש הקידוד הפרמטרי היא החיזוי הלינארי (LPC). חיזוי לינארי היא שיטה מוכרת, יעילה ופשוטה לשיערוך ספקטרום הספק של אותות, אשר מבנה הספקטרום שלהם הנו רציונלי, ומכיל קטבים בלבד.

המשך הפרק הנוכחי יוקדש רובו ככולו להצגתו וניתוחו של מקדד החיזוי הלינארי של אותות דיבור נקיים מרעש. בסופו של הפרק תוקדש תשומת לב לחסרונות השיטה, ולכעית ההתמודדות ברעש בשיטת קידוד זו, נושא שיורחב בפרקים האחרים.

2.2 מודל מערכת יצור הדיבור [2].

הנחת היסוד שעליו מבוסס מקדד החיזוי הלינארי (LPC), הוא המודל בציור (2.1) המתאר את מערכת יצירת הדיבור. כל קטע דיבור, להלן קטע אנליזה, נוצר ממערכת זאת.



ציור (2.1): מודל יצירת הדיבור.

Fig. (2.1): Speech production model.

תיאור המודל:

1. $s(n)$ (אות הדיבור) הוא תוצר של ערוך המסנן $H(z)$ על ידי אות הערוך $u(n)$.
2. $H(z)$ מערכת לינארית, וקבועה בזמן עבור כל קטע אנליזה. $H(z)$ משתנה לאט מקטע אנליזה אחד למשנהו.
3. $H(z)$ מערכת רציונלית בעלת קטבים בלבד. פונקצית התמסורת של $H(z)$ ניתנת לרישום:

$$H(z) = \frac{G}{A(z)} \quad (2.1)$$

כאשר,

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (2.2)$$

וכאשר a_k ו G מספרים ממשיים. $A(z)$ נקרא המסנן ההפוך (הסבר לשם זה יובא בהמשך).

הערה: הנחת הקטבים בלבד אינה הכרחית ואינה נכונה באופן כללי, וישנם שיטות לקידוד פרמטרי שאינם משתמשות בה. בשיטת החיזוי הלינארי הנחה זו הכרחית כדי להגיע לפתרון של מערכת משוואות לינארית. בהמשך יוכח שגם אם הנחה זו שגויה ניתן לכפר עליה במחיר סדר המודל (p) .

4. אות הערור $u(n)$ מתחלק לשני סוגים קולי (VOICE) ואל-קולי (UNVOICED):

- א. קולי: טור הלמי יחידה בעל תדר קבוע בכל קטע עבודה. תדר זה נקרא התדר היסודי של הדיבור, המכונה בספרות תדר ה- PITCH וכך יכונה לאורך חיבור זה. תדר זה משתנה לאט מקטע עבודה אחד למשנהו.
- ב. אל-קולי: רעש לבן.

שתי תכונות משותפות לשני סוגי הערור:

- א. לכל קטע עבודה, האנרגיה בקטע מנורמלת ליחידה, (האנרגיה האמיתית נלקחת בחשבון דרך מקדם ההגבר G של המסנן $H(z)$).
- ב. ספקטרום הערור לבן (בקטע קולי תכונה זו נכונה למחזור PITCH בודד).

בשימוש בהנחות היסוד שנמנו לעיל, תפקיד מערכת הקידוד הוא לחשב וליצג את הפרמטרים המאפשרים את שיחזור ספקטרום הדיבור.

הפרמטרים אותם יש לחשב בשיטת החיזוי הלינארי:

1. $a_k, k=1, \dots, p$ מקדמי הפולינום המגדיר את $H(z)$.
2. G מקדם ההגבר של $H(z)$.
3. החלטת (v/uv) - סוג הערור, חישוב תדר ה-PITCH במקרה של החלטת v .

בסעיף הבא ינתנו הכלים לחישוב הפרמטרים.

2.3 ניתוח האלגוריתם וחישוב הפרמטרים [2].

בסעיף זה ינותח אלגוריתם החיזוי הלינארי באופן עקרוני בלבד. הניתוח יתבסס על הכתוב ב-[2], כאשר הדיון יתמקד באותות דטרמניסטים בלבד, דבר המאפשר הכנת עקרונות השיטה. עבור אותות אקראים התוצאות הן דומות, ושונות רק בהגדרת פונקציית האוטוקורלציה. במקרה המעשי, למרות שאות הדיבור הוא אות אקראי, הרי מחשבים את פונקציית האוטוקורלציה מתוך פונקציית מדגם אחת, דבר המנון את התהליך לבעיה דטרמיניסטית.

הניתוח יתמקד בשיטת האוטוקורלציה (יוסבר בהמשך) מכיון שזו נבחרה למימוש בעבודה הנוכחית.

2.3.1 שערוך מקדמי החיזוי $(a_k, k=1, \dots, p)$.

2.3.1.1 מודל הקטבים.

יהי $s(n)$ אות המקיים

$$s(n) = - \sum_{k=1}^p a_k s(n-k) + Gu(n) \quad (2.3)$$

כאשר $s(n-k)$, $k=1, \dots, p$, הם ערכי האות בעבר, $u(n)$ אות הכניסה. במישור z מתקבל מ (2.3):

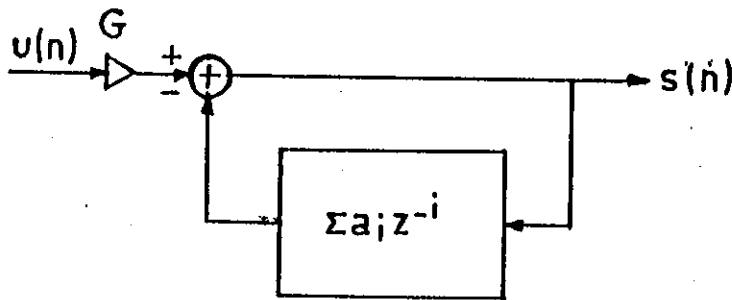
$$S(z) = H(z)U(z) \quad (2.4)$$

כאשר :

$$H(z) = \frac{G}{A(z)} \quad (2.5)a$$

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (2.5)b$$

ראה ציור (2.2).



ציור (2.2): מודל קטבים בלבד דיסקרטי.

Fig. (2.2): Discrete all pole model.

$A(z)$ נקרא המסנן ההפוך או המסנן המלכיון (לשם השני ינתן הסבר בהמשך). המודל נקרא קטבים בלבד מכיון שהמונה של $H(z)$ מתאפס ב- $z=0$ בלבד.

2.3.1.2 הצגת הבעיה ופתרונה.

נתון אות $s(n)$ המקיים את (2.3), $u(n)$ אות כניסה בלתי ידוע. נגדיר $\hat{s}(n)$ החזאי הלינארי של $s(n)$ מתוך $s(n-k)$, $k=1, \dots, p$, כלומר:

$$\hat{s}(n) = - \sum_{k=1}^p a_k s(n-k) \quad (2.6)$$

ונגדיר את אות השגיאה (השארית) $e(n)$:

$$e(n) = s(n) - \hat{s}(n) \quad (2.7)$$

ותהי E אנרגיית השגיאה על פני כל קטע העבודה

$$E = \sum_n e^2(n) \quad (2.8)$$

מצא את מקדמי החיזוי $a_k, k=1, \dots, p$, כך שיתקבל מינימום על E (בעיה זו מקבילה לבעיית מינימום שגיאה ריבועית ממוצעת בשערור לינארי של אותות אקראיים).

פתרון:

הפתרון הישיר לבעיה זו הוא בשיטת הגזירה של E לפי $a_k, k=1, \dots, p$ (קימת גם אפשרות להשתמש בכלל הניצבות בשיערוך לינארי). כלומר כדי למצוא מינימום על E לפי $a_k, k=1, \dots, p$, צריך לפתור את מערכת המשוואות מסדר p:

$$\frac{dE}{da_i} = 0, \quad i=1, \dots, p \quad (2.9)$$

ממשואות (2.6) ו (2.7) (2.8)

$$E = \sum_n e^2(n) = \left\{ \sum_n [s(n) + \sum_{k=1}^p a_k s(n-k)] \right\}^2 \quad (2.10)$$

ובגזירת (2.10) לפי (2.9) מתקבל

$$\sum_{k=1}^p a_k \sum_n s(n-k)s(n-i) = -\sum_n s(n)s(n-i) \quad i=1, \dots, p \quad (2.11)$$

זוהי מערכת משוואות לינארית מסדר p, הידוע בספרות כמערכת המשוואות הנורמלית.

אם נציב את התוצאה (2.11) ב (2.10) נקבל ביטוי לאנרגיה השגויה המינימלית E_p עבור חזאי מסדר p :

$$E_p = \sum_n s^2(n) + \sum_{k=1}^p a_k \sum_n s(n)s(n-k) \quad (2.12)$$

הפתרון מכאן והלאה יעקב אחרי שיטת האוטוקורלציה שנבחרה בעבודה זו. השיטה הנוספת, שיטת הקוריאנס, שונה למראית עין רק בהגדרת קטע העבודה הרלוונטי אולם שינוי זה מוביל הן לשינוי בתכונות הפתרון והן לשינוי בסיבוכיותו.

2.3.1.3 שיטת האוטוקורלציה

בשיטה זו ההנחה היא שהאות $s(n)$ קים בתחום $n=0, \dots, N-1$, ומתאפס מחוץ לתחום זה. בהצבה של תחום זה נגדיר את האות $R(i)$ כדלקמן:

$$R(i) = \sum_{n=0}^{n-i-1} s(n)s(n-i) \quad , \quad i \geq 0 \quad (2.13)$$

$R(i)$ היא פונקציית האוטוקורלציה של האות הדיטרמיניסטי $s(n)$ ומקימת את התכונות הבאות:

$$R(i) = R(-i) \quad (2.14)$$

$$R(0) \geq |R(i)| \quad (2.15)$$

אם נציב את (2.13) במערכת המשואות הנורמלית (2.11) תתקבל מערכת המשואות:

$$\sum_{k=1}^p a_k R(i-k) = -R(i) \quad , \quad i=1, \dots, p \quad (2.16)$$

או בצורה מטריצית

$$\underline{\underline{RA}} = -\underline{\underline{R}} \quad (2.17)$$

זוהי מערכת משואות לינארית אשר פתרונה הישיר דורש הפיכת המטריצה \underline{R} . אולם אם ננצל את התכונות (2.14), (2.15) של $R(i)$ אשר ממנה מורכבת המטריצה \underline{R} אזי נקבל את התכונות הבאות עבור המטריצה \underline{R} :

1. המטריצה סימטרית.

2. הערכים על כל אלכסון שווים (מטריצת טופליץ).

3. המטריצה מוגדרת חיובית (עבור חבורת אותות Σ^2 אשר נניח שהדיבור נמנה עמם).

התכונה השלישית מבטיחה פתרון למערכת המשואות הנורמלית שהתקבלה כיוון שהמטריצה הפיכה. תכונה זו גם מבטיחה (תאורטית) שכל השורשים של הפולינום $A(z)$, (2.5) נמצאים בתוך מעגל היחידה במישור z , דבר המבטיח יציבות למסנן $H(z)$ בשיטת חישוב זו.

הערה: במקרה המעשי היציבות של $H(z)$ אינה מובטחת בגלל בעיות של דיוק בחישובים.

העובדה ש- \underline{R} היא מטריצת טופליץ והעובדה שבאגף הימני של (2.17)

יושבת עמודה של אותה מטריצה, מאפשרת לפתור את מערכת המשואות (2.17) בצורה רקורסיבית. בסעיף הבא יובא הפתרון הרקורסיבי לפי Durbin.

2.3.1.4 חישוב רקורסיבי של מקדמי החיזוי הלינארי.

שיטתו של Durbin לחישוב מקדמי החיזוי תינתן להלן ללא הוכחה. הדגש יושם על התכונות והתוצאות של אלגוריתם זה.

האלגוריתם של Durbin.

$$E_0 = R(0) \quad (2.19)a$$

$$k_i = -[R(i) + \sum_{j=1}^{i-1} a_j^{(j-1)} R(i-j)] / E_{i-1}, \quad i=1, \dots, p \quad (2.19)b$$

$$a_i^{(i)} = k_i$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)}, \quad j=1 \dots i-1 \quad (2.19)c$$

$$E_i = (1 - k_i^2) E_{i-1} \quad (2.19)d$$

משוואות ב (2.19), ד (2.19) נפתרות רקורסיבית עבור $i=1, \dots, p$ והתוצאה הסופית המתקבלת:

$$a_j = a_j^{(p)} \quad (2.19)e$$

כאשר $a_j, j=1, \dots, p$, הם מקדמי החזאי הלינארי מסדר p .

תכונות ותוצאות האלגוריתם:

1. בפתרון של החזאי הלינארי מסדר p מחושבים גם החזאים מכל סדר נמוך יותר.
2. תוצאת לואי נוספת, היא חישוב אנרגיית השגיאה המינימלית, E_i , $i=0, \dots, p$, (תוצאה זו חשובה לחישוב, G , מקדם ההגבר). כן ניתן לראות ש- E_i קטנה ככל שהסדר i עולה, (בהנחה ש $|k_i| < 1$).
3. תוך כדי חישוב האלגוריתם מחושבים k_i , המכונים בספרות מקדמי החזרה. חשיבות פרמטרים אלו שהם מאפשרים בדיקה נוחה של קיום יציבות המערכת $H(z)$.

נעבור כעת לחישוב ההגבר G .

2.3.1.5 חישוב מקדם ההגבר G .

בהצגת הבעיה הונח שהכניסה $u(n)$ אינה ידועה. הנחה זו מונעת את האפשרות בחישוב מקדם ההגבר G .

אולם אם נחזור לדיבור ולמודל יצירת הדיבור אזי שם הונח ש- $u(n)$ יכול לקבל אחת משתי צורות אפשריות, כאשר בשתיים האנרגיה הכוללת בקטע העבודה מנורמל ל- 1.

הנחה זו מאפשרת לחשב את G . ממשוואות (2.6), (2.7), נקבל

$$e(n) = Gu(n) \quad (2.20)$$

$$E_p = \sum_{n=0}^{N-1} e^2(n) = G^2 \sum_{n=0}^{N-1} u^2(n) \quad (2.21)$$

ומהנחה שהאנרגיה שווה ל-1 נקבל

$$\sum_{n=0}^{N-1} u^2(n) = 1 \quad (2.22)$$

ולפיכך

$$G^2 = E_p \quad (2.23)$$

כלומר מקדם ההגבר עבור חזאי מסדר p שווה לאנרגיית אות השארית באותו קטע.

מסקנות

1. מקדם ההגבר G הוא תוצאת לוואי של חישוב אלגוריתם Durbin.

2. עבור אות המקינים את הנחת הקטבים בלבד מסדר p , אנרגיית השגיאה אינה

יכולה להיות קטנה מ- G^2 .

2.3.1.6 חיזוי לינארי ושערוך ספקטרום.

שיטת החיזוי הלינארי היא שיטה לשיערוך ספקטרום הדיבור, בסעיף זה

נציג את הקשר ביניהם.

נגדיר:

$P(w)$ ספקטרום ההספק של $s(n)$.

$\tilde{P}(w)$ ספקטרום ההספק של $\tilde{s}(n)$.

$E(w)$ ספקטרום השגיאה $e(n)$.

- 22 -

.H(z) של תגובת החדר של H(w)

.A(z) של תגובת החדר של A(w)

אזי מהגדרת s(n) יתקבל

$$P(w) = |H(w)|^2 = \frac{G^2}{|A(w)|^2} \quad (2.24)$$

מהגדרת e(n)

$$P(w) = \frac{|E(w)|^2}{|A(w)|^2} \quad (2.25)$$

ממשפט פרטוול

$$E = \sum_n e^2(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |E(w)|^2 dw \quad (2.26)$$

בהצבה של (2.24) ו (2.25) ב (2.26) מתקבל

$$E = \frac{G^2}{2\pi} \int_{-\pi}^{\pi} \frac{P(w)}{P(w)} dw \quad (2.27)$$

כלומר בחיזוי לינארי משוערך ספקטרום ההספק במוכן של מינימום על (2.27) ובשיערוך מושלם האינטגרל שווה לאחד.

מתוצאה זו ומהעובדה ש- p מקדמי האוטוקורולציה של אות המקור ואות החזאי שווים עבור חזאי מסדר p, מתברר ש- P(w) אינה חיבת לכלול קטבים בלבד, ואף יותר מזה, אינה חיבת להיות רציונלית, מספיק ש- P(w) תהיה פונקציה בעלת התמרת פוריה המקימת את תכונות פונקציית האוטוקורולציה כדי שניתן יהיה להגיע לשערוך מושלם. התשלום הוא כמוכן בסדר החזאי הדרוש.

2.3.1.7 בעית היציבות.

באופן תיאורטי ובהנחה שמקדמי האוטוקורולציה חושבו מאות השיך למשפחת L^2 , אזי מטריצת האוטוקורולציה R מוגדרת חיובית ולכן האפסים של המסנן ההפוך $A(z)$, המוגדר ב - b(2.5), שוכנים בתוך עיגול היחידה במישור z , דבר המבטיח את יציבות המסנן $H(z)$. אולם בגלל בעיות מעשיות של חישוב באורך מלה סופי של מקדמי האוטוקורולציה עלולה לצוץ בעית יציבות, ואכן נתקלים בד"כ בבעיה זו.

בגלל בעיות מעשיות אלו יש צורך בבדיקת המסנן $H(z)$ המתקבל או לחליפין לבדוק את מיקום האפסים של המסנן ההפוך $A(z)$ ולוודא שהם בתוך מעגל היחידה. פתרון האפסים של פולינום מסדר גבוה היא בדרך כלל בעיה מסובכת, יקרה בזמן ועדיף להימנע ממנה.

דרך ישירה ומקובלת לפיתרון בעית אי היציבות מסתרת באלגוריתם של Durbin ובמקדמי ההחזרה.

ניתן להראות שתנאי הכרחי ומספיק ליציבות המסנן $H(z)$ הוא חיוביות אנרגית השגיאה הנוותרת, כלומר מספיק לבדוק

$$E_i > 0, \quad i=1, \dots, p \quad (2.28)$$

וממשואות (2.19)a-(2.19)d מתקבל התנאי המקביל

$$|k_i| < 1, \quad i=1, \dots, p \quad (2.29)$$

כלומר אלגוריתם Durbin נותן כלי לבדיקת היציבות מערכת החיזוי הלינארי המתקבלת. בהמשך העבודה נראה דרך לשימוש בתכונות האלגוריתם לפתרון בעית אי היציבות.

2.3.1.8 קוונטיזציה של המקדמים.

בעית הקוונטיזציה מתחלקת לשניים:

1. כמה סיביות לתת לכל מקדם מתוך כלל הסיביות, כדי לקבל עיוות מינימלי בספקטרום.

2. כיצד להבטיח יציבות למסנן המתקבל לאחר הקוונטיזציה.

הבעיה הראשונה דורשת סדר כלשהו בין המקדמים, כך שתוכל להתבצע בדיקה סטטיסטית של המקדמים על העיוות הספקטראלי. השניה דורשת דרך ישירה ככל האפשר לבדיקת היציבות.

קימות מספר קבוצות מקדמים, אשר הקשר ביניהם הוא חד-חד ערכי, כדוגמא

נציג מספר קבוצות:

1. a_k , $k=1, \dots, p$, מקדמי הפולינום $A(z)$.
 2. z_k , $k=1, \dots, p$, האפסים של הפולינום הנ"ל.
 3. k_i , $i=1, \dots, p$, מקדמי ההחזרה.
 4. c_k , $k=1, \dots, p$, מקדמי הקפסטרום (יוסבר בהמשך).
- הבעיה היא במי מבין הקבוצות לבחור.

קבוצת מקדמי ההחזרה היא היחידה המקימת את שני התנאים, הסדר בה הוא ברור וגם על בעית בדיקת היציבות היא עונה (ראה סעיף מתאים). לכן מקובל לבחור קבוצה זו, או קבוצה הקשורה אליה ישירות ומשמרת את שתי התכונות, לקידוד והעברה בערוץ.

2.3.2 בעית V/UV וחישוב תדר ה-PITCH.

להשלמת מודל יצירת הדיבור נותרה עדין בעית ההחלטה לגבי סוג העירור הדרוש בקטע הרלוונטי, וחישוב תדר ה-PITCH, באם הוחלט שהקטע הוא קולי. בכל השיטות המקובלות שתי החלטות אלו תלויות אחת בשניה.

השאלה האם הקטע הנו קולי או לא, קשה מאוד לפתרון מושלם. בכל השיטות הקיימות הסכויים לטעות אינם מבוטלים. לעובדה זו יש השפעה על מובנות ואיכות הדיבור המשוחרר. אין בכונתי להרחיב את הדיבור על סוגיה זו, מכיון שאין

בעבודה זו תרומה כלשהי לבעיה. לכן אסתפק בהצגת מספר דרכים מקובלות לפתרון הבעיה, ואתן את הסיבות לבחירת המסוימת בעבודה זו.

- הפתרון של שתי הסוגיות נעשה כך, תחילה מחשבים מה שנראה כתדר ה-PITCH בקטע, ולאחר מכן בודקים אם תכונות התוצאה שהתקבלה עבור תדר ה-PITCH מתאימות לקטע קולי ואם לאו.

2.3.2.1 חישוב תדר ה-PITCH בשלט האוטוקורולציה [3].

עבור אות מחזורי, פונקציית האוטוקורולציה מחזורית גם כן ובעלת אותו זמן מחזור. בגלל תכונת המצוץ בחשוב הקורולציה ישנו אפקט של ניקוי מרעש ומחזוריות זו מודגשת יותר.

עבור פונקציית האוטוקורולציה

$$R(0) \geq |R(n)|, \quad \text{לכל } n \quad (2.30)$$

ואם האות מחזורי אזי

$$R(nT) = R(0) \quad (2.31)$$

כאשר T הוא זמן המחזור.

כלומר עבור אות מחזורי ממש, מספיק למצוא את המקדם אשר שווה בערכו למקדם האפס, והרי המחזור. ואם האות הוא דיבור, מספיק לבדוק בתחום המותר לתדר ה-PITCH ולבצע החלטה מתוך ערכי הפונקציה באותו תחום. מכיון שאות הדיבור אינו מחזורי ממש מוצאים מקסימום באותו תחום.

אלגוריתם ההחלטה הוא כדלקמן:

1. מצא מקסימום של פונקציית האוטוקורולציה בתחום המתאים למחזור ה-PITCH, של אות הדיבור.
2. בדוק את גודלו של האיבר הנבחר ב-1 ביחס לגודלו של $R(0)$ (האנרגיה בקטע הרלונטי), לפי יחס הגדלים החלט אם הקטע קולי או לא, ואם כן מצא את המחזור ב-1.

באופן מעשי הדיבור אינו מחזורי במדויק, בקטעים סמוכים ואפילו בקטע האנליזה עצמו מחזור ה-PITCH אינו קבוע. תכונות אילו גורמות לאיבר המתאים למחזור להיות קטן בהרבה מאיבר האפס, בפונקציה האוטוקורולציה ולכן לטעויות בהחלטת (v/uv) . מכנה הפונקציה האוטוקורולציה מושפע גם ממכנה ומיקום הפורמנטים של הדיבור, תכונה הגורמת להופעת ערכי שיא הנגרמים על ידי תדרי תהודה אילו (ראה ציור 2.4), ואשר גורמים להחלטות שגויות לגבי תדר ה-PITCH. קימות שיטות מספר להתגבר על בעיה זו ואחת מהן היא שיטת ה-SIFT.

2.3.2.2 אלגוריתם ה-SIFT [4].

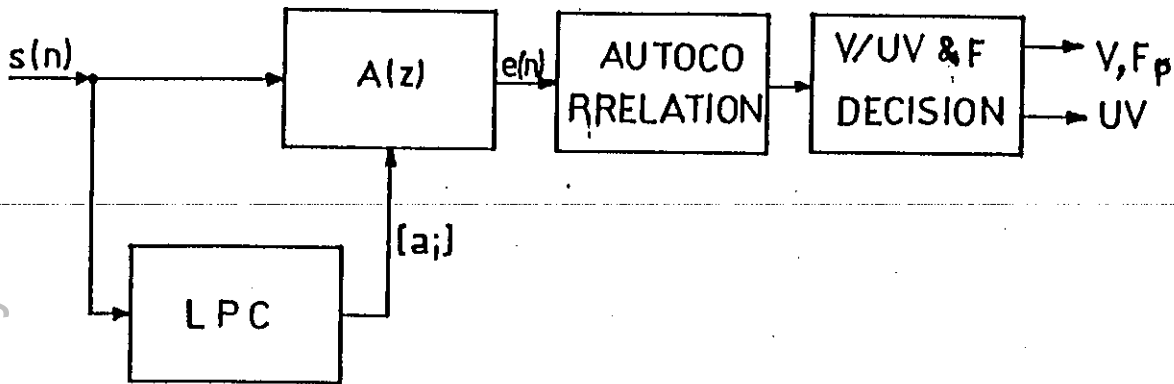
שיטה זו באה להתגבר על השפעת תדרי התהודה של הדיבור על פונקציה האוטוקורולציה, על ידי מציאתם והרחקתם. מתוך המידע שצברנו בפרק זה, ברור שאות השארית הוא בעל התכונות הרצויות.

האלגוריתם:

1. מצא את המסנן ההפוך $A(z)$ (מכיוון שרצוי לבטל בעקר את השפעתם של תדרי התהודה הראשונים בלבד, מסתפקים במסנן מסדר $p=4$), וחשב את אות השארית.
2. חשב את פונקציה האוטוקורולציה של אות השארית, ומצא את המקדם המקסימלי שלה בתחום המותר למחזור ה-PITCH.
3. חשב את פונקציה האוטוקורולציה של אות הדיבור, ובדוק את ערכו של המקדם המתאים במקומו למקום שחושב ב-2, השווה אותו לסף כפי שבוצע בצעד 2 בשיטת האוטוקורולציה והמשך כמו שם.

ראה ציור (2.3) להסבר התהליך.

שיטה זו יקרה בחישובים מקודמתה, אך בלצועיה משופרים בהרבה.



ציור (2.3): שיערוך ה-Pitch בשיטת ה-SIFT.

Fig. (2.3): SIFT Pitch estimation.

2.3.2.3 חישוב ה-PITCH במישור הקפסטרום [5].

הקפסטרום הממשי מוגדר כהתמרת פוריה של לוגריטם הערך המוחלט של הספקטרום.

יהי $s(n)$ אות כלשהו ויהי $c(n)$ הקפסטרום שלו אזי

$$c(n) = F^{-1}\{\log(|s(w)|)\} \quad (2.32)$$

הקפסטרום הוא למעשה ציר "זמן" חדש, אשר שומר על תכונת המחזוריות. יתרונו שעבור אות בעל תכונות של אות הדיבור, מתבצעת הפרדה בין אות הערור לבין המסנן היוצר את הדיבור, ובו קל לבדוק את האיברים התלויים במחזוריות.

בצורה אנליטית. יהי $s(n)$ אות הדיבור אזי

$$s(n) = h(n)*u(n) \quad (2.33)$$

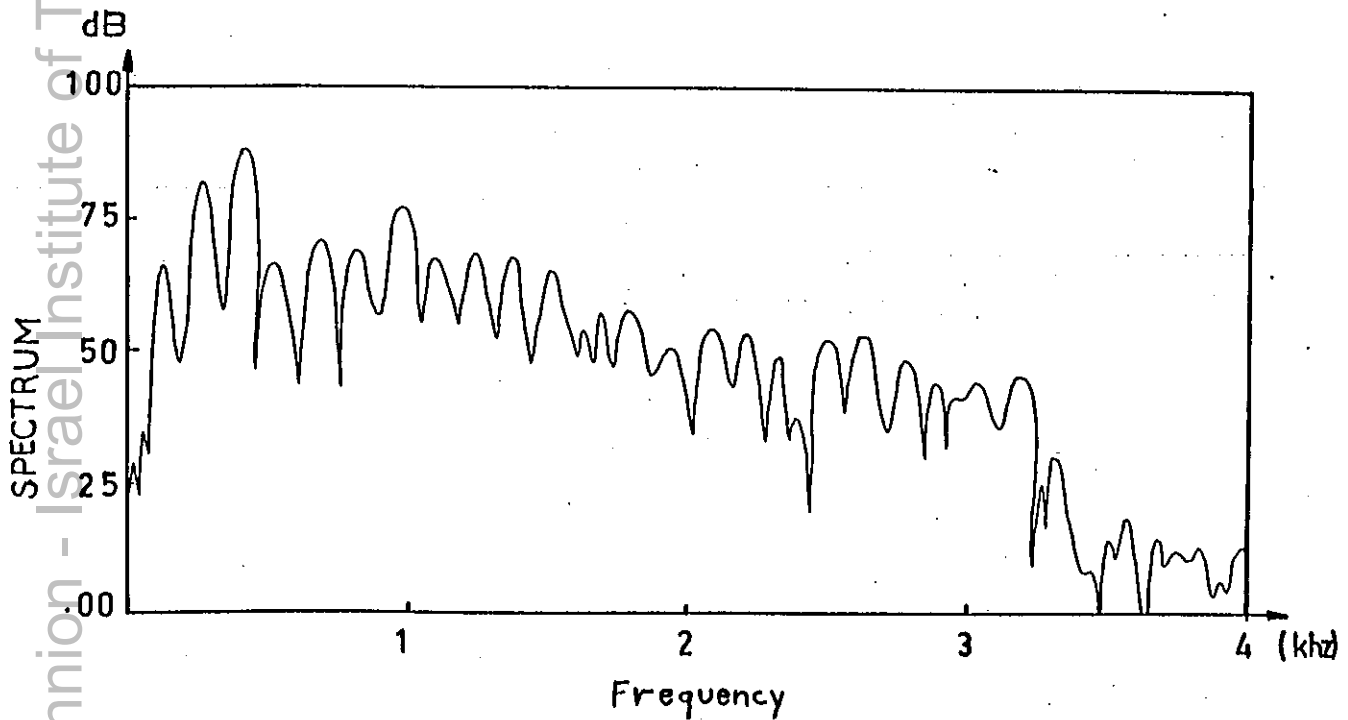
כאשר: $h(n)$ תגובה להלם של המסנן $H(z)$.
 $u(n)$ אות הערור.

אזי

$$S(w) = H(w)U(w) \quad (2.34)$$

$$\log|S(w)| = \log|H(w)| + \log|U(w)| \quad (2.35)$$

וכאן למעשה הופרדו ההשפעות של המסנן והערור (תופעה זו ידועה בשם דה-קונבולוציה). ואם נעיף מבט על ציור המכיל את $\log|S(w)|$ (ציור 2.4), הרי זה נראה כמו עוטפת איטית ($\log|H(w)|$) המתוספת לאות מהיר ($\log|U(w)|$) ועל ידי הפרדה בין שני אותות אילו בשיטות מקובלות בתחום הקפסטרום ניתן למצוא את האיבר המתאים למחזור ה-PITCH ולכצע את ההחלטות הרצויות.



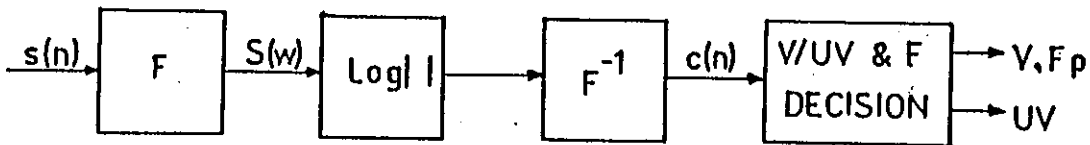
ציור (2.4): לוג הספקטרום של אות דיבור.

Fig. (2.4): Log spectrum of speech.

שלבי האלגוריתם:

1. חשב את הקפסטרום של אות הדיבור.
2. מצא בתחום זה את המקסימום בתחום המתאים למחזור ה-PITCH.
3. בדוק את גודלו של המקדם שחושב ב-2, ביחס ל $c(0)$ (איבר הנותן מוד לאנרגיה בקטע הרלונטי). אם גודלו מתאים (יחס ניסיוני) החלט - קטע קולי, והוסף אליה את חישוב המחזור כפי שחושב ב-2.

ראה ציור (2.5) להסבר התהליך.



ציור (2.5): זיהוי Pitch במישור הקפסטרום.

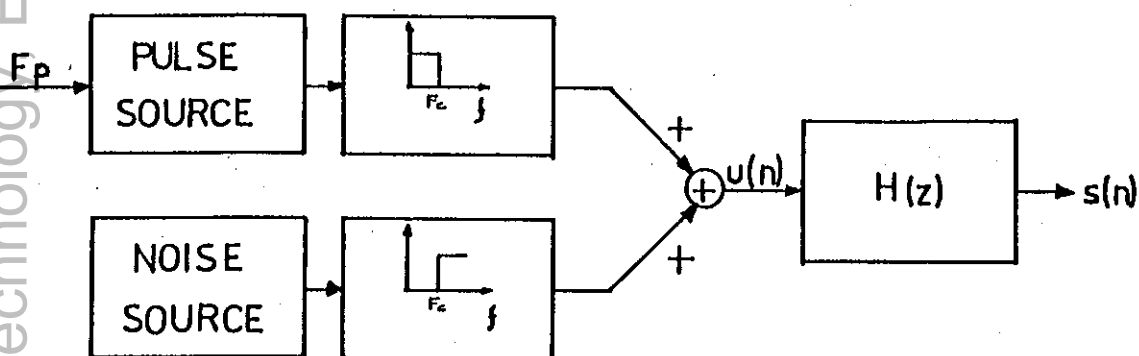
Fig. (2.5): Cepstral Pitch extraction.

ב- [6] הושגו השיטות שהובאו לעיל והמסקנה היתה ששתי השיטות האחרונות מקבילות בביצועיהן. החסרון בשיטת הקפסטרום הוא הסיבוכיות בחישובו של זה, אולם מכיון שבמערכת הניבדקת בעבודה זו, כפי שיוסבר בהמשך, הקפסטרום מחושב בין כה וכה, לכן נבחרה שיטת הקפסטרום.

2.3.2.4 שכלול מודל הערור MIXED SOURCE [7].

החלטת סוג הערור היא מסובכת בעלת סיכוי רב לטעות והשפעתה על איכות ומוכנות הדיבור המתקבל היא משמעותית. גם אם נניח שהחלטה היא מדויקת, עדין איכות הדיבור המתקבלת אינה מספקת. עובדה זו גררה בדיקת אפשרויות לשפור מודל הערור ואף לויתור עליו ועירור בעזרת אות השארית.

הסתבר לחוקרים שהערור מורכב למעשה מסכומם של שני אותות הערור המופיעים במודל הקלאסי, כאשר כל אחד מהם תופס תחום אחר בתחום התדרים של הערור. אות הערור המחזורי, המתאים לקטע קולי, תופס את תחום התדרים הנמוך, ואילו הרעש הלכן משלים את התדרים הגבוהים. ההצעה במודל הנוכחי היתה להצמר אומנם למודל הערור v/uv , אולם במקום לשלוח החלטה לגבי סוג הערור, ישלח פרמטר F_c אשר יגדיר את גבולות התדר של כל אחד משני סוגי הערור. ראה ציור (2.6).



ציור (2.6): מודל העירור המשוכלל.

Fig. (2.6): Mixed source excitation model.

שיטת לב שבהצבת F_c שווה ל $(0, MAX)$, המודל מתנוון למודל הקלאסי.

שיטה זו שיפרה במקצת את טבעיות הדיבור המשוחזר, אולם לא הוותה פריצת דרך בתחום זה. צעד נוסף קדימה בתחום זה היה קידוד אות השארית והשמוש בו כאות העירור, נושא שיורחב עליו הדיבור בסעיף הבא.

2.4 אות השארית בחיזוי לינארי [8-11,18].

אות השארית מכיל את המידע שנותר על אות הדיבור, לאחר מיצוי המידע על אודות מסנן יצור הדיבור, $H(z)$, בשיטת החיזוי הלינארי. וניתן לכן באמצעות להשלים את המודל לקבלת דיבור משוחזר בצורה מושלמת

$$E(z) = S(z)A(z) \quad (2.36)$$

כלומר

$$S(z) = E(z)H(z) \quad (2.37)$$

והרי $s(n)$ משוחזר בצורתו המדויקת.

המידע שנותר באות השארית:

1. מידע על אות הערור ועל הפאזה שלו, שלא נלקחה בחשבון במודל החיזוי הלינארי.
2. מידע על הקטבים הנוספים במידה וסדר המודל האמיתי גדול מזה שנבחר לשיערוכו
3. מידע על אפסים במדה והם קימים, שלא נלקחו בחשבון בהצגת המודל.
4. מידע על איברים הנגרמים על ידי חריגה של האות ממודל רציונלי.

עובדות אלו מצביעות על האפשרות להשתמש במידע המסתתר באות השארית, כדי לשפר את ביצועי מערכת קידוד הדיבור. ואכן נעשו מספר רב של עבודות בכיוון זה. השיטות שנבדקו לקידוד השארית הן במישור הזמן, במישור התדר ובמישור הקפסטרום. נושא זה יורחב בפרק השני בחיבור זה.

2.5 חסרונות שיטת החיזוי הלינארי בקידוד דיבור [12].

בסעיף זה יסקרו מקצת מחסרונות השיטה, חסרונות אלו מובילים לפיתוח מודל הכולל, נוסף על החזאי הלינארי, גם את קידוד השארית במישור הקפטרום. מודל זה יתואר בהרחבה בפרק הבא, ועונה על חלק מהחסרונות הקיימים בחיזוי לינארי כדלקמן:

1. החזאי הלינארי הנו מודל קטבים בלבד, אי לכך הוא לוקה ביצוג האפסים הקיימים בדיבור אמיתי. נבדקו אומנם שיטות לחישוב אפסים אילו, אולם התוצאות לא הצדיקו את הסיבוכיות הנוספת.
2. רוחב הסרט של הפורמנטים (תדרי התהודה של הקטבים) הנוצרים שגוי, דבר היוצר עיוות בדיבור המשוחזר.
3. מיקום הפורמנטים נוטה להיפגע, כיון שאלגוריתם החיזוי הלינארי, נוטה אל השיאים של הרמוניות ה-PITCH. הדבר בולט במיוחד כאשר תדר השיאים גבוה כן שהמרחק בין ההרמוניות גדול יחסית.
4. מקדמי המסנן ההפוך $A(z)$ משתנים עם מיקום מקום חלון האנליזה יחסית למקום הלמי ה-PITCH [17].
5. מודל הערור, v/uv , (BUZZ-HISS), מאבד מידע רב הכלול באות השארית, הכולל מידע על הפאזה של הדיבור, ועל ספקטרום האמור להשלים את $H(z)$.
6. גם אם נניח שמודל הערור v/uv הנו מושלם, אזי הטעויות בגילוי סוג הערור, ובשערוך תדר ה-PITCH, יוצרים עיוותים מאוד מורגשים באות המשוחזר.

2.6 חיזוי לינארי של דיבור ברעש לבן [13].

בחיזוי לינארי של אותות דיבור מניחים שספקטרום אות הדיבור $s(n)$, מכיל קטבים בלבד. כאשר בוחנים את איכות הדיבור המתקבלת, אזי יחסית לקצב הסיביות, 2400-4800 סיביות לשניה, איכות הדיבור משביעת רצון. אולם כאשר מתווסף לאות הדיבור רעש, הירידה באיכות הדיבור היא דרסטית, כבר ביחסי אות-לרעש גבוהים יחסית.

כאשר בוחנים את ספקטרום האות המשוחרר מתבררות שתי בעיות עקרוניות (גם אם נתעלם מבעיית שערור ה- PITCH והחלטת v/uv):

1. הפורמנטים, תדרי התהודה של הקטבים, נמרחים ואף נעלמים, ביחסי אות לרעש של 10dB ומטה. בדרך כלל נוטים להעלם הפורמנטים הגבוהים.
2. הפורמנטים זזים בתדר, ומתקבלים בתדרים שגויים.

כפי שיוסבר בפרק הרביעי, הירידה הדרסטית באיכות הדיבור נובעת מתרומת הרעש, המתבטאת בעקור כתוספת אפסים לספקטרום הדיבור, אשר מקשים על אלגוריתם ה- LPC בעקיבה אחרי הקטבים של הדיבור.

פרק שלישי: LPC עם קידוד השארית הקפסטרלית.

3.1 מבוא

LPC הנה שיטה יעילה לקידוד דיבור בקצב סיביות נמוך, אולם הדיבור המשוחזר לוקה בחסר באיכותו ובטיבעיותו. חסרונות אלו נובעות מחסרונות שיטת ה-LPC כפי שהוסברו בסעיף 2.5. אם נתעלם מהעיוותים הנובעים ממודל העירור, אזי חסרונות השיטה שנמנו בסעיף הנ"ל, מצביעות על העובדה, שהתרומה העיקרית לחסר באיכות הדיבור, נובע מאי ההתאמה בין ספקטרום הדיבור הממשי לזה של מסנן השיחזור $H(z)$. כן הוכח, ניסויות [14], שמובנות הדיבור נפגעת בעיקר מאי ההתאמה באמפליטודה של הספקטרום, ופחות מאי ההתאמה בפאזה.

אי התאמה זו נובעת מכך שמודל הקטבים בלבד, אינו מציג נכונה את אות הדיבור. אות דיבור טבעי מכיל בנוסף לקטבים אפסים ואיברים החורגים ממודל רציונלי. תוספות אלו נתרמות מתכונות הדובר ומסוגי סינון שעובר הדיבור עד הגיעו למערכת קידוד הדיבור.

על אי ההתאמה הספקטראלית, הנגרמת על ידי אפסים, ניתן להתגבר על ידי הצגת מודל הדיבור כבעל קטבים ואפסים, ופתרון בעית חיזוי לינארי זו. אולם מודל זה עדיין אינו מכסה על הגורמים הנוספים לאי ההתאמה הספקטראלית. השיפור באיכות הדיבור, באופן יחסי לסיבוכיות בפתרון המודל החדש (שאינו לינארי ברוב השיטות הידועות), אינו מצדיק מעבר למודל זה [15].

בפרק הקודם ראינו שהמידע החסר במסנן $H(z)$, כדי להתאימו לספקטרום הדיבור, נמצא באות השארית, עובדה המצביעה שיש מקום לנסות ולהשתמש באות זה כדי לתקן את עיוותי מודל ה-LPC. קיימות מספר גישות לשימוש באות השארית, להלן אמנה אחדות מהן.

שיטות קידוד אות סטנדרטיות בתחום הזמן נשללות מיד בגלל קצב הסיביות הגדול שהן דורשות, המוריד את כדאיות השימוש במודל החיזוי הלינארי בכללותו [15]. גישה שונה לקידוד אות השארית במישור הזמן, היא לקוד תחום תדרים צר, בתחום הנמוך, ובצידו השני של הערוץ, לשחזר את מלוא תחום התדר, של השארית, על ידי שיטות לשחזור תדרים גבוהים [11,12].

חסרונות שיטה זו:

1. פעולת שיחזור התדרים הגבוהים יוצרת עיוות מורגש בדיבור.
2. אי ההתאמה הספקטרלית מתקזזת רק כתחום התדרים שקודד.

שיטות אחרות לקידוד השארית התרכזו בקידוד העוטפת הספקטרלית של השארית, במישור התדר, ללא התחשבות בפאזה, ושימוש במודל העידוד הקלאסי (v/uv), כאשר העוטפת הספקטרלית של השארית משמשת לעצוב ספקטרום אות העידוד לפני הכנסו למסנן $H(z)$ [10].

כל השיטות שנסקרו לעיל, מנצלות את אות השארית כדי לשפר את ביצועי מערכת קידוד הדיבור. אולם אף לא אחת מהן אינה נותנת תשובה לכל המקורות לאי ההתאמה הספקטרלית. הקידוד במישור הזמן אינו מנצל את העובדה שניתן לוותר על קידוד הפאזה, ולכן קצב הסיביות בשיטות אלו גבוה. במודל קטבים-אפסים לא ניתנת תשובה מספקת לאיברים הנובעים מתכונות אי-רציונליות של ספקטרום אות הדיבור. בעוד שקידוד המעטפת הספקטרלית אינו יעיל בליצוג האפסים.

במודל שעליו מבוססת עבודה זו [15], מוצע להשתמש בקידוד השארית במישור הקפסטרום, מישור שבו ניתן להפריד בדרך נוחה, בין הגורמים השונים המאפיינים את אי ההתאמה הספקטרלית שעליה דובר לעיל, ולקודד כל אחד מהם בנפרד.

בהמשך הפרק יוצג מודל זה של מקודד חיזוי לינארי עם קידוד השארית הקפסטרלית. לפני הצגת המודל אקדיש מספר שורות להצגת הקפסטרום, ומנית מספר מתכונותיו שישמשו אותנו בהמשך.

3.2 הקפטרומ [3].

3.2.1 הקפטרומ מהו ?

הקפטרומ הקומפלקסי הינו התמרת פוריה של לוג הספקטרומ של האות.

יהי $s(n)$ אות ויהי $\hat{h}(n)$ הקפטרומ הקומפלקסי שלו, אזי

$$\hat{h}(n) = F^{-1}\{\log[S(w)]\} \quad (3.1)$$

כאשר $S(w)$ הוא הספקטרומ של $s(n)$. מבחינת יחידות מקביל תחום הקפטרומ לתחום הזמן.

הערה: יש לשים לב שפעולת ה- \log מתבצעת על מספר קומפלקסי, עובדה המצריכה זהירות יתר בטיפול בפאזה, בכדי שההתמרה הנ"ל תהיה הומומורפית.

בדרך דומה מוגדר הקפטרומ הממשי (להלן הקפטרומ), אלה שהפעם פעולת ה- \log מתבצעת על הערך המוחלט של הספקטרומ $S(w)$.

$$c(n) = F^{-1}\{\log[|S(w)|]\} \quad (3.2)$$

3.2.2 תכונות הקפטרומ והקפטרומ הקומפלקסי

1. דה-קונוולוציה - במישור הקפטרומ הקומפלקסי, אות המתקבל מקונוולוציה של שני אותות, מופרד לסכום הקפטרומים המתאימים לאותות אלו. יהי $s(n)$ אות המתקבל מקונוולוציה בין $u(n)$ לבין $x(n)$.

$$s(n) = x(n)*u(n) \quad (3.3)$$

$$S(w) = X(w)U(w) \quad (3.4)$$

$$\log[S(w)] = \log[X(w)] + \log[U(w)] \quad (3.5)$$

$$\hat{h}_g(n) = F^{-1}\{\log[S(w)]\} = F^{-1}\{\log[X(w)]\} + F^{-1}\{\log[U(w)]\} \quad (3.6)$$

$$\hat{h}_g(n) = \hat{h}_x(n) + \hat{h}_u(n) \quad (3.7)$$

כלומר במישור הקפסטרום הקומפלקסי, הקפסטרום של $s(n)$ מורכב מסכום הקפטרומים השיכים ל $x(n)$ ו $u(n)$.

2. יהי $s(n)$ אות מינימום פאזה, אזי הקפסטרום הקומפלקסי שלו הוא חר צדדי.

$$\hat{h}(n) = 0, \quad n < 0 \quad (3.8)$$

3. הקפסטרום (הממשי) הוא אות זוגי כלומר $c(n) = c(-n)$.

4. עבור אות מינימום פאזה מתקיים הקשר הבא בין הקפסטרום הקומפלקסי לממשי.

$$c(n) = \begin{cases} \hat{h}(-n)/2 & n < 0 \\ \hat{h}(n) & n = 0 \\ \hat{h}(n)/2 & n > 0 \end{cases} \quad (3.9)$$

5. יהי $H(z)$ מסנן בעל קפסטרום $c(n)$, אזי הקפסטרום המתאים ל-
 $G(z) = 1/H(z)$ הוא $-c(n)$.

6. יהי $H(z)$ מסנן יציב ובעל קטבים בלבד.

$$H(z) = \frac{G}{1 + \sum_{i=1}^p a_i z^{-i}} \quad (3.10)$$

היציבות מבטיחה כי $H(z)$ יוא מינימום פאזה. יהי $\hat{h}(n)$ הקפסטרום הקומפלקסי המתאים ל $H(z)$ אזי מתקיים הקשר הבא [19]:

$$\hat{h}(n) = \begin{cases} \log(G) & n=0 \\ a_1 & n=1 \\ a_n + \sum_{k=1}^{n-1} (k/n) \hat{h}(k) a_{n-k} & 2 \leq n \leq p \\ \sum_{k=1}^p [(n-k)/n] \hat{h}(n-k) a_k & n > p \end{cases} \quad (3.11)$$

כלומר ניתן לחשב בצורה רקורסיבית את הקפסטרום הקומפלקסי, מתוך הפרמטרים $a_k, G, k=1, \dots, p$, המגדירים את $H(z)$.

7. יהי $\hat{h}(n)$ קפסטרום קומפלקסי המתאים למסנן יציב בעל קטבים בלבד $H(z)$, המקיים את (3.10) אזי מתקלים הקשר [19]:

$$a_n = \begin{cases} \hat{h}(1) & n=1 \\ \hat{h}(n) - \sum_{k=1}^{n-1} (k/n) \hat{h}(k) a_{n-k} & 2 \leq n \leq p \end{cases} \quad (3.12)$$

$$G = e^{\hat{h}(0)}$$

8. מכיון שקונוולוציה בתחום הזמן מקבילה לסכום במישור הקפסטרום, אזי פעולה המקבילה לסינון אות $s(n)$ על ידי מסנן $H(z)$, היא חיבור מקדמי הקפסטרום המתאימים ל $s(n)$ ו $H(z)$.

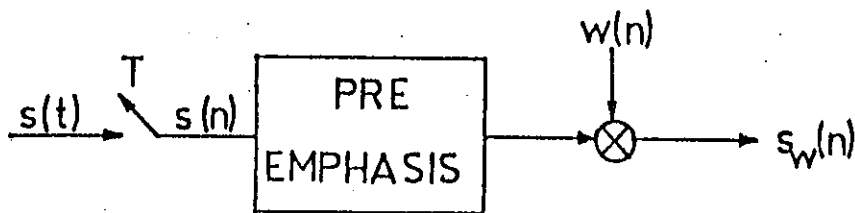
3.3 הצגת מודל האנליזה.

המודל המוצע מבוסס ביסודו על מודל ה-LPC, כאשר בנוסף לפרמטרים הקבועים במודל זה, נוספים איברים הנובעים מהשארית הקפסטרלית, על מנת לשפר את איכות וטבעיות הדיבור המשוחזר. קידוד זה גובה את מחירו בקצב הסיביות הנדרש לקוד תוספת זו, מדובר במקרה זה על העלאת קצב הסיביות מ-2400 סיביות לשניה, עד לקצב כפול. יש להדגיש, שהשיפור המתקבל באיכות הדיבור, אינו תוצאה ישירה של עלית הקצב בלבד. אם נקדיש את אוחס הסיביות, הניתנות למקדמי השארית הקפסטרלית, כדי להעלות את סדר מודל החיזוי הלינארי, השיפור באיכות הדיבור יהיה נמוך בהרבה, בהשוואה לשיפור המושג במודל החדש. הסיבה לכך היא העובדה שאי ההתאמה הספקטרלית אינה נובעת בעקרה מקטבים, ולכן אילה אינם יעילים ביצוגה.

בהמשך סעיף זה יוצג המודל המוצע צעד אחר צעד עד לקבלת המודל השלם.

3.3.1 עיבוד מקדים והגדרת קטע האנליזה (WINDOWING).

אות הדיבור הדגום, לפני היכנסו למערכת קידוד הדיבור, מועבר דרך מסנן קדם ולאחריו מוכפל בחלון ראה ציור (3.1).



ציור (3.1): מערכת הכניסה של מערכת האנליזה.

Fig. (3.1): Input stage of the analysis system.

המסנן המקדים (pre emphasis), מכיל אפס בודד. ב-LPC משתמשים בקדם עבוד זה, כדי "לישר" (Flatten) את הספקטרום של הדיבור לפני שערוך פרמטרי המודל. ספקטרום הדיבור נוטה לרדת עם עלית התדר. ככל שספקטרום זה פחות

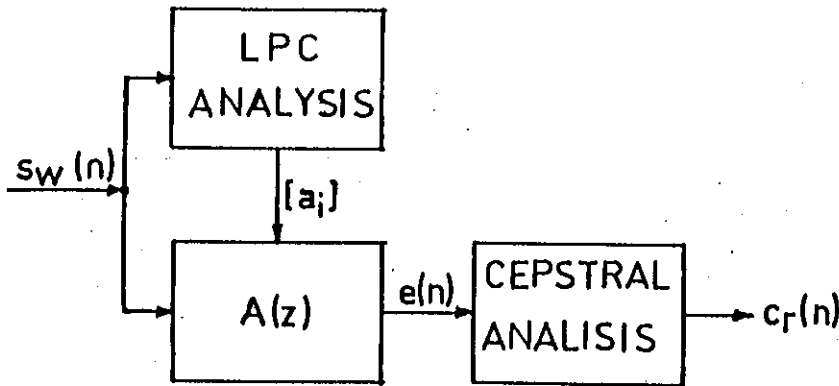
שטוח, כלומר ככל שהתחום הדינמי שלו גדול יותר, נוטה פונקציית האוטוקורולציה להיות רגישה יותר לחישובים באורך מלה סופי. עובדה הנובעת מכך שמטריצת האוטוקורולציה המתאימה, היא יותר "ill-conditioned", תופעה הגוררת בעיות אי יציבות במסנן המחושב. "ישור" הספקטרום מקטין בעיות אלו.

פונקציית החלון $w(n)$, תפקידה להגדיר את הקטע לאנליזה בצורה שתמנע בעיות אי יציבות בקצוות הקטע.

3.3.2 מציאת פרמטרי ה-LPC וחישוב השארית הקפסטרלית [15].

המסנן $H(z)$, המגדיר את מסנן החיזוי הלינארי, מחושב בשיטת Durbin, כפי שתוארה בפרק הקודם.

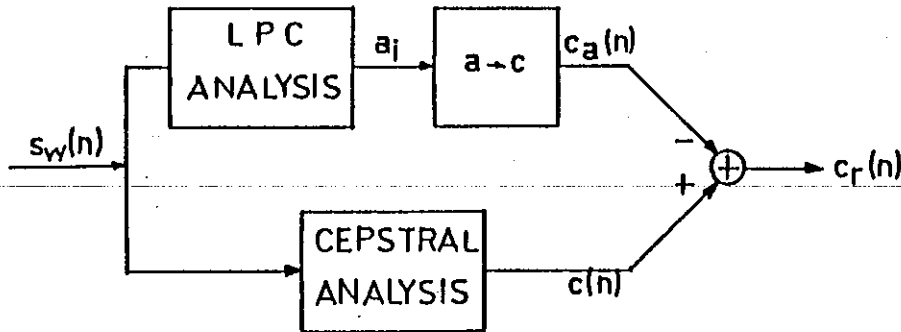
השיטה הישירה לחישוב השארית הקפסטרלית היא סינון אות הדיבור, $s(n)$, דרך המסנן ההפוך $A(z)$ (ראה פרק קודם), ומתוך אות השארית המתקבל במישור הזמן, חישוב הקפסטרם על ידי התמרות מתאימות, ראה ציור (3.2).



ציור (3.2): חישוב השארית הקפסטרלית בדרך הישירה.

Fig. (3.2): Cepstral residual computation - direct way.

דרך שניה לחישוב השארית הקפסטרלית תתואר להלן והיא שנבחרה לשימוש בעבודה זו, ראה ציור (3.3).



ציור (3.3): חישוב השארית הקפטרלית - דרך מקבילה.

Fig. (3.3): Cepstral residual computing - parallel way.

בגישה זו במקביל לחישוב מודל ה-LPC מחושב הקפטרום, $c(n)$, של אות הדיבור, על ידי שימוש בתכונות 4 ו-6 של הקפטרום (ראה סעיף 2. פרק זה), מחושב הקפטרום $c_a(n)$ המיצג את המסנן $H(z)$. על ידי חיטור $c_a(n)$ מתוך $c(n)$ מתקבלת השארית הקפטרלית $c_r(n)$.

$$c_r(n) = c(n) - c_a(n) \quad (3.13)$$

יתרונות השיטה השניה על הראשונה הן:

1. סיבוכיות החישוב קטנה יותר.
2. ההשהיה הנגרמת על ידי המערכת קטנה יותר, עובדה החשובה בקידוד דיבור בזמן אמיתי.

יש לשים לב שהשיטה השניה דורשת אמצעים המאפשרים את חישוב שני הענפים במקביל.

לאחר חישוב השארית הקפטרלית ניתן לגשת ולהפרידה לשני מרכיבים:

1. איברים המתאימים לאפסים הקימים באות זה.
2. איברים המתאימים למסננים קבועים או משתנים לאט, אשר קימים במערכת (כמו מסנן הכניסה שדרכו עובר הדיבור עד הגיעו למערכת הקידוד), ולא יוצגו כלאות על ידי מקדמי ה-LPC.

בעבודה [15] הוצג שכדי לקבל יצוג יעיל של האפסים, יש להקדים ולהפחית את השפעת אותם סינונים קבועים הקיימים בשארית הקפסטרלית.

3.3.3 חישוב הממוצע של השארית הקפסטרלית.

תכונה 8 מתכונות הקפסטרום שמניתי, מבהירה שסינון בתחום הזמן מתאים לחיבור במישור הקפסטרום.

הבעיה הניצבת בפנינו היא למצוא את מקדמי הקפסטרום המיצגים את אותו סינון קבוע, או משתנה לאט, שהדיבור עבר דרכו ואחר להפחיתם כדי לבטל את השפעתם על השארית.

עפ"י ההנחה שמסננים אלו משתנים לאט מאוד או אף קבועים (במקרים מסוימים), פרושה שהשפעתם תופיע בשארית הקפסטרלית כגורם קבוע הנתון ע"י הממוצע הזמני של השארית הקפסטרלית.

לחישוב ממוצע זה אפשריות שתי דרכים:

1. אם נניח שממוצע זה קבוע בזמן, אזי ניתן לדאוג שהוא ימצא במקלט.
2. אם ממוצע זה משתנה בזמן אזי יש לחשב ממוצע זה בצורה דינמית ולהעבירו כערוץ.

דרך אפשרית לחשב ממוצע זה בצורה דינמית היא על ידי שימוש ברקורסיה:

$$\langle c_r(n) \rangle_m = \alpha \langle c_r(n) \rangle_{m-1} + (1-\alpha)[c_r(n)] \quad (3.14)$$

כאשר m הוא מונה הקטעים, α מספר קבוע $0 < \alpha < 1$ המותאם לקצב ההשתנות של אותו ממוצע.

כמוכן שיש צורך להעביר ממוצע זה כערוץ. בגלל היותו משתנה לאט בזמן, קצב העידכון שלו יכול להיות נמוך, וניתן להעבירו בקטעי שקט, או לנצל את הסיביות המיועדות לקידוד תדר ה-PITCH בקטעים לא קוליים.

לאחר חישוב ממוצע זה ניתן להפחיתו ולחשב את השארית הקפסטרלית חסרת

ההטיה $\hat{c}_r(n)$ (unbiased).

$$\tilde{c}_r(n) = c_r(n) - \langle c_r(n) \rangle \quad (3.15)$$

3.3.4 חישוב האפסים.

לאחר שחושבה השארית הקפסטרלית הנוותרת $\tilde{c}_r(n)$, ניתן לגשת לחישוב האפסים.

3.3.4.1 הצגת הבעיה.

יהי $\tilde{c}_r(n)$ קפסטרום המתאים לאות בעל אפסים בלבד אשר פונקצית התמסורת שלו $G(z)$ מקימת:

$$G(z) = \frac{B(z)}{G_b} \quad (3.16)$$

$$B(z) = 1 + \sum_{i=1}^q b_i z^{-i} \quad (3.17)$$

מצא את $G_b, b_i, i=1, \dots, q$, המגדירים מסנן זה.

3.3.4.2 פתרון.

יהי $\tilde{c}_r(n)$ קפסטרום המתאים למסנן $G(z)$ כפי שמוגדר ב- (3.16), אזי $\tilde{c}_r(n)$, הינו קפסטרום המתאים למסנן $G(z)=1/G(z)$ בעל קטבים בלבד. מ $\tilde{c}_r(n)$ נחשב את $\hat{h}_g(n)$, הקפסטרום הקומפלקסי המתאים למסנן $\tilde{G}(z)$, בעל מינימום פאזה

$$\hat{h}_g(n) = \begin{cases} 0 & n < 0 \\ -\tilde{c}_r(n) & n = 0 \\ -2\tilde{c}_r(n) & n > 0 \end{cases} \quad (3.18)$$

מכאן קימות שתי שיטות לחישוב b_i :

1. שימוש בתכונת הקפסטרום (3.12), וחישוב b_i בצורה רקורסיבית.

2. לחשב מ- $\hat{h}_g(n)$ את האות הזמני המתאים לו, וממנו באלגוריתם ה-LPC למצוא את b_i כפרמטרים של מודל קטבים בלבד.

השיטה השניה עדיפה, למרות שבשניהם ההנחה היא ש- $\hat{c}_r(n)$ מתאים למודל אפסים בלבד, אזי חריגה מעטה מהנחה זו מקלקלת את התוצאה. בשיטה השניה ניתן לחשב אפסים גם אם $\hat{c}_r(n)$ מתאים למודל מסובך יותר וזאת במגבלות אלגוריתם ה-LPC.

3.3.5 חישוב השארית הקפסטרלית הנותרת.

לאחר חישוב מקדמי האפסים b_i , ניתן לחשב את השארית הקפסטרלית הנותרת שעדין אינה מיוצגת על ידי אותם מקדמים שכבר חושבו. בשימוש בתכונות 4 ו 6, של הקפסטרום (ראה סעיף 3.2), ניתן לחשב את $c_b(n)$, הקפסטרום המתאים למקדמים b_i , ומכאן נגדיר את השארית הנותרת:

$$\hat{c}_r(n) = \hat{c}_r(n) - c_b(n) \quad (3.19)$$

לאחר חישוב השארית הנותרת ניתן, אם רוצים, לקדר חלקים ממנה. לפני קידודה מוצע להכפיל את $\hat{c}_r(n)$ בחלון התלוי באורך מחזור ה-PITCH כדי שהאיברים הנבחרים לקידוד, לא יכילו מידע הקשור לאות העירור ול-PITCH, מכיון שמידע זה מועבר במקביל.

לאחר קידוד אותו חלק מהשארית הנותרת, שנסמנו ב- $c_t(n)$, נשארת שארית $c_e(n)$, המגדירה סופית את אי ההתאמה הספקטרלית בין האות המשוחזר על ידי המערכת המשופרת, לבין אות הדיבור האמיתי:

$$c_e(n) = \hat{c}_r(n) - c_t(n) \quad (3.20)$$

כלומר $c_e(n)$ נותנת מדד לאי ההתאמה בין אות הכניסה לאות המשוחזר.

3.3.6 החלטת v/uv וחישוב תדר ה- PITCH.

לא נרחיב את הדיבור על בעיה זו. נציין רק שמכיון שתוך כדי חישוב השארית הקפטרלית מחושב הקפטרום של אות הדיבור, ניתן לבצע את מצוי ה- PITCH, בתחום הקפטרום, כפי שתואר בפרק השני.

3.3.7 המודל השלם.

במודל השלם, כפי שהוצע ב- [15], נבדקו שתי אפשרויות המכונות PCR ו

PZCR:

1. PZCR - הוא מודל בו מקודדים קטבים, אפסים ומקדמי השארית הקפטרלית הנותרת (חלקם).

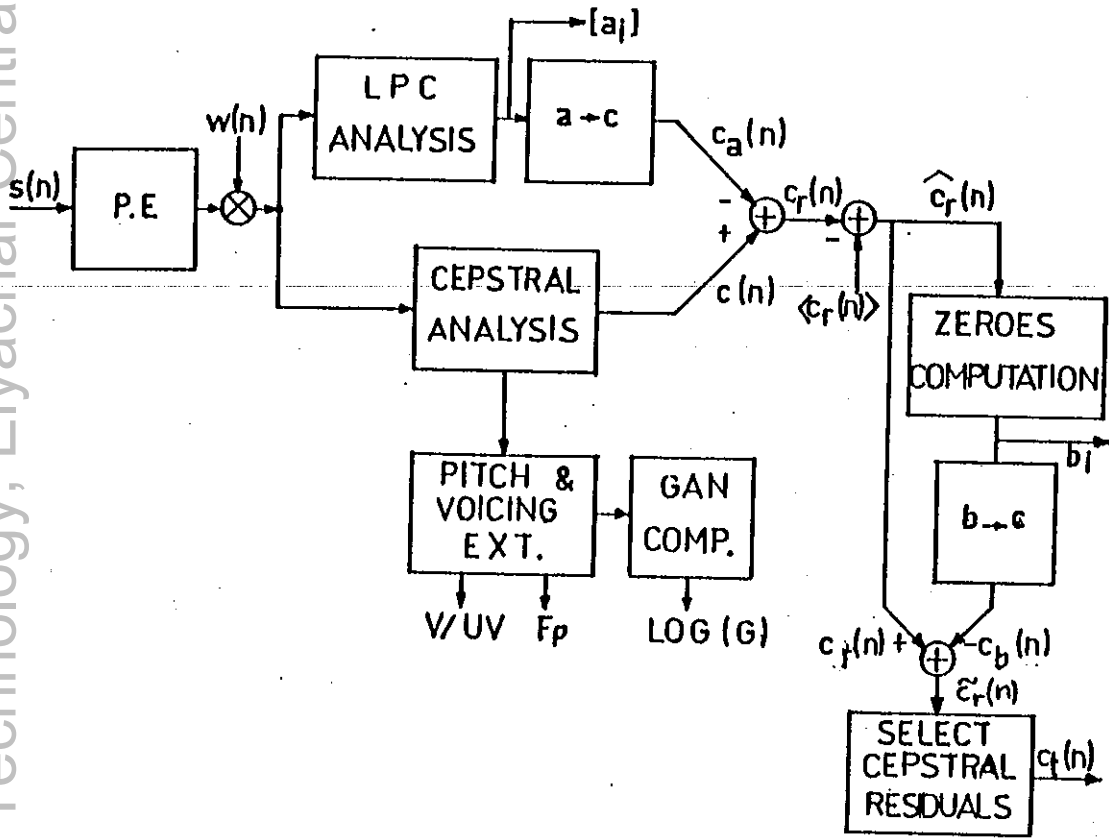
2. PCR - מודל המותר על קידוד האפסים, ומפנה את הסיביות שהתפנו,

לקידוד מספר גדול יותר של מקדמי השארית הקפטרלית הנותרת.

בשני המודלים שנמנו לעיל ממשיכים לקוד את $\langle c_x(n) \rangle$, הממוצע של השארית הקפטרלית.

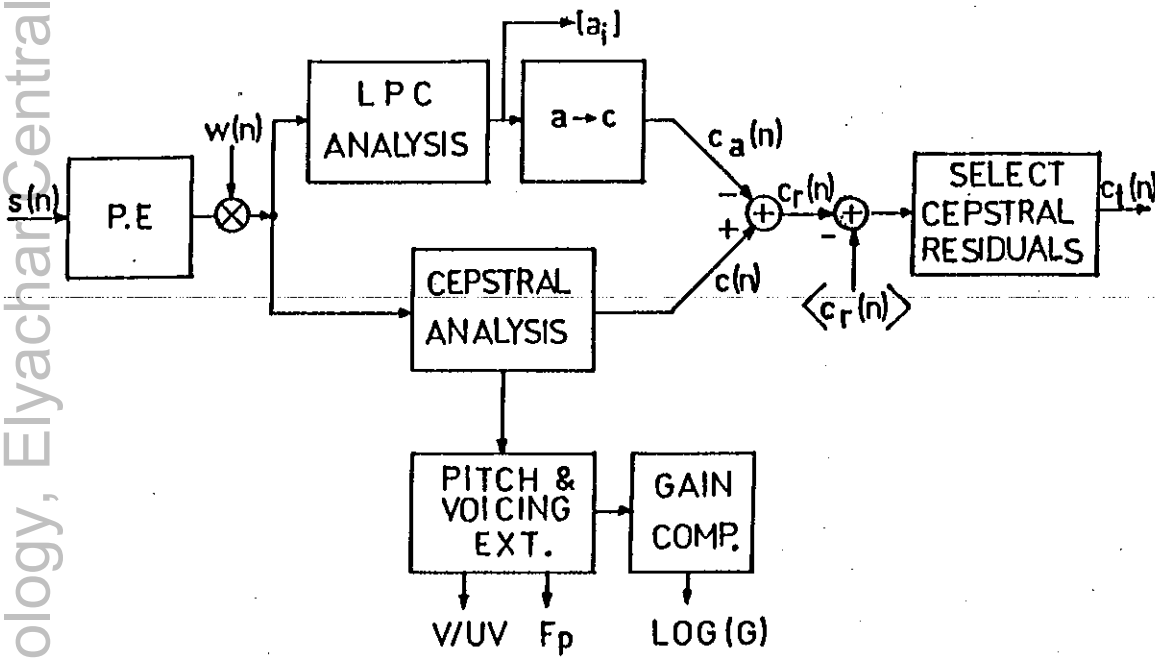
ציורים (3.4) ו (3.5) מתארים את מודל האנליזה השלם עבור שני

האפשרויות שפורטו לעיל.



ציור (3.4): אנליזה PZCR - קודד קטבים אפסים ושארית ספקטרלית.

Fig. (3.4): PZCR analysis - poles, zeros and cepstral residual coding.



צילור (3.5): אנליזת PCR - קודוד קטבנים ושאריות טפקטרליות.

Fig. (3.5): PCR analysis - poles and cepstral residual coding.

3.4 מערכת הסינטזה.

מערכת הסינטזה במודל המוצע, דומה למערכת הסינטזה הקלסית במקרה LPC, אלא שנוסף לה מסנן הבנוי מאותם תוספות שנוספו על גבי המודל הקלסי. אלה הם השארית הקפסטרלית הממוצעת והנותרת. מסנן זה תפקידו לעצב את ספקטרום אות הערור הקלסי לפני הגיעו למסנן $H(z)$, מסנן מודל ה-LPC. במידה ומקודדים גם האפסים, יש לסנן את האות המשוחזר גם עם המסנן הבנוי מאפסים אילו. דאה ציורים (3.6) ו (3.7), המתארים את שתי מערכות הסינטזה הנובעות ממערכות האנליזה המתאימות.

חלקי המערכת, המופיעים בציורים (3.6) ו (3.7) הוסברו, אך נותר להסביר כיצד מתקבל המסנן המעצב את ספקטרום אות העירור, מתוך מקדמי הקפסטרום המקודדים. בכדי לקבל מסנן זה יש לבצע את הצעדים הבאים:

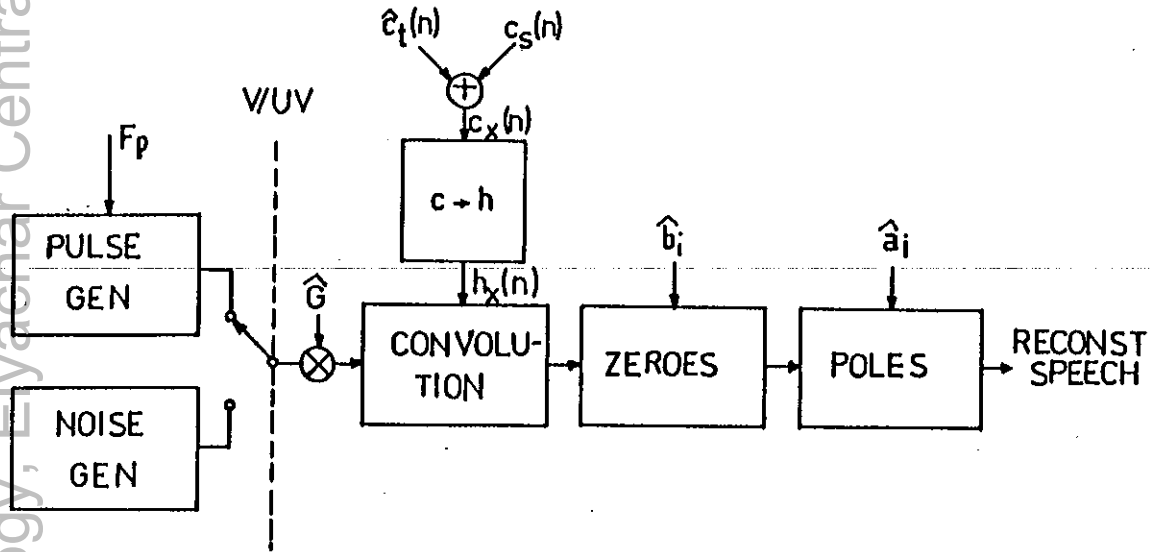
1. מסכמים את $c_t(n)$ ו $\langle c_r(n) \rangle$ כדי לקבל $c_x(n)$

$$c_x(n) = c_t(n) + \langle c_r(n) \rangle \quad (3.21)$$

2. מ $c_x(n)$ מחשבים את הקפסטרום הקומפלקסי בעל מינימום פאזה $\hat{h}_x(n)$ המתאים לו, וממנו אות זמני $h(n)$.

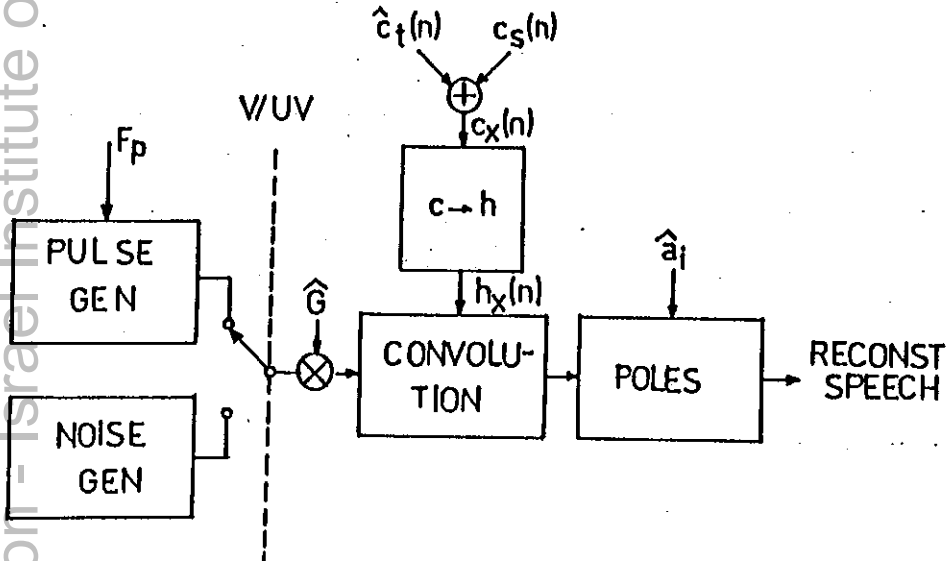
$h(n)$ הוא התגובה להלם יחידה של המסנן המחליק את אות הערור לפני היכנסו למערכת ה-LPC.

בסעיף הבא נגדיר מדד איכות מקובל לאי ההתאמה הספקטרלית, כדי לאפשר השוואת המודל המוצע כאן, למודל החזוי הלינארי המקובל, וכן בכדי להשוות את האפשרויות השונות שהוצעו: PZCR ו PCR.



ציון (3.6): סינתזת PZCR - אפסים קטנים ושארית קפטטלית.

Fig. (3.6): PZCR synthesis - poles zeros and cepstral residual.



ציון (3.7): סינתזת PCR - אפסים קטנים ושארית קפטטלית.

Fig. (3.7): PCR synthesis - poles and cepstral residual.

3.5 מדד איכות E_c

מטרת מערכת קידוד דיבור בקצב נמוך, היא ליצג בצורה יעילה את העוטפת הספקטרלית של אות הדיבור, אשר היא, כפי שהוכח בעבודות קודמות, האחראית העיקרית לאיכות הדיבור המשוחזר.

בעבודה [15] נמצא שהאנרגיה של השארית הקפסטרלית הנותרת $c_e(n)$, כפי שהוגדרה בסעיפים קודמים, מהווה מדד יעיל לאי ההתאמה הספקטרלית.

המדד E_c יוגדר להלן.

$$E_c = (\epsilon_c^2)^{1/2} E_0 \quad (3.22)$$

כאשר ϵ_c^2 הוא האנרגיה של $c_e(n)$ על פני N_c האיברים הראשונים

$$\epsilon_c^2 = \sum_{n=-N_c}^{N_c} c_e^2(n) \quad (3.23)$$

ו $E_0 = 20/\log(10)$ הוא קבוע ההופך את יחידות E_c ליחידות ב-dB. N_c נבחר כך ש E_c לא יושפע ממקדמי הקפסטרם המיצגים את אות הערוך, כך ש- E_c נותן מדד לאי ההתאמה של העוטפת הספקטרלית, ולא של הספקטרם בכללותו.

אם נשתמש במשפט פרסוול מתקבל.

$$\epsilon_c^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} [\log |E(w)|]^2 dw \quad (3.24)$$

$$E_c = 20/[\log(10)] \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} [\log |E(w)|]^2 dw \right\}^{1/2} \quad (3.25)$$

כלומר E_c נותן מדד לאי ההתאמה ב db בין \log הספקטרם של הדיבור האמיתי לבין \log הספקטרם של הדיבור המשוחזר.

באותה עבודה [15] נמצא שעבור $E_c=1$ dB ההבדל בין האות המשוחזר לאות משוחזר עם כל מקדמי הקפסטרום אינו ניתן להבחנה במיבחני שמיעה טוביקטיבים.

3.6 סימולציה של מודל קידוד השארית

מערכת האנליזה-סינטזה, כפי שתוארה בסעיפים הקודמים, נבדקה ב-[15] גם בסמולציה על מחשב. ניבדקו ביצועיה ביחס לפרמטרים אחדים. בדיקות האיכות התחלקו לשניים:

1. שימוש במדד האיכות שהוגדר בסעיף הקודם.

2. בדיקות שמיעה סלקטיביות.

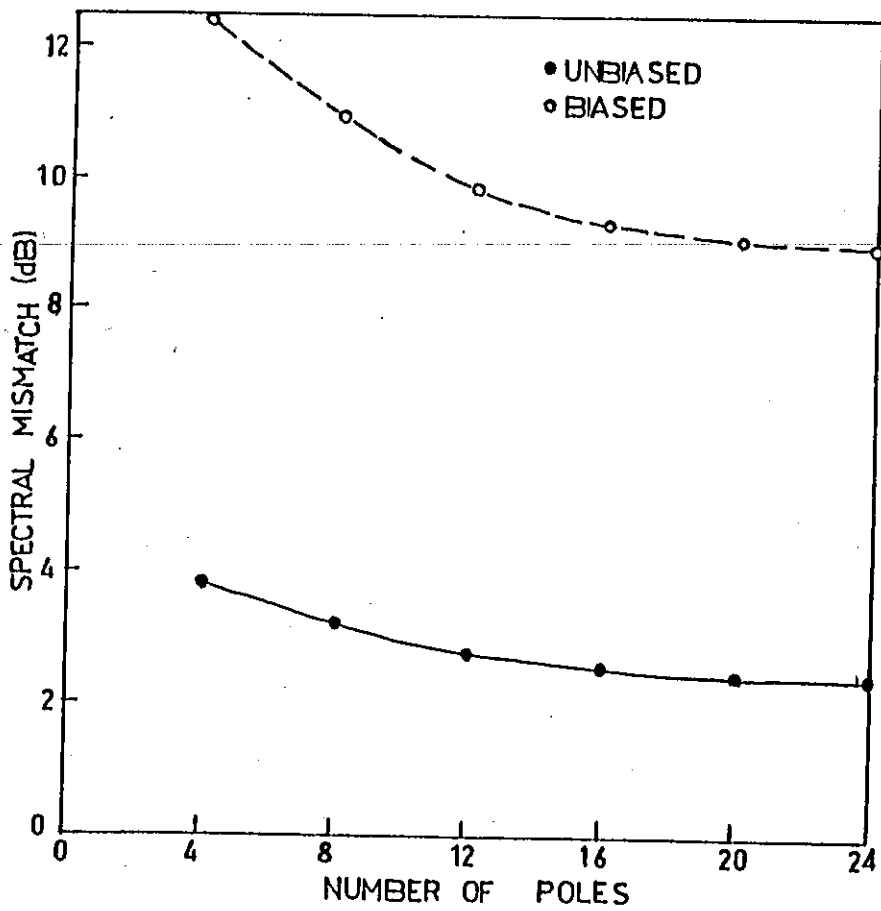
הבדיקה נעשתה על מגוון של משפטים.

ב [15,16] מתוארת סימולציה של שני המקדדים PCR ו PZCR. תוצאות אותן סימולציות יתוארו להלן.

3.6.1 תוצאות הסימולציה.

בתוצאות הסימולציה נתרכז בביצועי המערכת, בגילגוליה השונים, כפי שמתאר אותם מדד האיכות E_c . יש לזכור לכל אורך הדרך שהמטרה היא להגיע ל- $E_c=1\text{db}$. מטרה המגדירה איכות דיבור מספקת (בדרך כלל).

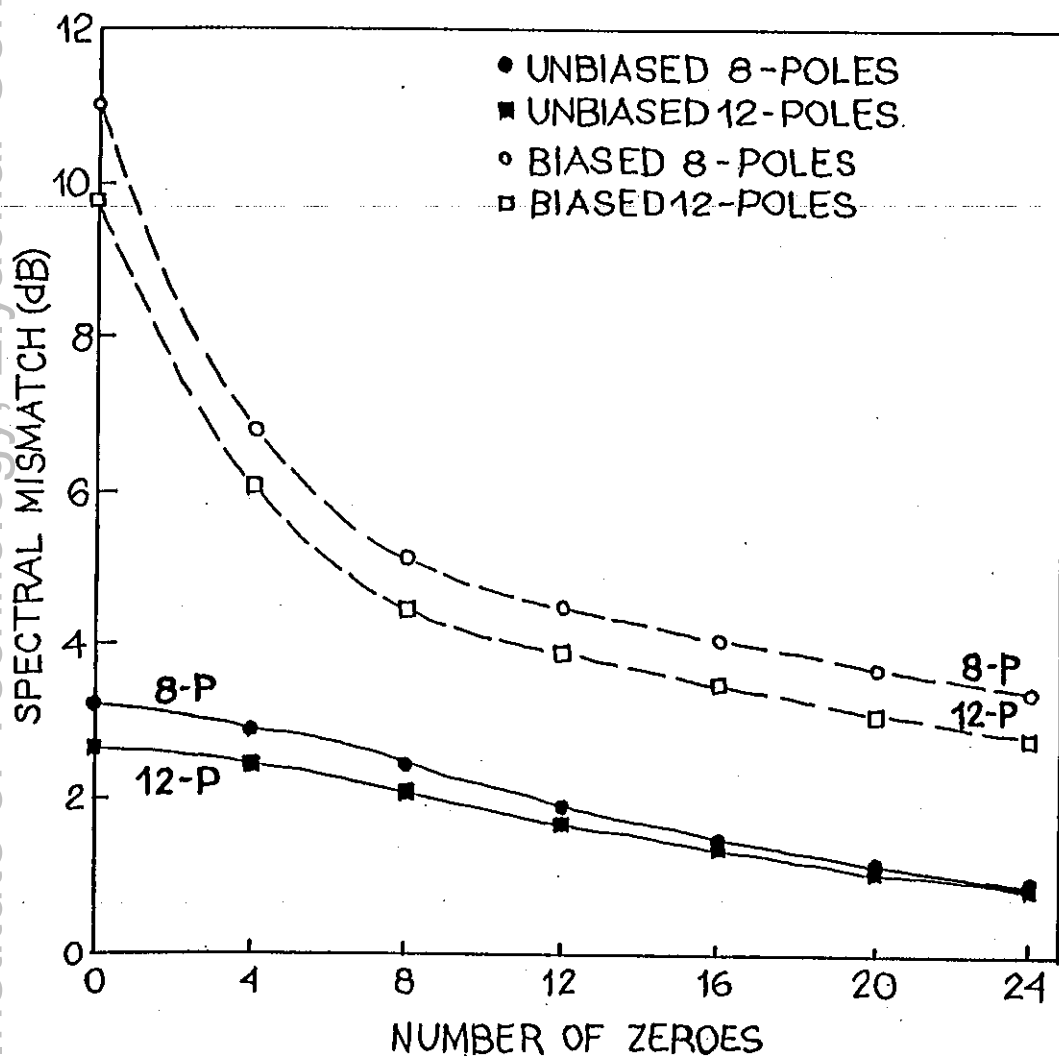
ציור (3.8) מתאר את תלות E_c ב- p , סדר מודל ה-LPC. מציור זה ברור שהגדלת סדר המודל מעל $p=8$ משפרת במעט, אך הגדלת סדר המודל מעל $p=12$ אינה משפיעה כמעט על מדד האיכות. בציור זה גם מתואר ההבדל בין מערכת עם הורדת ממוצע השארית הקפסטרלית (unbiased), לבין זו בעלת הממוצע (biased), והשיפור הדרסטי בולט בראשונה.



ציור (3.8): מדד אי ההתאמה הספקטורלית בתלות במספר הקטבים.

Fig. (3.8): Spectral mismatch versus LPC model's order.

ציור (3.9) מבהיר את תרומתם של האפסים על ביצועי המערכת (כפי שמתוארים על ידי E_c). מציור זה ברורה תרומתם של אלו אולם מתברר שהגדלת מספרם מעל $q=8$ משפיעה אך במעט על E_c . כן מתברר שההבדלים בין $p=8$ לבין $p=12$ הולכים ומתקזזים עם הגדלת מספר האפסים.



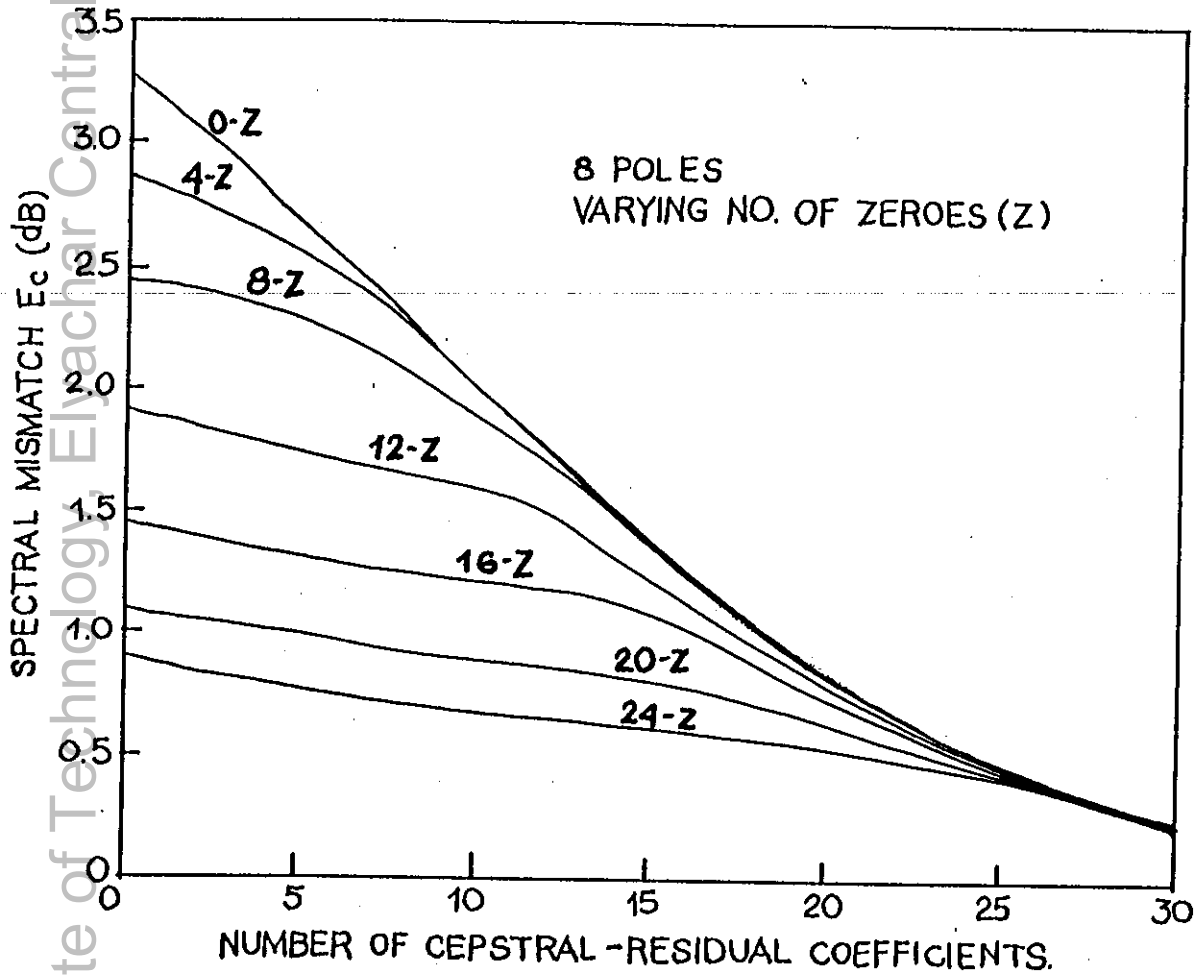
ציור (3.9): השפעת האפסים על מדר אי ההתאמה הספקטרלית.

Fig. (3.9): The influence of the zeros on the spectral mismatch.

ציור (3.10) מתאר, עבור $p=8$, השוואה בין מספר האפסים למספר מקדמי

השארית הקפסטרלית. מתברר שעבור מודל המקדד 25 מקדמי שארית קפסטרלית

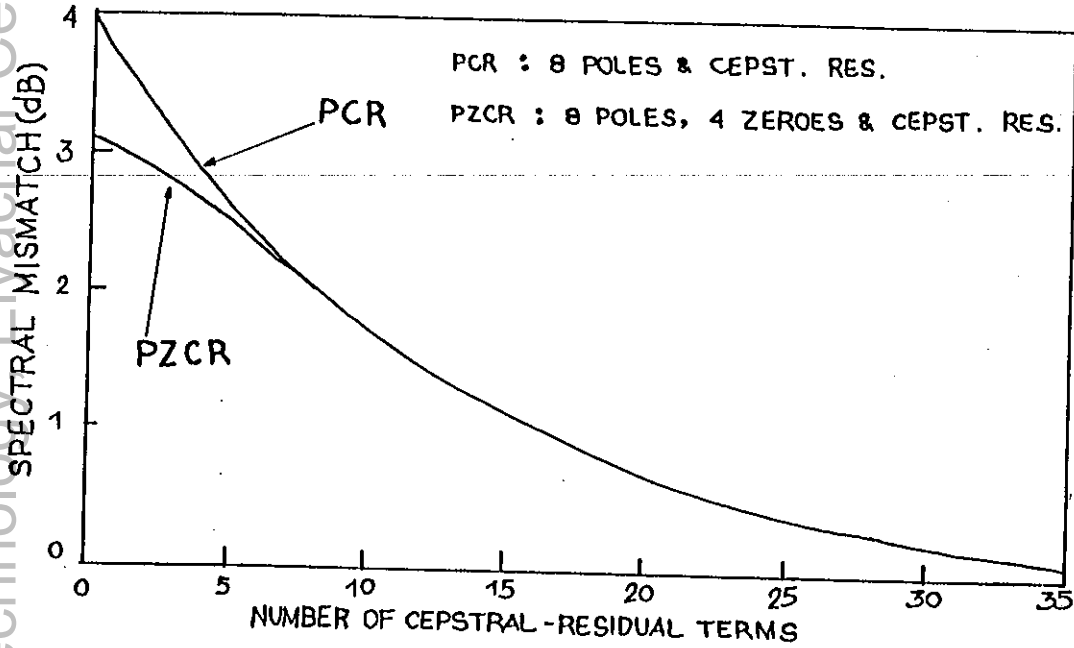
ומעלה, קידוד האפסים מיותר, לפחות ביחס למוד האיכות E_c .



ציור (3.10): השוואה בין תרומת האפסים לתרומת מקדמי שארית הקפטטלית על ההתאמה הספקטלית.

Fig. (3.10): Comparison between the contribution of zeros and cepstral coefficients to the spectral matching.

ציור (3.11) משווה סופית בין שני המודלים PCR ו PZCR. ממנו מתברר סופית שאפשר לוותר על קידוד האפסים, אם נזכור שניתן להפנות את המשאבים (סיביות) שהופנו לקידוד האפסים, לקידוד מקדמי שארית קפטטלית נוספים, הרי ברורה עדיפות מודל PCR.



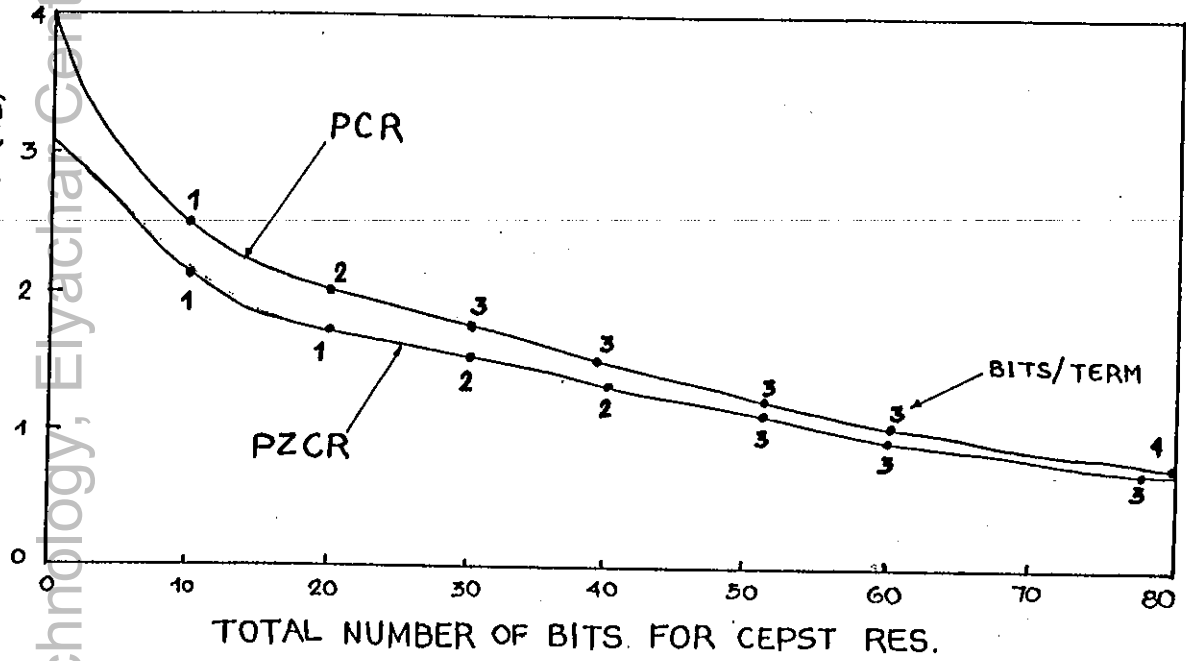
צילור (3.11): השוואה בין מקדד PZCR לבין מקדד PCR.

Fig. (3.11): Comparison between PZCR model and PCR model.

3.6.2 מסקנות.

1. העלאת סדר מודל הקטבים k אינה מקבילה באיכותה לקידוד אפסים ו/או מקדמי שארית קפסטרלית. כלומר ביצועי המודל החדש עולים על מודל LPC קלסי, גם אם הם מקודרים במספר סיביות שווה.
2. מודל PCR עדיף על מודל PZCR, כאשר הבסיס להשוואה הוא מדד אי ההתאמה הספקטרלית E_c . אולם במבחני שמיעה סלקטיבים מתברר שעבור דוברים מסוימים ובחלק מהקטעים, מודל PZCR עדיף בביצועיו. אלה הם בדרך כלל קטעי דיבור בעלי אפסים מודגשים.

צילור (3.12) מתאר את חלוקת הסיביות עבור מערכת PCR. עבור קצב של 4800 סיביות לשניה מתקבלים קרוב מאוד למערכת שביצועיה הממוצעים מתאימים ל $E_c = 1dB$.



צילור (3.12): השוואה בין PZCR לבין PCR ביחס לקצב הסיביות הדרוש להשיג מדרג אי התאמה של 1db.

Fig.(3.12): comparison between PZCR and PCR, the bit rate needed to get 1dB spectral mismatch.

פרק רביעי: חיזוי לינארי של אותות דיבור
הטבולים ברעש לבן

4.1 מבוא

LPC הינה שיטה יעילה לקידוד דיבור בקצב נמוך. יעילותה נובעת מעצם הנחת מודל הקטבים בלבד עבור אות הדיבור והעובדה שאות זה קרוב בתכונותיו למודל זה.

כאשר מתווסף לאות הדיבור רעש, איכותו של אות הדיבור המשוחזר ממקודד LPC יורדת בצורה חדה מאוד כתלות ביחס אות לרעש (SNR). ירידה זו באיכות הדיבור הינה תלולה יותר בהשוואה לזו המתקבלת בשיטות קידוד האות. הסיבה לירידה החדה באיכות הדיבור המשוחזר היא העובדה שאות הדיבור הרועש אינו מקיים את הנחת מודל הקטבים בלבד [13].

קימות שיטות רבות להקטנת השפעתו של הרעש על ביצועיו של מקודד ה-LPC. חלקן מטפלות בניקוי אות הדיבור כעיבוד מוקדם לאלגוריתם ה-LPC, בעוד אחרות מבוססות על מודיפיקציה של האלגוריתם כדי לטפל ברעש.

בפרק הנוכחי תובא סקירה על שיטות אלה ותכונותיהם, אולם תקדים אותה הצגת מודל האות הרועש ותכונותיו.

4.2 אות הדיבור הרועש.

4.2.1 הצגת מודל האות הרועש [13].

יהי $s(n)$ אות דיבור המקיים (בקרווב) את הנחת מודל הקטבים בלבד כלומר

$$S(z) = \frac{G}{A(z)} \quad (4.1)$$

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (4.2)$$

יהי $n(n)$ רעש לבן בעל צפיפות ספקטרום הספק N_0 . נגדיר את $x(n)$ אות הדיבור הרועש כדלקמן

$$x(n) = s(n) + n(n) \quad (4.3)$$

יהי $P_x(z)$ ספקטרום ההספק של $x(n)$ אזי

$$P_x(z) = \frac{G^2}{A(z)A(z^{-1})} + N_0 \quad (4.4)$$

או

$$P_x(z) = \frac{G^2 + N_0 A(z)A(z^{-1})}{A(z)A(z^{-1})} \quad (4.5)$$

נגדיר את $B(z)$ כדלקמן

$$G_b^2 B(z)B(z^{-1}) = G^2 + N_0 A(z)A(z^{-1}) \quad (4.6)$$

אזי

$$P_x(z) = \frac{G_b^2 B(z)B(z^{-1})}{A(z)A(z^{-1})} \quad (4.7)$$

כלומר אות הדיבור הרועש הוא אות המקיים מודל של קטבים ואפסים, כאשר הקטבים שלו הם אותם הקטבים כמו של $s(n)$ ואילו האפסים הם תרומתו של הרעש שמתבטא בפולינום $B(z)$.

למרות האמור לעיל ניתן עדיין להציג את $x(n)$ כאות המיוצג על ידי מודל קטבים בלבד.

נגדיר

$$C(z) = \frac{A(z)}{B(z)} = 1 + \sum_i c_i z^{-i} \quad (4.8)$$

אזי

$$P_x(z) = \frac{G_b^2}{C(z)C(z^{-1})} \quad (4.9)$$

יש לשים לב שלמרות $A(z)$ הוא פולינום מסדר סופי וכנ"ל $B(z)$, $C(z)$ יהיה פולינום מסדר גדול ובדרך כלל מסדר אינסופי.

מכיוון שלפונקציות האוטוקורולציה תפקיד חשוב באלגוריתם ה-LPC, תוכא גם השפעתו של הרעש על פונקציות האוטוקורולציה המתקבלת [13].

יהיו $R_x(k)$ ו $R_s(k)$ פונקציות האוטוקורולציה של $x(n)$ ו $s(n)$ בהתאמה בהנחה ש $n(n)$ הוא רעש לבן, מתקיים הקשר

$$R_x(k) = \begin{cases} R_s(0) + N_0 & k=0 \\ R_s(k) & k \neq 0 \end{cases} \quad (4.10)$$

כלומר רעש לבן משפיע רק על מקדם האוטוקורולציה הראשון. רעש פיזיקלי אינו לבן ממש, לכן הוא ישפיע על מקדמים נוספים, בנוסף פונקציות האוטוקורולציה מחושבת על פני קטע זמן סופי וספקטרום של קטע כזה אינו לבן.

4.2.2 השפעתו של הרעש על מודל ה-LPC

בניתוח המודל שהתקבל עבור אות רועש, ניתן להראות [13] את השפעתם של האפסים הנ"ל ובדיעבד השפעתו של הרעש, על המסנן המתקבל באלגוריתם ה-LPC. השפעה זו תלויה ביחס σ^2/N_0 , יחס התלוי ישירות ביחס אות לרעש.

1. עבור יחס גדול מאוד, האפסים מתרכזים בראשית ואינם משפיעים על הקטבים של המודל המשוער.

2. ככל ש- יחס זה הולך וקטן, האפסים נעים מהראשית לעבר הקטבים ומשפיעים על מיקומם ורוחב הסרט של הקטבים (פורמנטים).

3. עבור יחס קטן מאוד מיקום האפסים מתאחד עם מיקום הקטבים, גורם לכיטול הקטבים ולקבלת מודל משוער בעל ספקטרום שטוח.

לאור השפעתו המפריעה של הרעש על ביצועי מקודד LPC, יש צורך לטפל ברעש כבר ביחסי אות לרעש גבוהים יחסית, שעבורם, בשיטות קידוד האות ניתן עדיין להתעלם מהרעש. מכאן ברור שיש צורך לבדוק את השיטות לניקוי הרעש ביסודיות, מכיוון שלא להישאר על הרעש גדולה על איכות הדיבור המשוחזר.

4.3 שיטות להקטנת השפעתו של הרעש על ביצועי מקדור LPC.

השיטות לטיפול ברעש מתבססות על מודל האות הרועש. חלקן משתמשות בשיטות ניקוי מקובלות, כעיבוד מוקדם, בעוד שאחרות מנקות תוך כדי ממוש אלגוריתם ה- LPC עצמו.

בהמשך הפרק תובא סקירה של השיטות הנ"ל, יתרונותיהן וחסרונותיהן.
השיטות שיסקרו:

1. הגדלת סדר המודל [13].
2. הפחתת ספקטרלית (SPECTRAL SUBTRACTION) [20-23].
3. הפחתת אוטוקוולציה [24].
4. משוואות Yule Walker מסדר גבוה [25].
5. שיערוך של מקדמי ההחזרה החלקית [25].
6. שיערוך קטבים ואפסים [26].

4.3.1 הגדלת סדר המודל. [13].

שיטה זו מתבססת על (4.8), אשר מכהירה שניתן להתייחס אל $x(n)$, האות הרועש, כאל אות המיצג מודל קטבים בלבד. אי לכך ניתן לשערך את העוטפת הספקטרלית של $x(n)$, על ידי אלגוריתם LPC. אולם כדי לשערך את העוטפת הספקטרלית הנ"ל יש צורך לבחור מודל LPC מסדר גבוה יותר מזה המתאים לאות הנקי, $s(n)$.

יתרונה המידי של שיטה זו היא פשטותה, בעצם אין צורך לשנות את המערכת המתאימה לדיבור נקי.

חסרונות השיטה:

1. סדר המודל המתאים ל $x(n)$ עלול להיות גדול מאוד, דבר היוצר בעיות דיוק ויציבות באלגוריתם ה- LPC.
2. גם אם סדר המודל הנבחר מספיק גדול, הרי ספקטרום האות המשוחרר אינו יכול להיות טוב יותר מזה של האות הרועש.

4.3.2 הפחתה ספקטרלית (SPECTRAL SUBTRACTION) [20-23].

הפחתה ספקטרום הספק, להלן SPS, הינה שיטה מקובלת לניקוי דיבור מרעש. שיטה זו מתבססת על העובדה שספקטרום ההספק של האות הרועש, $P_x(w)$, שווה לסכומם של $P_s(w)$ ו- $N(w)$ ספקטרומי ההספק של האות הנקי והרעש בהתאמה. כלומר

$$P_x(w) = P_s(w) + N(w) \quad (4.11)$$

שלבי אלגוריתם ה- SPS:

1. מצא את $\hat{N}(w)$, המשערך של ספקטרום ההספק של הרעש $N(w)$, על ידי מיצוע ספקטרומי ההספק של קטעי רעש בלבד.
2. חשב את $\hat{P}_s(w)$, המשערך של ספקטרום ההספק של אות הדיבור הנקי, לפי הנוסחה להלן

$$\hat{P}_s(w) = \begin{cases} P_x(w) - \hat{N}(w) & P_x(w) - \hat{N}(w) > K_0 \\ K_0 & \text{אחרת} \end{cases} \quad (4.12)$$

- כאשר K_0 מספר ממשי לא שלילי, המוגדר באופן ניסויי ותפקידו להבטיח חיוביות של הספקטרום המשוערך.
3. בשלב זה קימות שתי אפשרויות לקבלת אות הדיבור המנוקה: מוסיפים לספקטרום שהתקבל את הפאזה של האות הרועש ומחשבים את האות הזמני, או כאשר הניקוי נעשה כעיבוד מקדים לאלגוריתם ה- LPC, מחשבים את פונקציות האוטוקורולציה ישירות מספקטרום ההספק.

כאשר מממשים אלגוריתם זה מתעוררות מספר בעיות:

1. $\hat{N}(w)$, משערך ספקטרום ההספק של הרעש, אינו שווה בהכרח לספקטרום ההספק של פונקציה המדגם הרלוונטית של הרעש, לכן יש גם תדרים בהם $\hat{N}(w)$ גדול מדי ואחרים בהם הוא קטן מדי. עבור המקרים בהם $\hat{N}(w)$ בהם גדול מדי שמוש ב- $K_0 > 0$ המבטיח את חיוביותו של ספקטרום ההספק. באזורים בהם $\hat{N}(w)$ קטן מדי עלולים להווצר אזורי ספקטרום צרים

ובולטים. אזורים אלו מפולגים אקראית על כל תחום התדר ונעים אקראית מקטע אנליזה אחד למשנהו. קטעי תדר אילו נשמעים כצלילי רקע באות המשוחזר. הפתרון לבעיה זו הוא למצע את הדיבור על פני מספר קטעים, כך שספקטרום הרעש בכל קטע אנליזה גם הוא יהיה ממוצע ולכן קרוב יותר ל- $\hat{N}(w)$ כך שכמות השיאים הספקטראלים הנ"ל תלך ותקטן. נהוג למצע בין שלושה לחמישה קטעים. יש לשים לב שמיצוע זה גורם בד"כ למריחת ספקטרום הדיבור ועלול ליצור עיוות באות הדיבור המשוחזר.

2. K_0 נקבע באופן נסיוני ובחירתו קשה. מצד אחד ככל שהוא קטן הפחתת הרעש יעילה יותר, מצד שני ככל שהוא גדול יותר התחום הדינמי של הספקטרום קטן יותר עובדה התורמת ליציבות מסנן ה-LPC. בנוסף K_0 גדול מקטין את השיאים האקראיים שנוצרים ואשר נשמעים כצלילי רקע.

שיטת ה-SPS הינה שיטת ניקוי דיבור השיכת לקבוצה של שיטות אשר מתוארות בצורה כללית בנוסחאה להלן [23]:

$$\hat{S}(w) = \hat{P}_g(w)^{1/2} \exp\{\theta_x(w)\} \quad (4.13)a$$

כאשר

$$|\hat{S}(w)| = \begin{cases} [P_x(w)^{a/2} - b(\hat{N}(w))^{a/2}]^{1/a} & \text{אחרת} \\ K_0 & [P_x(w)^{a/2} - b(\hat{N}(w))^{a/2}]^{1/a} > K_0 \end{cases} \quad (4.13)b$$

כאשר:

a, b קבועים ממשיים.

$\theta_x(w)$ הפאזה של האות הרועש.

$\hat{S}(w)$ משערך הספקטרום הקומפלקסי המתאים ל- $s(n)$.

אם נציב $a=2$ ו- $b=1$ נקבל את שיטת ה-SPS, אשר ידועה בספרות גם כ-

POWER SPECTRAL SUBTRACTION, מכיוון שמפחיתים בה את הספקים.

4.3.3 הפחתת אוטוקורולציה [24].

הפחתת אוטוקורולציה הינה שיטת ניקוי דיבור מרעש המיועדת לעבודה עם אלגוריתם ה-LPC.

פונקציית האוטוקורולציה היא התמרת פוריה של ספקטרום ההספק. התמרת זו היא לינארית ולכן מ-(4.11)

$$R_x(k) = R_g(k) + R_n(k) \quad (4.14)$$

כאשר $R_x(k)$, $R_g(k)$ ו $R_n(k)$ הן פונקציות האוטוקורולציה של $x(n)$, $g(n)$ ו $n(n)$ בהתאמה. (4.14) מאפשרת שערך ישיר של $R_g(k)$. מוצאים משערך לפונקציית האוטוקורולציה של הרעש, $R_n(k)$, על ידי מיצוע של קטעי רעש בלבד ומחשבים את $\hat{R}_g(k)$ על ידי:

$$\hat{R}_g(k) = R_x(k) - \hat{R}_n(k) \quad (4.15)$$

הבעיה בשיטה זו היא שלא ידועה דרך ישירה להבטיח ש- $\hat{R}_g(k)$ תהיה מוגדרת חיובית, אי לכך יש צורך לעבור למישור התדר בכדי לבדוק תכונה זו ולשנות הפתרון אם הספקטרום אינו חיובי. בגלל בעיה זו ממוש שיטת האוטוקורולציה מתבצע בדרך שונה הניתנת להלן:

1. חשב $R_x(k)$, $k=0, \dots, p$.
2. חשב $\hat{R}_n(k)$, $k=0, \dots, p$, על ידי מיצוע פונקציית האוטוקורולציה בקטעי רעש בלבד.
3. חשב את $P_x(w)$ ו $\hat{P}_n(w)$ מ- $R_x(k)$ ו $\hat{R}_n(k)$ בשימוש בהתמרת פוריה.
4. חשב את $\hat{P}_g(w)$ על ידי פעולה מקבילה ל- (4.12) באלגוריתם SPS.
5. חשב את $\hat{R}_g(k)$, $k=0, \dots, p$, מתוך $\hat{P}_g(w)$.
6. המשך באלגוריתם ה-LPC עם $p+1$ המקדמים של $\hat{R}_g(k)$.

למראית עין שיטה זו מקבילה לשיטת ה-SPS ורק מסבכת אותה. אולם שיטה זו הרבה יותר חסכונית בחישובים, מכיוון שכל החישובים מתבצעים על מספר קטן של רכיבי תדר ($p+1$ מקדמים), וזולוציה נמוכה זו מאפשרת את חישוב מקדמי

האוטוקורלציה המשתתפים בחישוב מודל ה-LPC, מספר זה קטן בסדר גודל מזה הדרוש לחשב את כל רכיבי התדר כפי שמתבצע בשיטת ה-SPS.

4.3.4 משוואות Yule Walker מסדר גבוה [25].

כפיתוח אלגוריתם ה-LPC מגיעים לפתרון מערכת המשוואות הבאה:

$$\begin{bmatrix} R_x(0) & ,R_x(1) & , \dots ,R_x(p-1) \\ R_x(1) & & R_x(p-2) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ R_x(p-1) & ,R_x(p-2) & , \dots ,R_x(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} = - \begin{bmatrix} R_x(1) \\ R_x(2) \\ \cdot \\ \cdot \\ \cdot \\ R_x(p) \end{bmatrix} \quad (4.16)$$

מערכת משוואות זו נודעת בספרות כמערכת המשוואות הנורמלית או משוואות Yule Walker מסדר אפס.

בדרך דומה ניתן להראות [25] שפתרון משוואות Yule Walker, להלן Y.W., מסדר גבוה נותן את אותו הפתרון ל- $a_k, k=1, \dots, p$. להלן משוואות Y.W. מסדר גבוה יותר (מסדר ראשון):

$$\begin{bmatrix} R_x(p) & ,R_x(p-1) & , \dots ,R_x(1) \\ R_x(p+1) & & R_x(2) \\ \cdot & & \cdot \\ \cdot & & \cdot \\ \cdot & & \cdot \\ R_x(2p-1) & ,R_x(2p-2) & , \dots ,R_x(p) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{bmatrix} = - \begin{bmatrix} R_x(p+1) \\ R_x(p+2) \\ \cdot \\ \cdot \\ \cdot \\ R_x(2p) \end{bmatrix} \quad (4.17)$$

המיוחד במשוואות (4.17) הוא שהן נותנות פתרון ל- $a_k, k=1, \dots, p$, מבלי להשתמש ב- $R_x(0)$. אם נזכר שאיבר זה מושפע באופן החזק ביותר על ידי הרעש (4.10), הרי שבעצם פתרון (4.17) במקום (4.16) הופחתה השפעת הרעש על

הפיתרון. בדרך דומה ניתן לעלות בסדר של משוואות y.w. ולבטל את השפעתם של $R_x(1), R_x(2)$ וכך הלאה. בשיטה זו להקטין עוד ועוד את השפעת הרעש על חישוב המקדמים $a_k, k=1, \dots, p$.

חסרונות השיטה:

1. הזהות בפתרונות של משוואות (4.16) ו (4.17), עבור $a_k, k=1, \dots, p$, נכונה עבור פונקציית האוטוקורולציה $R_x(k)$ אשר מחושבת בצורה תיאורטית על כל ציר הזמן, כלומר

$$R_x(k) = \sum_{n=-\infty}^{\infty} s(n)s(n+k) \quad (4.18)$$

- במגבלות החישוב ובגלל אי הסטציונריות של אות הדיבור, $R_x(k)$ מחושב על קטע סופי ולפי (2.13) ולכן (4.16) ו (4.17) לא יתנו את אותו הפתרון.
- בגלל צורת חישוב $R_x(k)$, ככל שעולים בסדר המקדם משתמשים בפחות ופחות ערכים של $x(n)$, ולכן שיערוכו כמקדם קורולציה פחות טוב (שונות גבוהה יותר).
- המטריצה ב- (4.17) אינה סימטרית ולכן אין אפשרות להשתמש באלגוריתם של DUBIN.
- המטריצה הנ"ל אינה בהכרח מוגדרת חיובית, עובדה העלולה ליצור בעיות אי-יציבות במסנן ה-LPC.

4.3.5 שיערוך קטבים ואפסים [16].

שיטה זו מבוססת על תכונה (4.7), אשר מגדירה את המודל המתאים ל- $x(n)$ האות הרועש, כמודל קטבים-ואפסים, כאשר הקטבים זהים לקטבים של אות הדיבור הנקי, בעוד שהאפסים הם תרומת הרעש.

הכוונה בשיטה זו היא לחשב את הקטבים והאפסים במקביל, בעיה הידועה כבעית ARMA, ולהשתמש בקטבים שחושבו למסנן הסינטזה.

חסרונותיה העיקריים של שיטה זו הם שניים:

1. פתרון בעית מודל קטבים-אפסים היא קשה ולא לינארית במהותה.
2. אות הדיבור המעשי מכיל אפסים אשר יעלמו בשיטה זו, דבר היוצר עיוות בדבור המשוחזר, להבדיל מ-LPC אשר מקרב במידת מה גם את האפסים של ספקטרום האות.

4.4 סכום.

בפרק הנוכחי נסקרו מספר שיטות להקטנת ההשפעה של הרעש על ביצועי מקדד LPC. אין בכוונתו של חיבור זה לבחון השוואתית את השיטות הנ"ל, אלא להציגן ולהשתמש בחלקן.

בפרק החמישי והשישי יוצגו שני מודלים לקידוד אות, אחד ישתמש בשיטה של הפרדה בין קטבים לאפסים ושימוש בקטבים לאנליזה, השני ירחיב את מקדד LPC עם קידוד השארית הקפסטורלית, אשר הוצג בפרק השלישי, גם לקדוד אותות דיבור רועשים.

פרק חמישי: מקודר דיבור עם הפרדה בין קטבים
ואפסים במישור נגזרת הפאזה (DPS).

5.1 מבוא

בפרק הקודם הוצגו מספר שיטות להקטנת השפעת הרעש על ביצועי מקודר LPC. אחת השיטות שהוזכרה התבססה על הפרדת הקטבים והאפסים של המודל המיצג את האות הרועש ושימוש בקטבים בלבד לממוש מסנן הסינטזה.

קימות מספר שיטות למיצוי מקביל של קטבים ואפסים של מודל. [27] הראה דרך להפרדה כזאת עבור אות מינימום פאזה על ידי שימוש בנגזרת של ספקטרום הפאזה (DERIVATIVE OF PHASE SPECTRUM - DPS). שיטה זו נראית אטרקטיבית בגלל פשטותה וכן עקב התאמתה הנוחה למקודר הדיבור שהוצג בפרק השלישי.

בהמשך הפרק תוצגנה הבדיקות שנעשו בעזרת סימולציות על מחשב לכחינת האפשרות לשימוש בשיטה זו, (DPS) לקודר אותות דיבור. נבחנו שתי אפשרויות:

1. שימוש בשיטה זו למיצוי קטבים ואפסים של אות דיבור נקי ומימוש מסנן סינטזה הכולל קטבים ואפסים.
2. מיצוי דומה של קטבים ואפסים של אות דיבור רועש ושימוש בקטבים בלבד לסינטזה של אות הדיבור.

5.2 הפרדה בין קטבים ואפסים במישור DPS [27].

5.2.1 הצגת הבעיה.

יהי $x(t)$ אות מינימום פאזה הניתן ליצוג על ידי מודל קטבים ואפסים ותהי $X(z)$ התמרת z שלו, אזי

$$X(z) = G \frac{B(z)}{A(z)} \quad (5.1)$$

כאשר:

$$B(z) = 1 + \sum_{k=1}^q b_k z^{-k} \quad (5.2)$$

$$A(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (5.3)$$

G - מקדם ההגבר.

$x(t)$ מוגדר כאות מינימום פאזה ולכן הערך המוחלט של הספקטרום, $|x(w)|$ או הפאזה של הספקטרום מגדירים את הספקטרום חד ערכית. הקשר בין שני אלה הוא דרך התמרת הילברט [3].

הבעיה היא למצוא דרך להפריד בין איברים בספקטרום הקשורים לפולינום המכנה (קטבים), לבין איברים הקשורים לפולינום המונה (אפסים).

5.2.2 הצגת ה-DPS ותכונותיו [27].

מסנן בעל קטבים בלבד ניתן להפרדה למכפלה של מסננים מסדר ראשון, בעלי קוטב ממשי בודד, ומסננים מסדר שני בעלי זוג קטבים קומפלקסים צמודים. אם המסנן יציב כלומר כל הקטבים שוכנים בתוך מעגל היחידה, אזי הוא מינימום פאזה.

ה- DPS של מסנן יציב מסדר ראשון ניתן על ידי הנוסחה [27] (ראה נספח

:א)

$$\frac{d\theta(w)}{dw} = \theta_1'(w) = -\frac{\delta}{w^2 + \delta^2} \quad (5.4)$$

כאשר δ הוא חדר הברך של המסנן, ברדיאנים.

ה- DPS של מסנן מסדר שני בעל זוג קטבים קומפלקסים צמודים ניתן על

ידי הנוסחה

$$\theta_2'(w) = -\frac{2\alpha(\alpha^2 + \beta^2 + w^2)}{(\alpha^2 + \beta^2 - w^2)^2 + 4w^2\alpha^2} \quad (5.5)$$

כאשר β, α הם חצי ההספק והחדר המרכזי של המסנן, בהתאמה. בדרך כלל $\beta^2 \gg \alpha^2$. β, α ו δ מספרים חיוביים ולכן $\theta_1'(w)$ ו $\theta_2'(w)$ יהיו תמיד שלילים.

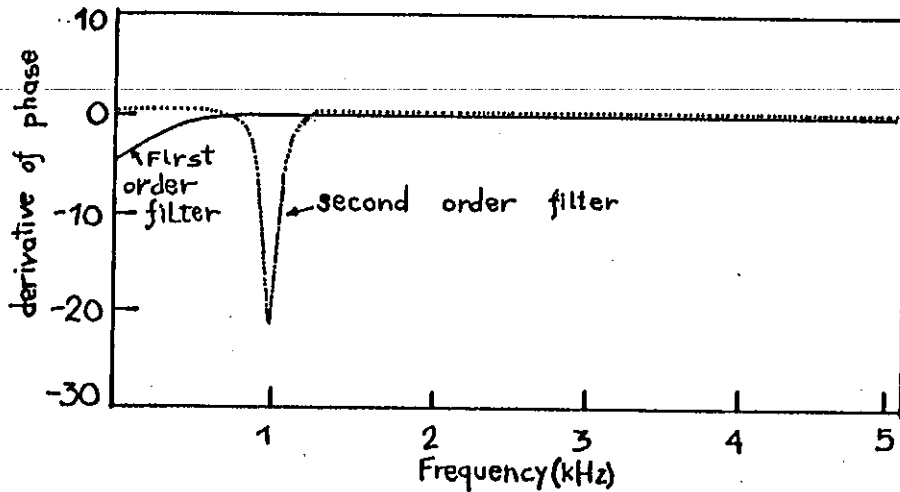
תכונות ה- DPS עבור מסננים מסדר ראשון ושני:

1. $\theta_1'(w)$ פונקציה מונוטונית עולה.
2. בתדרים נמוכים, w קטנה $\theta_1'(w) \approx -1/\delta$.
3. בתדרים גבוהים, w גדולה $\theta_1'(w) \approx -\delta^2/w^2$.
4. $\theta_2'(w)$ היא יחסית לערך המוחלט בריבוע של הספקטרום של מסנן בעל קוטב כפול עם אותם קטבים ואותו רוחב סרט.
5. בתדרים נמוכים $\theta_2'(w) \approx -2\alpha/\beta^2$ קבוע קטן.
6. עבור תדרים גבוהים $\theta_2'(w) \approx -2\alpha/w^2$, קטן מאוד.

בציור (5.1) מצויר DPS של $\theta_1'(w)$ ו- $\theta_2'(w)$ מהתכונות שהובאו לעיל ניתן להסיק מספר מסקנות:

1. עבור δ גדול מאוד $\theta_1'(w)$ קטן בכל תחום התדרים. עבור δ קטן $\theta_1'(w)$ יהיה בעל ערכים משמעותיים רק בתדרים הנמוכים מאוד. מהנאמר לעיל ניתן להסיק שלקוטב ממש? בודד תהיה השפעה מועטה מאוד על מכנה השיאים של $X(w)$, הנבנה על ידי קטבים קומפלקסים צמודים.

2. $\theta_2'(w)$ קטנה מהר מאוד עם מתרחקים מהתדר המרכזי, לכן השפעה של קוטב כפול אחד על השני תהיה קטנה אם הם אינם קרובים מאוד אחד אל השני.



צור (5.1): נגזרת ספקטרום הפאזה (DPS) עבור מסנן מסדר ראשון ומסנן מסדר שני.

Fig. (5.1): Derivative of Phase Spectrum (DPS) for first and second order filters.

ניתן להראות שעבור מסנן המיצג אפסים בלבד, פונקציה ה-DPS מתנהגת באותה צורה. השוני היחיד בין אפסים לקטבים הוא הסימן של $\theta_1'(w)$ ו $\theta_2'(w)$, עבור אפסים פונקציות אלו תהיינה תמיד חיוביות.

עבור מסנן הבנוי מקטבים ואפסים אשר אינם סמוכים מדי אחד אל השני ניתן להפריד בתוך פונקציה DFS בין רכיבים המתארים קטבים לבין אלו המתארים אפסים על ידי הפרדה בין ערכים חיוביים לערכים שליליים, החיוביים מתארים את האפסים בעוד השליליים מתארים את הקטבים.

5.2.3 הקשר בין DFS לבין הקפסטרום.

הקפסטרום של אות הוגדר בפרק השלישי. כן נמנו מספר מתכונותיו.

להי $\hat{h}(k)$ הקפסטרום הקומפלקסי של $x(w)$, כאשר $x(w)$ הוא ספקטרום בעל מינימום פאזה אזי:

$$\log[X(w)] = F[\hat{h}(k)] = \sum_{k=0}^{\infty} \hat{h}(k) \exp(-jkw) \quad (5.5)$$

כלומר

$$\log[X(w)] = \sum_{k=0}^{\infty} \hat{h}(k) \cos(kw) - j \sum_{k=0}^{\infty} \hat{h}(k) \sin(kw) \quad (5.6)$$

כאשר ה- \log הוא על פי הבסיס הטבעי.

$x(w)$ ניתן לביטוי

$$x(w) = |x(w)| \exp[-j\theta_x(w)] \quad (5.7)$$

כלומר

$$\log[X(w)] = \log|x(w)| - j[\theta_x(w) + 2\pi\delta] \quad (5.8)$$

δ שלם קבוע הנועד לשמור על רציפות הפאזה.

כאשר משווים את החלק הדימיוני והחלק הממשי בין משוואות (5.6) ו (5.8) מתקבל

$$\log|x(w)| = \sum_{k=0}^{\infty} \hat{h}(k) \cos(kw)$$

וכן

$$\theta_x(w) + 2\delta\pi = \sum_{k=0}^{\infty} \hat{h}(k) \sin(kw) \quad (5.10)$$

אם נגזור את (5.10) לפי w נקבל

$$\theta_x'(w) = - \sum_{k=1}^{\infty} k \hat{h}(k) \cos(kw) \quad (5.11)$$

הערה: לתוצאה (5.11) ניתן להגיע גם דרך תכונות התמרת פוריה ולא רק דרך גזירה ישירה.

5.2.4 אלגוריתם ההפרדה בין קטבים ואפסים.

יהי $x(n)$ אות מינימום פאזה בעל התמרת Z , $X(z)$, רציונלית המקימת את נוסחא (5.1). מצא במישור הקפטרום הקומפלקסי את הסדרה המתארת את הקטבים ואת זו המתארת את האפסים.

האלגוריתם:

1. חשב את הקפטרום הקומפלקסי, $\hat{h}(k)$, של האות $x(n)$ (מינימום פאזה).
2. חשב את $\theta_x'(w)$ בשימוש בנוסחא (5.11).
3. יהיו $[\theta_x'(w)]^+$ ו- $[\theta_x'(w)]^-$ איברים של DFS השיכים לאפסים ולקטבים בהתאמה אזי:

$$[\theta_x'(w)]^+ = \begin{cases} \theta_x'(w) & \theta_x'(w) \geq 0 \\ 0 & \text{אחרת} \end{cases} \quad (5.12)$$

$$[\theta_x'(w)]^- = \begin{cases} \theta_x'(w) & \theta_x'(w) < 0 \\ 0 & \text{אחרת} \end{cases} \quad (5.13)$$

כאשר קים הקשר:

$$\theta_x^{\sim}(w) = [\theta_x^{\sim}(w)]^+ + [\theta_x^{\sim}(w)]^- \quad (5.14)$$

4. חשב את $\hat{h}^+(k)$, $\hat{h}^-(k)$, סדרות הקפסטרום המיצגות את האפסים והקטבים בהתאמה כאשר:

$$\hat{h}^+(k) = 2/k F^{-1}\{[\theta_x^{\sim}(w)]^+\} \quad k=1, \dots, q \quad (5.15)$$

$$\hat{h}^-(k) = 2/k F^{-1}\{[\theta_x^{\sim}(w)]^-\} \quad k=1, \dots, p \quad (5.16)$$

כאשר $\{ \} F^{-1}$ מיצג התמרת פוריה ההפוכה.

בסיכום התהליך שהוצג לעיל התקבלו שתי סדרות קפסטרום המיצגות אחת את הקטבים והשניה את האפסים של הספקטרום $X(w)$.

כדי להשלים את הצגת $X(z)$ על ידי קטבים ואפסים יש לחשב את a_k ,

b_k , $k=1, \dots, p$ ו $k=1, \dots, q$, מתוך אותן שתי סדרות ספקטרום שהתקבלו

בנוסחאות (5.15), (5.16). את המקדמים האלה ניתן לחשב על ידי שימוש בנוסחא

(3.12) כפי שהוגדרה במסגרת תכונות הקפסטרום.

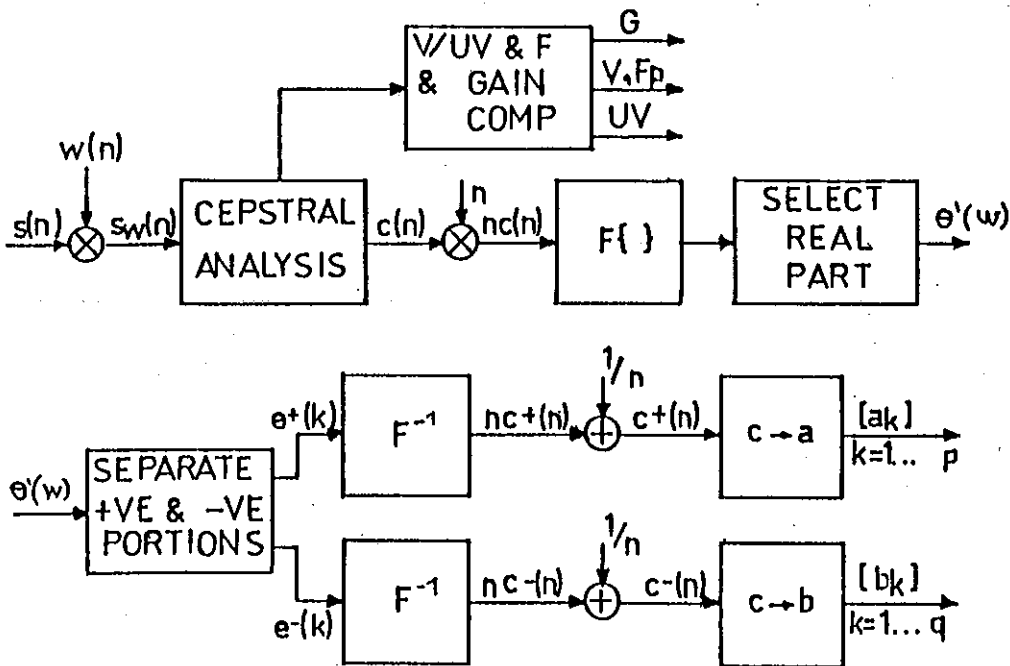
5.3 מקודד "YANG" של אות דיבור נקי.

ב [27] נבדקה יעילותו של האלגוריתם שתואר לעיל (להלן אלגוריתם "YANG"), ליצוג העוטפת הספקטרלית של אות דיבור. בסעיף הנוכחי יבדק האלגוריתם כמקודד דיבור. ביצועיו יושוו לביצועי מקודד LPC. ההשוואה תהיה על פי הממד לאי התאמה ספקטרלית, E_c , שהוצג בפרק השלישי. כן יושוו שני מקודדים אלו בבדיקות שמיעה סוביקטיבית.

המקודדים נבדקו והושוו על ידי סימולציה במחשב.

5.3.1 הצגת מערכת האנליזה ("YANG").

ציור (5.2) מתאר את מערכת האנליזה לפי אלגוריתם "YANG".



ציור (5.2): מערכת האנליזה לפי אלגוריתם "YANG".

Fig. (5.2): Analysis system - "YANG" algorithm.

הסבר:

אות דיבור $s(n)$, דגום בתדר 8000Hz מוכפל בחלון (Hanning) בעל 256 מקדמים. החלון נועד להחליק את קטע האנליזה הרלונטי בדרך שתלמנע בעיות אי רציפות בקצוות. לאחר הכפל בחלון מתקבל $s_w(n)$ אשר עובר אנליזה קפסטרלית וממנו לחשוב הקפסטרום הקומפלקסי $\hat{h}(k)$. $\hat{h}(k)$ גם הוא מוכפל בחלון, $w_c(n)$ שמטרתו לבודד את הקפסטרום, $\hat{h}_w(n)$, המיצג את העוטפת הספקטרלית של $s_w(n)$. פעולה זו ידועה בספרות בשם החלקה קפסטרלית.

מ- $\hat{h}_w(k)$ מחשבים את $\theta_g^+(k)$ (DPS), כפי שתואר בסעיפים הקודמים. על ידי שימוש במשוואות (5.12), (5.13), מחשבים את $[\theta_g^+(k)]^+$, $[\theta_g^+(k)]^-$. על ידי שימוש במשוואות (5.15), (5.16), מחשבים את $\hat{h}^+(k)$, $\hat{h}^-(k)$, ומהם על ידי הקשר המתואר בנוסחא (3.12), מחושבים הפרמטרים a_k , b_k , $k=1, \dots, q$.

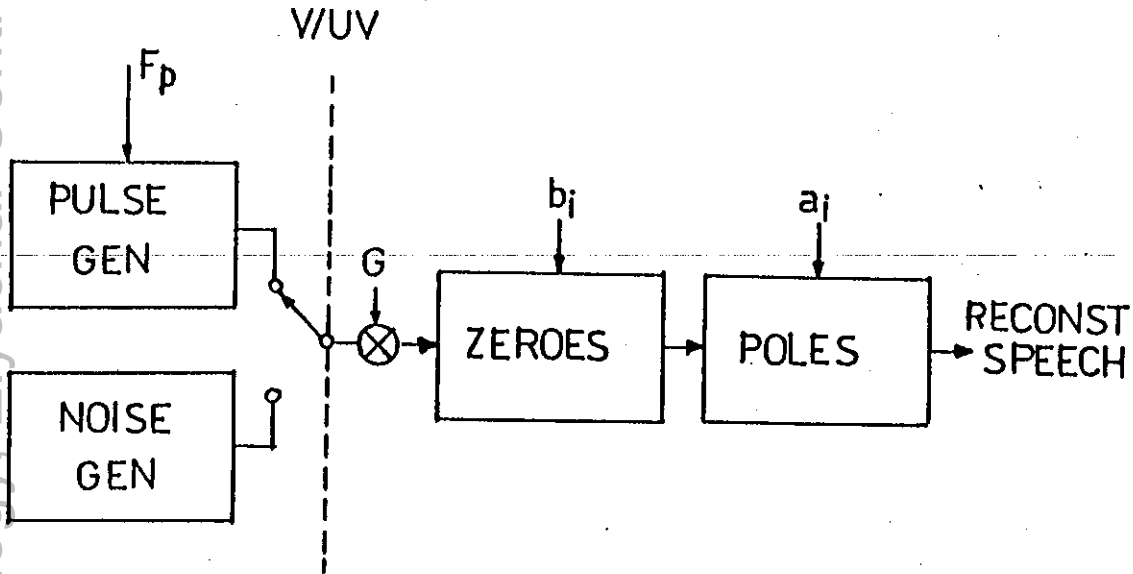
מקדם ההגבר של מערכת הקידוד, G , מחושב ישירות מ- $\hat{h}(0)$:

$$G = \log\{\hat{h}(0)\} \quad (5.17)$$

החלטת V/UV וחישוב תדר ה- PITCH הונחו כידועים ולא נבחנה השפעתם על ביצועי המקודדים.

5.3.2 מערכת הסינטזה ("YANG").

ציור (5.3) מתאר את מערכת הסינטזה המתאימה למערכת האנליזה שתוארה בסעיף הקודם.



ציור (5.3): מערכת הסינטזה - אלגוריתם "YANG".

Fig. (5.3): Synthesis system - "YANG" algorithm.

עבור מקדד LPC ניתן להשתמש באותה מערכת סינטזה כאשר מציבים $b_k=0$,
 $k=1, \dots, q$.

5.3.3 תוצאות הסימולציה.

לאחר בדיקות ראשונות של המערכת המתוארת בציור (5.2), הסתבר שקימת בעיה אי יציבות. הקטבים של $1/A(z)$ לא התקבלו בתוך מעגל היחידה. כלומר הסידרה $\hat{h}^{-(k)}$, $k=1, \dots, m$, וחת הסידרה $\hat{h}^{(k)}$, $k=1, \dots, p$, אשר ממנה חושבו a_k , $k=1, \dots, p$, אינה נותנת תמיד מסנן יציב בשימוש בקשר (3.12). בעיה זו נובעת מכך שאחת ההנחות לחישוב (3.12) היא ש- a_k , $k=1, \dots, p$, מיצגים מסנן מינימום פאזה. אך לא כל קפסטרום חד צדדי מתאים למסנן ALL POLE מינימום פאזה לפי הקשר (3.12).

דוגמא:

בכדי להסביר זאת תוכא דוגמא.

$$A(z) = 1 + 3/2z^{-1} + 5/8z^{-2} \quad (5.18)$$

$A(z)$ פולינום במישור z בעל אפסים בתוך מעגל היחידה. מ (3.11) מתקבל:

$$\hat{h}(1) = 3/2 \quad (5.19)a$$

$$\hat{h}(2) = 21/8 \quad (5.19)b$$

נחשב את $\bar{A}(z)$ הפולינום מסדר ראשון המתקבל מהסדרה $\hat{h}(n)$ המתוארת ב- (5.19). מ- (3.12) נקבל:

$$\bar{A}(1) = 3/2 \quad (5.20)$$

ומתקבל הפולינום:

$$\bar{A}(z) = 1 + 3/2z^{-1}$$

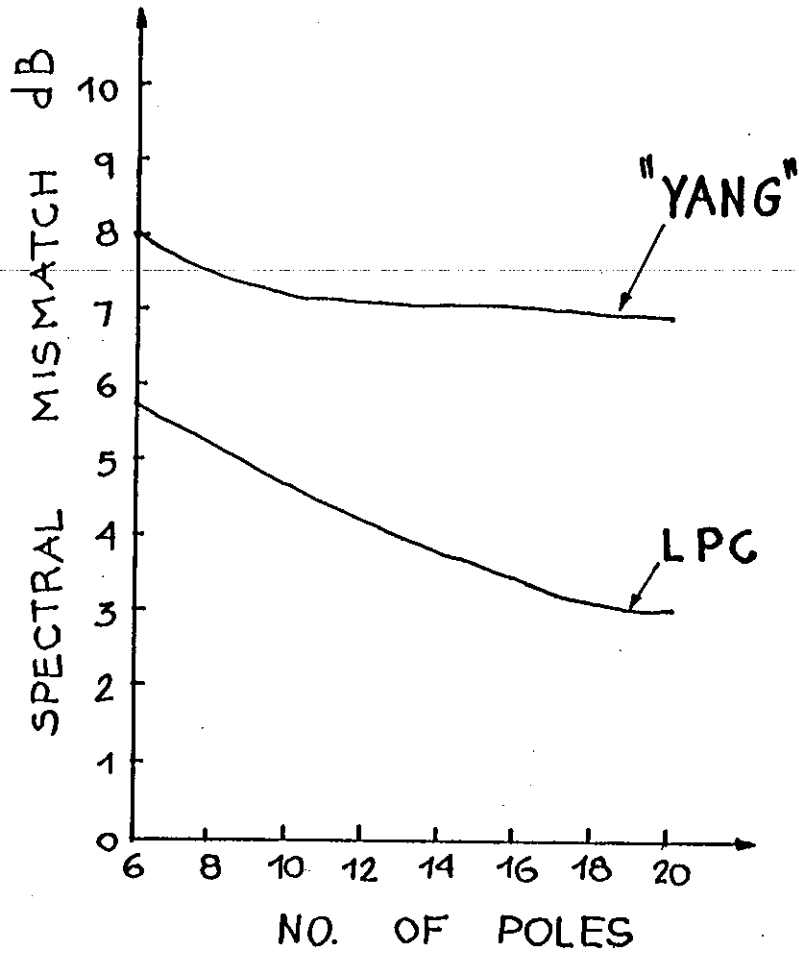
פולינום זה מכיל אפס מחוץ למעגל היחידה ולכן כאשר הנו במכנה של מסנן מתקבל מסנן לא יציב.

כדי להתגבר על בעיית היציבות שהוזכרה לעיל, הוחלפה תת המערכת במערכת האנליזה, המחשבת את a_k מתוך $\hat{h}^-(k)$ ישירות על פי (3.12), במערכת המחשבת ראשית את אות התגובה להלם הזמני המתאים ל- $\hat{h}^-(k)$ וממנו חושבו a_k , $k=1, \dots, p$. על ידי אנליזת LPC.

באותה דרך גם חושבו מקדמי האפסים מחשש שסדרת הקפסטרום $\hat{h}^+(k)$, אינה מתאימה לספקטרום בעל מינימום פאזה.

לאחר פתרון בעיית היציבות כפי שהוצג לעיל, נבדקו ביצועי מקדוד "YANG" לעומת ביצועי מקדוד LPC.

ציור (5.4) מתאר את E_c כתלות במספר הקטבים עבור שני האלגוריתמים, ללא אפסים עבור אלגוריתם "YANG".



ציור (5.4): השפעת סדר המודל על מדרד אי ההתאמה הספקטרלית עבור אלגוריתם ה-LPC ואלגוריתם "YANG" (ללא אפסים).

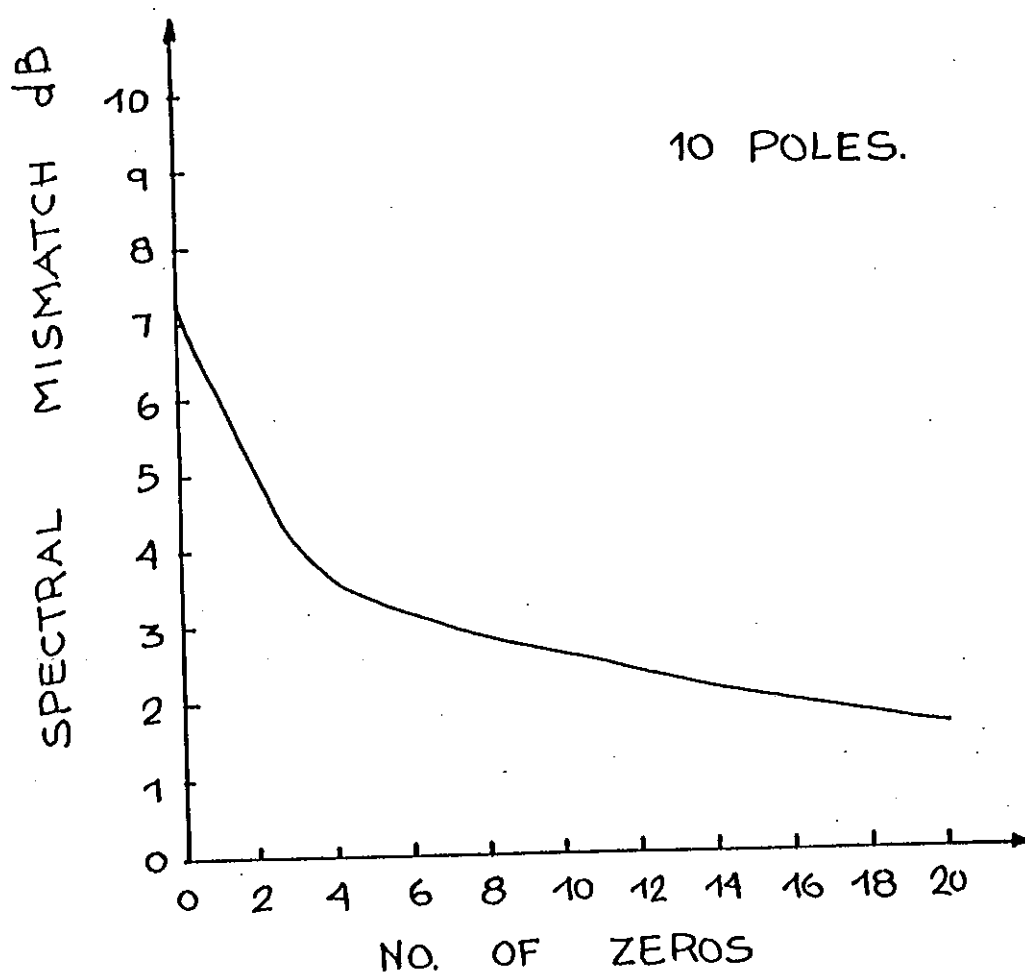
Fig. (5.4): Spectral mismatch versus model order: LPC algorithm and "YANG" algorithm (no zeros).

מצויר זה ניתן להסיק את המסקנות הבאות:

1. אלגוריתם "YANG" מציג את הקטבים הקיימים בעוטפת הספקטרלית של הדיבור. הגדלת מספר הקטבים מעל 10, מספר המספיק ליצג את הפורמנטים הקיימים בדיבור ברוחב הסרט הרלונטי, אינה משפרת את איכות ההתאמה הספקטרלית.

2. ללא שימוש באפסים, אלגוריתם LPC מיצג את העוטפת הספקטרלית טוב יותר מאשר אלגוריתם "YANG".

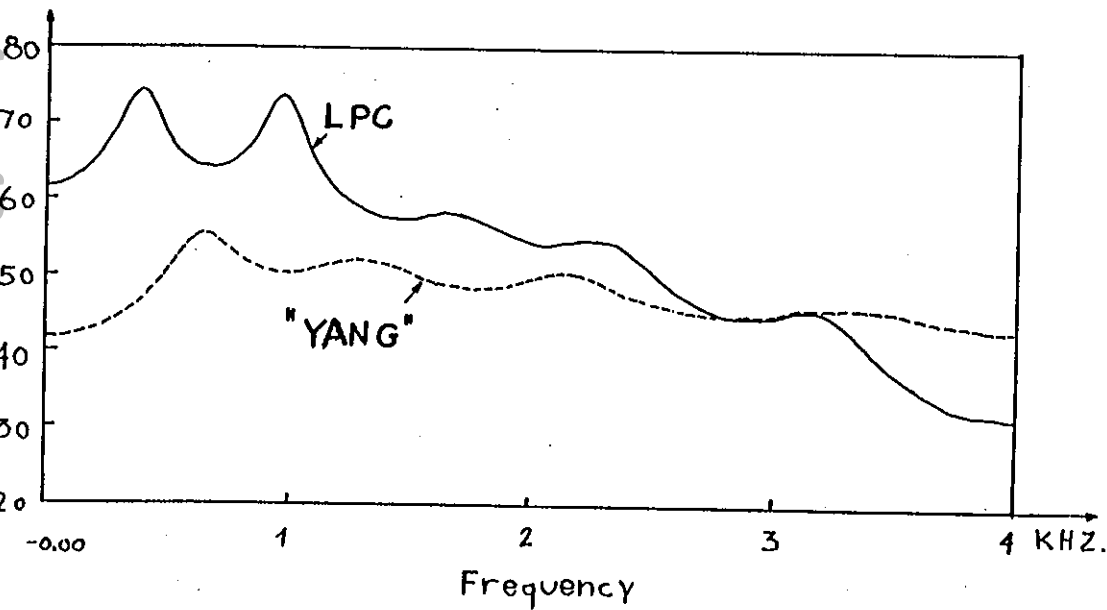
בציור (5.5) מתוארת אי ההתאמה הספקטרלית עבור מערכת "YANG" בעלת 10 קטבים ומספר משתנה של אפסים.



ציור (5.5): השפעת האפסים על ההתאמה הספקטרלית עבור אלגוריתם "YANG", מספר הקטבים הוא 10.

Fig. (5.5): The contribution of zeros to the spectral matching - "YANG" algorithm, number of poles is 10.

בציור (5.5) ברור השיפור בהתאמה הספקטרלית עם הגדלת מספר האפסים. כאשר מספר האפסים מספיק גדול אלגוריתם "YANG" מקרב את העוטפת הספקטרלית טוב יותר מאשר אלגוריתם LPC. אולם כדי להגיע לאיכות התאמה זו יש לקודד את האפסים ולשלם בקצב הסיביות. כדי לבצע השוואה הוגנת יש להשוות ביצועי מקודד "YANG" עם מספר גדול של אפסים, לביצועי מקודד PCR, אשר הוצג בפרק השני, בעל אותו קצב סיביות.



ציור (5.6): תגובת החדר של מסנן הסינטזה עבור אלגוריתם LPC ואלגוריתם "YANG", 10 קטבים וללא אפסים.

Fig. (5.6): The spectrum of synthesis filter for LPC and "YANG" algorithm, 10 poles and no zeros.

מציורים (5.4) ו (5.5) אלגוריתם "YANG" נראה כמיצג יעיל של העוטפת הספקטרלית, עבור קצב סיביות מינימלי המאפשר קידוד 10 קטבים וכ 20 אפסים יעילותו של אלגוריתם "YANG" נמדדת במקרה זה עבור מדד אי ההתאמה הספקטרלית, אולם כאשר באים לבחון את ביצועי מקודד "YANG" במבחני שמיעה סובייקטיבית, מסתבר שאיכות הדיבור המשוחזר ממקודד זה ירודה והדיבור יוצא מעוות למדי. כאשר יורדים לשורש העניין מסתבר שאלגוריתם "YANG" מקרב בצורה די יעילה את העוטפת הספקטרלית בממוצע על פני כל תחום החדר, אולם לוקה במיקום מדויק וברוחב הסרט של הפורמנטים. מכיון שלא להתייחס לתרומה העיקרית

לאיכות הדיבור, נכשל אלגוריתם "YANG" כמקדד. בצירוף (5.6) מושווה ספקטרום הקטבים לפי אלגוריתם "YANG" לזה של אלגוריתם LPC.

תוצאה זו מסבירה את הלקוי במדד מספרי כמו E_c ומגדירה את הצורך בבדיקות שמיעה בנוסף למדד מספרי.

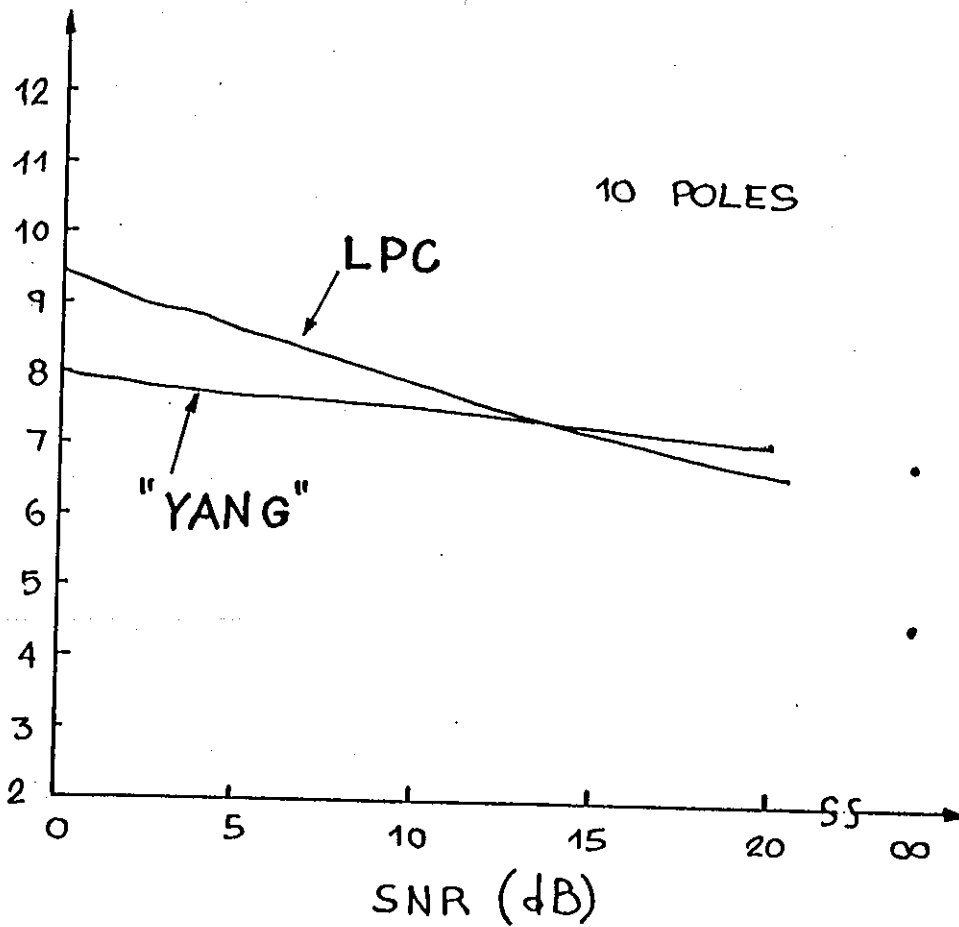
5.3.4 מסקנות.

בכל הקשור ל- E_c , המדד לאי ההתאמה הספקטרלית, אלגוריתם "YANG" נראה מתאים לשמש כמקדד דיבור, כאשר נבחר מספר מספיק גדול של אפסים בנוסף על 10 קטבים. ביצועיו אף ניתנים להשוואה לביצועיהם של המודלים, PCR ו PZCR, שהוצגו בפרק השלישי.

כאשר ניגשים להשוואה ביניהם כמבחני שמיעה סלקטיבית מסתבר שאיכות הדיבור המשוחזר ממקדד "YANG" ירודה. הסיבה לאיכות הירודה היא אי הדיוק בשערוך מיקום הקטבים ורוחב הסרט שלהם.

5.4 מקדד "YANG" עבור אות דיבור רועש.

בחינת ביצועיו של אלגוריתם "YANG" כמקדד דיבור רועש היא אקדמית בלבד וזאת לאור ביצועיו הגרועים בקידוד אות נקי. אולם כדי להשלים את בדיקת האלגוריתם ותכונותיו, תובאנה בקצרה תוצאות ומסקנות של סימולציות של מקדד זה על אות רועש.

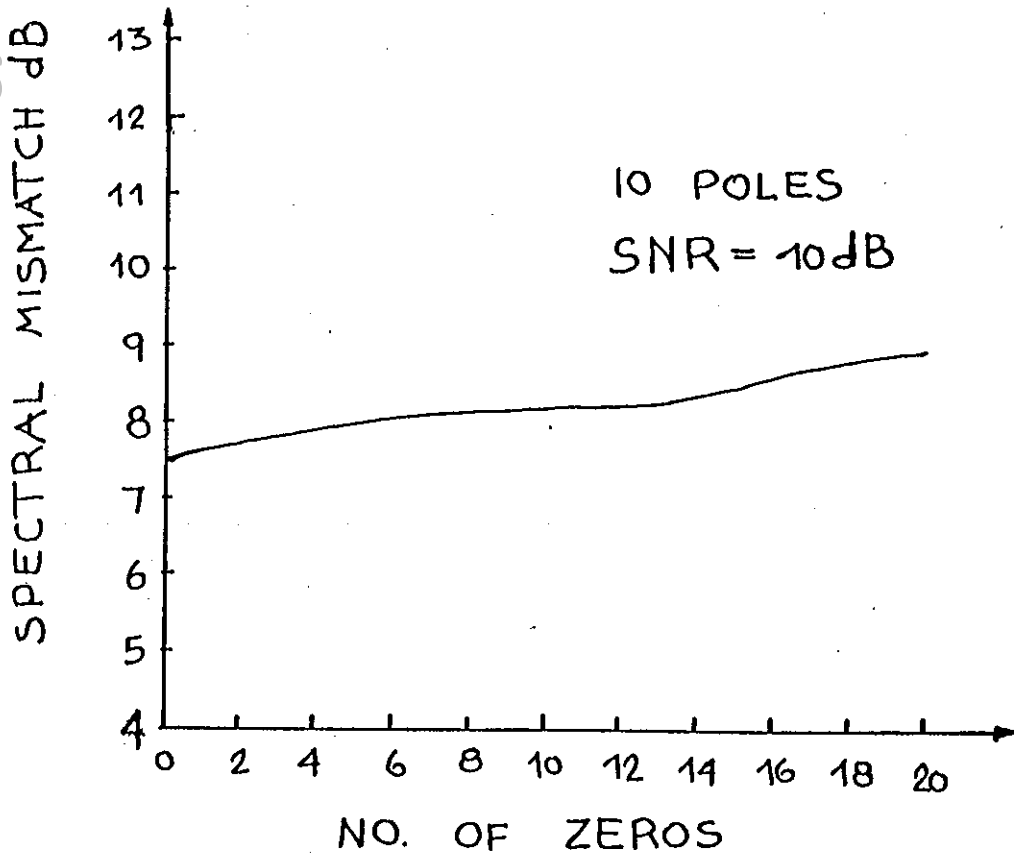


צילור (5.7): ההתאמה הספקטרלית עבור LPC ו- "YANG" בתלות ביחס אות לרועש עבור דבור רועש, 10 קטבים וללא אפסים.

Fig. (5.7): Spectral mismatch versus SNR for LPC and "YANG" algorithm, 10 poles and no zeros.

ציור (5.7) מתאר את אי ההתאמה הספקטרלית בתלות ביחס אות לרעש, (SNR), כאשר זה מבוטא ב dB, עבור מקודד "YANG" ומקודד LPC. שני המקודדים נבדקו עם 10 קטבים וללא אפסים.

מציור זה מתברר שביחסי אות לרעש של 15dB ומטה, אלגוריתם "YANG" מקרב את העוטפת הספקטרלית טוב יותר מאשר LPC. הסיבה לכך היא ש-"YANG" המכיל קטבים בלבד, מתעלם (או כמעט מתעלם), מתכולת האפסים בספקטרום. מכיוון שהרעש תורם בעיקר לאפסים אזי הוא משפיע מעט על אלגוריתם זה.



ציור (5.8): ההתאמה הספקטרלית בתלות במספר האפסים עבור אלגוריתם "YANG", עבור אות דבור רועש בעל SNR=10dB. מספר הקטבים הוא 10.

Fig. (5.8): Spectral mismatch versus number of zeros for "YANG" algorithm, noisy speech with SNR=10dB. number of poles is 10.

צ'ור (5.8) מתאר את אי ההתאמה הספקטרלית עבור מקדד "YANG" בעל 10 קטבים ומספר משתנה של אפסים, זאת עבור אות דיבור בעל $SNR=10dB$. מצ'ור זה מסתבר שתוספת אפסים מקלקלת את ההתאמה הספקטרלית בין ספקטרום האות המשוחרר לאות הנקי, מכיוון שהם, האפסים, מחזירים את הרעש אל האות המשוחרר.

בבדיקת ביצועיו של מקדד "YANG" עבור אותות רועשים, מסתבר שאיכותו של אות הדיבור המשוחרר אינו נופלת מזה של מקדד LPC. הבעיה היא ששניהם גרועים מדי לשימוש.

5.5 סיכום

מתוצאות הסימולציה שהובאו לעיל, הורדה מן הפרק האפשרות לשימוש באלגוריתם "YANG" כמקדד דיבור. איכות הדיבור המשוחזר ממקדד זה היתה נמוכה עקב שגיאות במיקום הפורמנטים וברוחב הסרט שלהם.

עדיין נשארה האפשרות להשתמש באלגוריתם "YANG" כאשר זקוקים להשוואה של חתימה ספקטרלית ממוצעת בתדר, של אות דיבור רועש.

המסקנות לגבי מקדד "YANG" מחזירות אותנו ל-LPC עם ניקוי הרעש. בפרק הבא יוצג מקדד דיבור המבוסס על מקדד PCR שהוצג בפרק השני, מלווה בניקוי דיבור בשיטת ההפחת הספקטרלית, SPS.

פרק שישי: מקוד דיבור רועש עם הפחתה ספקטרלית
וקידוד השארית הקפסטורלית.

6.1 מבוא

בפרק השלישי הוצגה מערכת קידוד עם קידוד השארית הקפסטורלית, עבור אות דיבור ללא רעש.

בפרק הרביעי הוצגו מספר שיטות לסינון דיבור רועש ולשיפור יעילותו של מקוד דיבור.

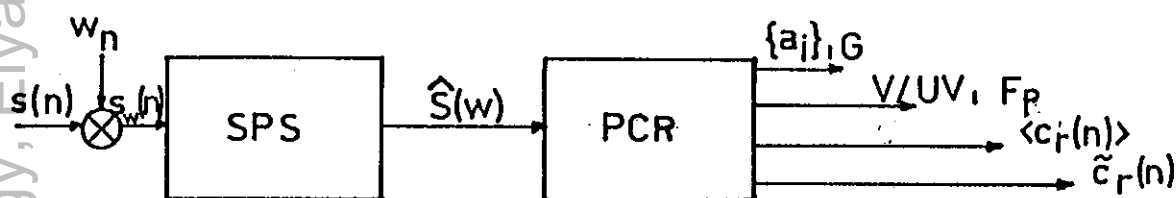
בפרק הנוכחי תוצג מערכת המשתמשת בתוצאות הפרקים השלישי והרביעי, כלומר מערכת קידוד עם קידוד השארית הקפסטורלית, כאשר מקדים אותה תהליך של סינון אות הדיבור בשיטת ההפחתה הספקטרלית (SPS).

בסעיף הבא יוצג המקודר המוצע ובסעיפים שלאחריו יוצגו מספר בעיות שהתעוררו והפתרון המוצע לכל בעיה.

6.2 הצגת מערכת האנליזה.

6.2.1 המבנה הכללי של המערכת.

ציור (6.1) מתאר את מערכת האנליזה בצורה כללית.



ציור (6.1): מערכת האנליזה של מקודר LPC עם קודר השארית הקפסטרלית עבור דיבור רועש - מבנה כללי.

Fig. (6.1): LPC with cepstral residual coding for noisy speech - analysis system.

הסבר:

אות הדיבור הרועש נדגם לקבלת $s(n)$ אשר מוכפל בחלון להגדרת קטע האנליזה הרלונטי ומתקבל $s_w(n)$. על האות $s_w(n)$ מופעל אלגוריתם ההפחתה הספקטרלית, (SPS), לניקוי מרעש. אלגוריתם הניקוי מסתפק בשיערוך הספקטרום של אות הדיבור המנוקה $\hat{S}(w)$. $\hat{S}(w)$ משמש ככניסה למערכת קידוד מסוג PCR כפי שתוארה בפרק השלישי, כאשר זו עוברת שינוי כדי שתפעל ישירות על ספקטרום אות הדיבור להבדיל מזו שפועלת על האות הזמני.

בהמשך יתואר כל בלוק המופיע בציור (6.1) בנפרד.

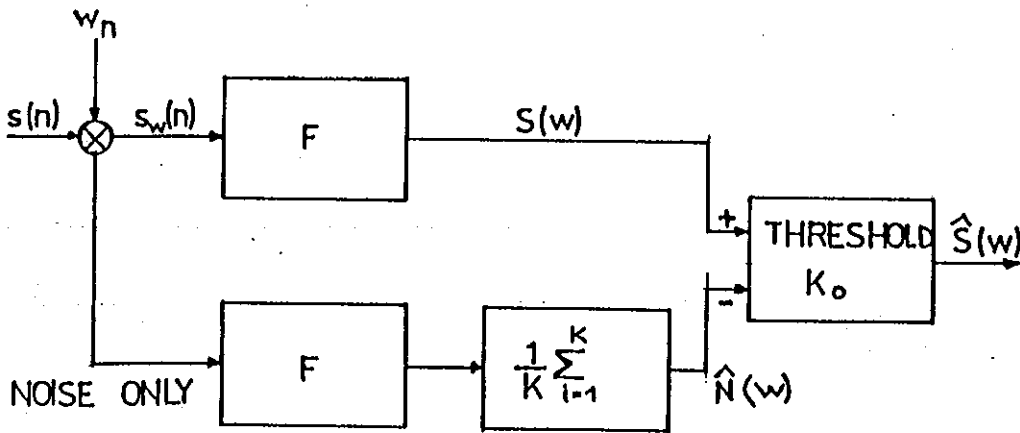
6.2.2 פירוט של מערכת האנליזה.

6.2.2.1 דגימה והפחתה ספקטרלית (SPS).

אות הדיבור הרועש $s(t)$ נדגם בקצב של 8000 דגימות בשניה ומתקבל $s(n)$, אות זה מוכפל בחלון Hanning בעל 256 מקדמים, פעולה המגדירה קטע אנליזה בעל אורך של 256 דגימות כלומר 32 msec. נבחר חלון Hanning בכדי למנוע בעיות אי רציפות בקצוות הקטע.

הקידום מקטע לקטע מתבצע בקפיצות של 128 דגימות ומגדיר חפיפה של חצי בין קטע אנליזה אחד לשני.

מערכת הניקוי בשיטת ההפחתה הספקטרלית ממומשת על פי העקרונות שתוארו בפרק הרביעי, ומתוארת בציור (6.2).



ציור (6.2): מעגל הכניסה וההפחתה הספקטרלית עבור מקודד הדיבור הרועש.

Fig. (6.2): Input filter and SPS for the noisy speech Vocoder.

ספקטרום הרעש משוערך על ידי מיצוע הספקטרום על קטעי רעש בלבד.

ספקטרום האות הרועש מחושב מ- $s_w(n)$ ומוצע על פני שלושה קטעים עוקבים להחלקת הרעש לפני תהליך ההפחתה.

ההפחתה הספקטרלית ממומשת לפי נוסחא (5.1) המתאימה לנוסחא (4.12), כאשר מתווסף אליה הפרמטר b מנוסחה (4.13), ונבחר $b=2$, בכדי להגדיל את כמות הרעש המופחת, זאת מתוך סימולציות המוכיחות שיפור בביצועים על פני $b=1$. K_0 , הסף בהפחתה הספקטרלית נבחר 10^{-3} .

$$\hat{S}(w) = \begin{cases} S(w) - 2\hat{N}(w) & S(w) - 2\hat{N}(w) > 10^{-3} \\ 10^{-3} & \text{אחרת} \end{cases} \quad (6.1)$$

באלגוריתם ההפחתה הספקטרלית מסתפקים בשיערוך הספקטרום המנוקה $\hat{S}(w)$ (אין ממשיכים בחישוב האות הזמני המשוחזר), אשר משמש ככניסה למערכת קידוד מסוג PCR. מ- $S(w)$ תחושב פונקציית האוטוקורלציה (על ידי התמרת פוריה) הדרושה לחישוב מקדמי ה-LPC. בדרך זו יש חסכון גדול בחישובים יחסית לשיטה המקובלת של מעבר דרך האות הזמני.

6.2.2.3 מערכת PCR.

ציור (6.3) מתאר את מערכת הקידוד PCR המתאימה למקדר הנוכחי.

הסבר:

אות הכניסה למערכת הקידוד המתוארת בציור (6.3) הוא הספקטרום המנוקה $\hat{S}(w)$. מ- $\hat{S}(w)$ מחושבת פונקציית האוטוקורלציה $\hat{R}_g(k)$, על ידי התמרת פוריה, כאשר לפני ביצוע ההתמרה מוסיפים אפסים לספקטרום בכדי למנוע חישוב אוטוקורלציה סיבובית.

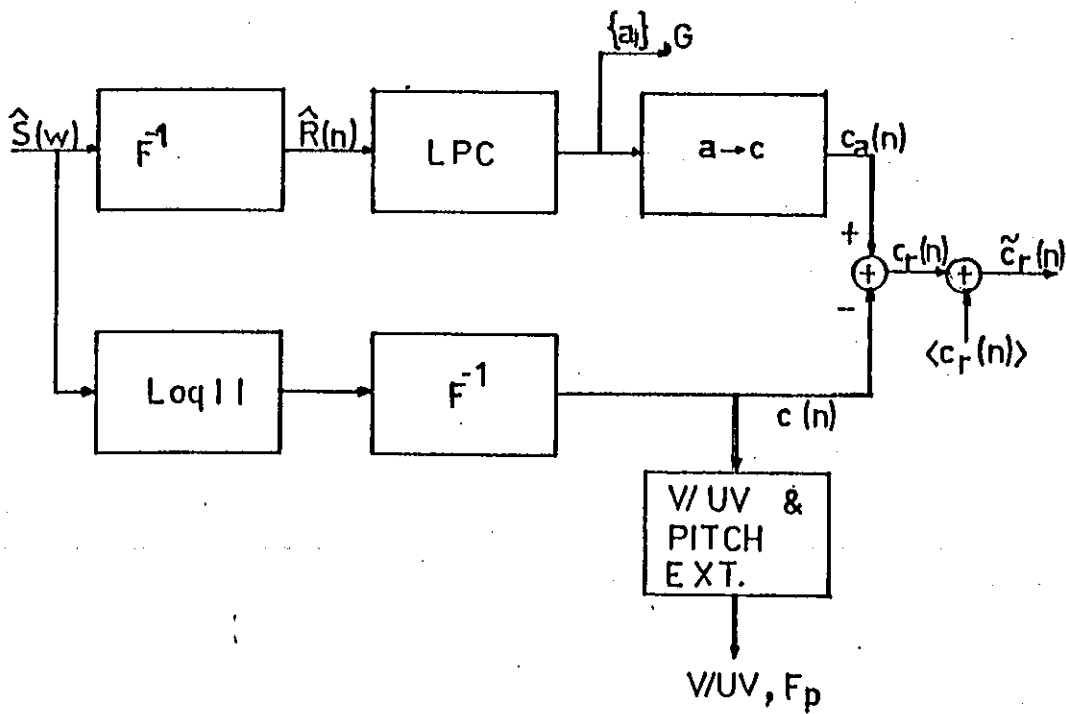
על פונקציית האוטוקורלציה שהתקבלה $\hat{R}_g(k)$, מתבצע אלגוריתם ה-Durbin כפי שתואר בפרק השני. סדר מודל ה-LPC נבחר 10 . מתוך a_i , $i=1, \dots, p$, -1 , G מחושב $c_a(n)$ הקפסטרום המיצג את מסנן ה-LPC שחושב.

כמקביל מחושב הקפסטרום המתאים לספקטרום המנוקה $\hat{S}(w)$, $c(n)$. על ידי הפחתת $c_a(n)$ מ- $c(n)$ מתקבל אות השארית הקפסטרלית $c_r(n)$. הממוצע לאורך זמן של $c_r(n)$, $\langle c_r(n) \rangle$, מחושב בצורה רקורסיבית כפי שהוצגה בנוסחה (3.14).

לאחר הפחתה מתקבלת השארית הקפסטרלית חסרת ההטיה $\hat{c}_r(n)$.

לאחר סיום החישובים שתוארו לעיל מקודדים פרמטרי ה- a_i LPC, $i=1, \dots, p$, ומקדמי השאריות הקפסטריית $\tilde{c}_r(n)$. כל אלה מועברים למערכת סינטזה PCR רגילה כפי שתוארה בצירור (3.7).

כאשר מומשה המערכת שתוארה לעיל בסיומולציה על מחשב התעוררו מספר קשיים שעליהם ועל פתרונם נעמוד בסעיפים הבאים.



צירור (6.3): מערכת PCR המקבלת ככניסה את הספקטרום של האות.

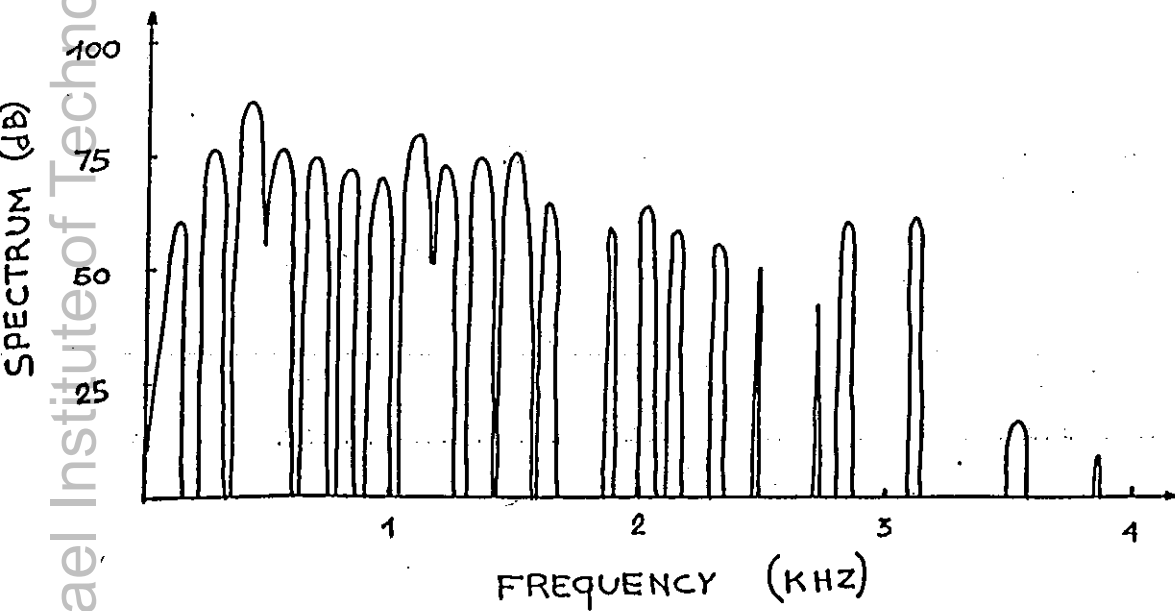
Fig. (6.3): PCR system changed to get the spectrum as input.

6.3 בעיות אי היציבות של מסנן LPC

כאשר מומשה המערכת שהוצגה בסעיפים הקודמים בפרק הנוכחי, התגלתה תופעה של אי יציבות במסנן LPC שחושב. שתי סיבות נראות כגורמות העיקריות לאי יציבות זו ושתייהן נובעות ממכנה הספקטרום המנוקה שחושב בתהליך ההפחתה הספקטרלית:

1. התחום הדינמי של הספקטרום הנ"ל גדול תופעה המשפיעה בדרך כלל על יציבות מסנן LPC [2].
2. נקודות אי הרציפות הקימות בספקטרום הנ"ל.

ציור (6.4) מתאר את הספקטרום הנ"ל.



ציור (6.4): ספקטרום שנוקה על ידי אלגוריתם ההפרדה הספקטרלית.

Fig. (6.4): A spectrum that was cleaned with SPS algorithm.

קימים מספר פתרונות מקובלים לבעיות אי היציבות במסנן ה-LPC. הפתרון שנבחר בעבודה זו ומימושו יתוארו להלן.

בכדי לפתור את בעיית אי היציבות נשתמש בשתי תכונות של אלגוריתם

DURBIN כפי שתוארו בפרק השני:

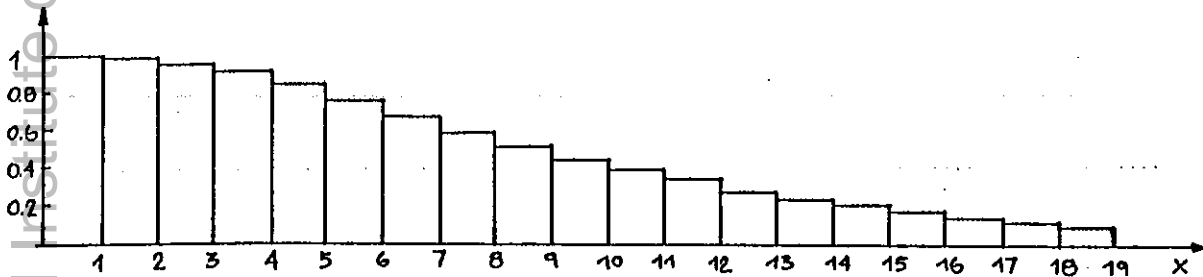
1. תוך כדי חישוב מסנן LPC מסדר p כלשהו מחושבים המסננים מכל סדר הקטן $m-p$.

2. תנאי הכרחי ומספיק ליציבותו של מסנן ה-LPC הוא שכל מקדמי החזרה החלקית k_i , $i=1, \dots, p$, יקומו $|k_i| < 1$.

משתי תכונות אלו נגזרת השיטה המוצעת:

חשב את מודל ה-LPC לפי אלגוריתם DURBIN עד לקבלת מודל LPC מסדר p הרצוי, במקרה הנוכחי $p=10$. אם תוך כדי ביצוע האלגוריתם התקבל מקדם החזרה חלקית k_j כך ש- $|k_j| \geq 1$ אזי מוקפא המודל לזה שחושב בצעד הקודם ($j-1$) באלגוריתם ואת כל יתר מקדמי החזרה מאפסים.

בכדי לבדוק את השפעת השיטה שתוארה לעיל על סדר המסנן הנבחר, בוצע חישוב סטטיסטי של מספר המקדמים שנבחרו בקטעים השונים.

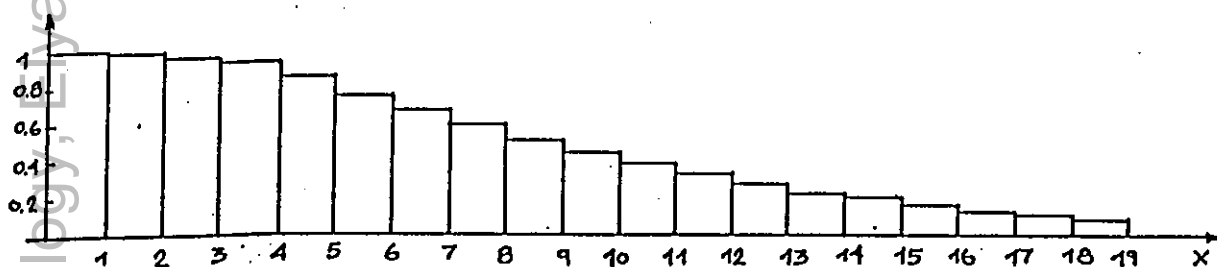


ציור (6.5): ההסתברות לסדר המודל הנבחר (בדיקה על כל הקטעים).

Fig. (6.5): The probability of chosen model order (taking into consideration all the frames).

בציורים (6.5) ו (6.6) מתוארת פונקציית ההסתברות $P(x > p)$ כאשר p הוא סדר המסנן הנבחר. סטטיסטיקה זו נעשתה עד מודל מסדר 20 כאשר אנו מתעניינים בעקר עד סדר 10.

בציור (6.5) מחושבת פונקצית ההסתברות הנ"ל על פני כל הקטעים שעברו אנליזה, בעוד שבציור (6.6) מדובר על קטעים שהאנרגיה שלהם גדולה מסף מסוים וזאת בכדי לבדוק את ההשפעה על הקטעים המכילים אות דיבור בנוסף לרעש. מציורים אלו מסתבר שעבור יותר מ- 75% מהקטעים נבחר מודל מסדר גדול מ- 8 ועבור יותר מ- 90% מהקטעים נבחר מודל מסדר הגדול מ- 6.



ציור (6.6): ההסתברות לסדר המודל הנבחר (בדיקה רק על קטעים בעלי אנרגיה מינימלית מסוימת המגדירה קטע המכיל דבור).

Fig. (6.6): The probability of chosen model order (taking into consideration only frames with large enough energy to indicate that the frame us not noise alone).

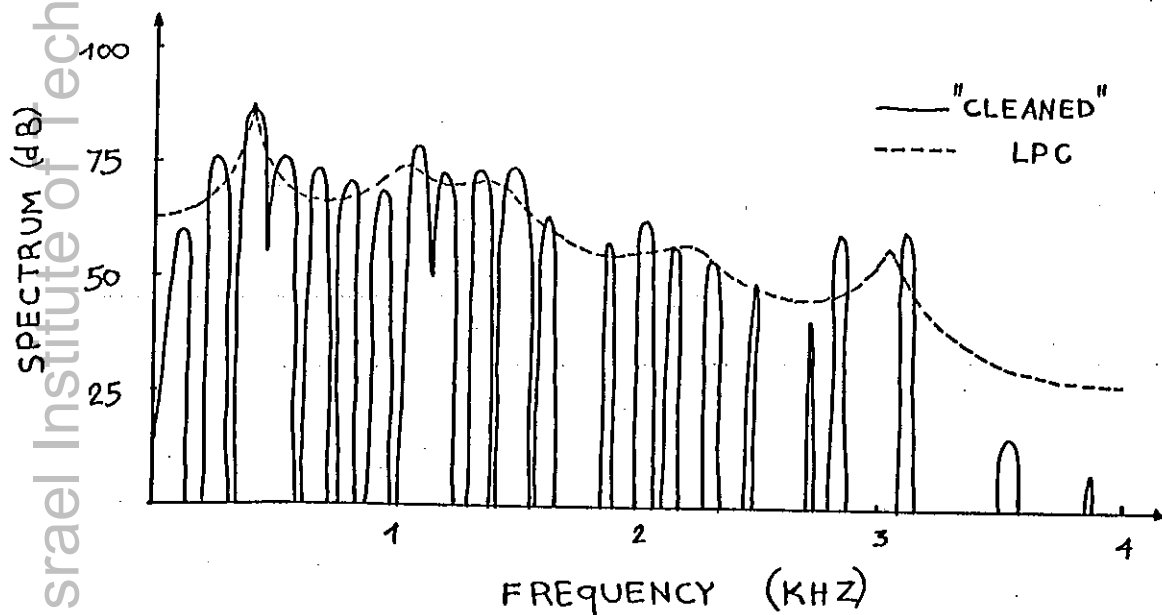
הפתרון לבעיה אי היציבות שתואר לעיל הנו פשוט מאד למימוש ומציוורים (6.5) ו-(6.6) מסתבר שהנו יעיל, דבר המקבל חיזוק נוסף כאשר מתבצעות בחינות שמיעה לדיבור המשוחזר. ניתן להוסיף ולהקטין את השפעת הקטעים בהם נבחר מסנן מסדר נמוך, על ידי ניצול הסיביות המתפנות ממקדמי ההחזרה לקידוד יעיל יותר של השארית הקפסטרילית.

6.4 מקדד LPC ישיר על הספקטרום המנוקה

לפני בחינת איכותו של מקדד ה-PCR הכולל על הספקטרום המנוקה, נבדקה איכותו של מקדד LPC לבדו וזאת מפאת חשיבותו הרבה של זה כגרעין מקדד ה-PCR.

כאשר מושווה מקדד LPC הפועל על הספקטרום המנוקה לזה הפועל על האות הזמני המנוקה באותה שיטה - הפחתה ספקטרלית, מסתבר שאיכותו של השני טובה יותר, כלומר מימוש אלגוריתם LPC ישירות על הספקטרום המנוקה מקטין את איכות ביצועי המקדד.

בציור (6.7) מתואר הספקטרום המנוקה והעוטפת הספקטרלית המתארת את מסנן ה-LPC.



ציור (6.7): הספקטרום המנוקה ותגובת החדר של מסנן ה-LPC שהתקבלו עבור אלגוריתם ההפחתה הספקטרלית המקובל.

Fig. (6.7): Estimated spectrum and the spectrum of the LPC model as recieved from ordinary SPS algorithm.

בציור זה בולט מבנהו של הספקטרום המנוקה ומתבררות הסיבות לאיכותו הגרועה של מקדד ה-LPC עליו. התחום הדינמי הגדול של הספקטרום המנוקה

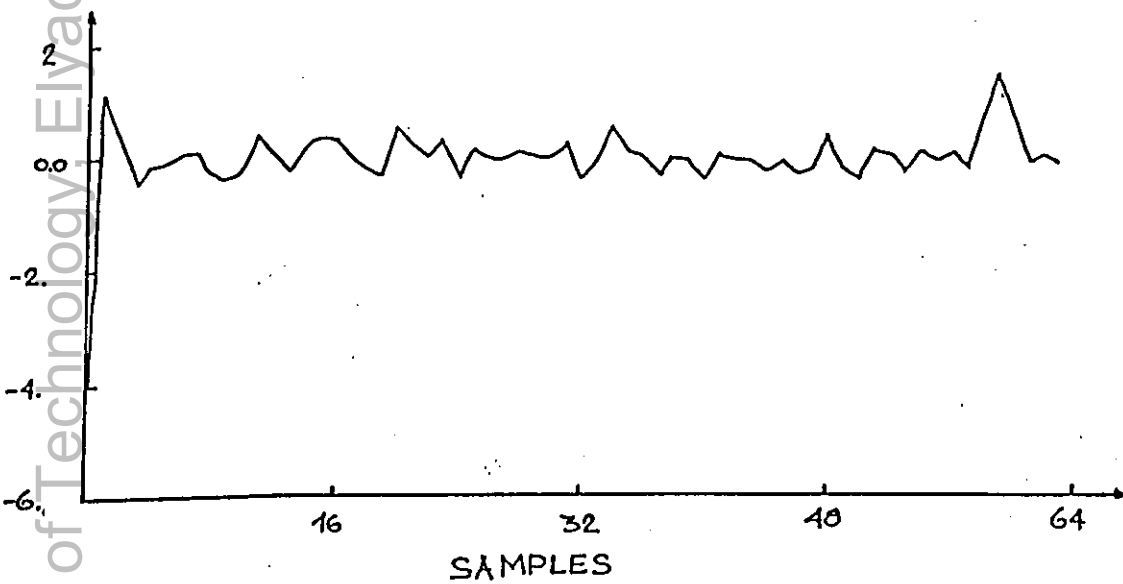
והקפיצות הגדולות בנקודות אי הרציפות הן הגורמות לעיוות בדיבור המשוחזר. תופעות אלו נעלמות באות הזמני המנוקה תודות להחלקה הנובעת מפעולת ה-OVERLAP & ADD שבסיום אלגוריתם ההפחתה הספקטרלית המקובל, וכן סינון הכניסה של האות לפני ביצוע אלגוריתם ה-LPC.

התוצאות שתוארו לעיל מצביעות על יתרון במימוש המקדד תוך מעבר דרך האות הזמני וללא קיצור הדרך שנבחר. מצד שני יתרונו של קיצור הדרך מתבטא בחיסכון ניכר בחישובים. עובדה זו גרמה לחיפוש דרכים לביטול התופעה שהוצגה לעיל.

בסעיף הבא תואר בעיה נוספת הנובעת ממבנהו של הספקטרום המנוקה ובסעיף שלאחריו יוצג הפתרון לבעיות האלו.

6.5 מבנה השארית הקפסטרלית.

ציור (6.8) מתאר את אות השארית הקפסטרלית שנוצר תוך כדי מימוש מערכת ה-PCR. ציור (6.9) את אותו הנוצר תוך כדי קידוד אות דיבור נקי (שום להבדל בקנה המידה בשני ציורים אלו).

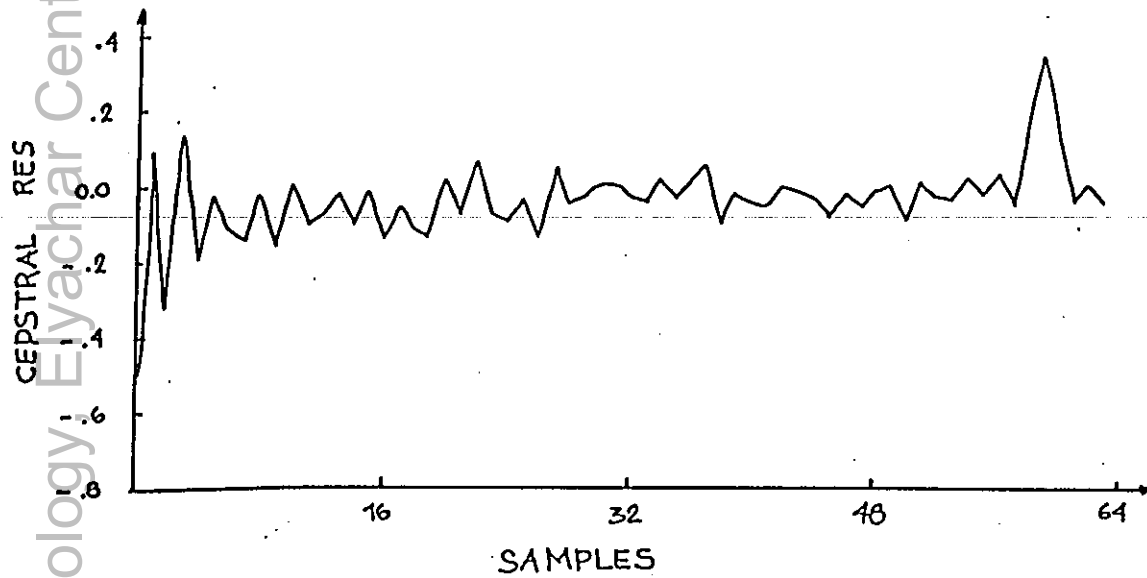


ציור (6.8): השארית הקפסטרלית המתקבלת עבור מודל PCR כאשר מופעל אלגוריתם ה-SPS המקובל.

Fig. (6.8): Cepstral residual received in PCR model after using ordinary SPS.

מציורים (6.8) ו (6.9) מסתבר שאות השארית הקפסטרלית עבור הספקטרום המנוקה הנו בעל אנרגיה הרבה יותר גבוהה מזו של אות השארית הקפסטרלית עבור דיבור נקי, איבריו מסדר גבוה אינם דועכים מספיק מהר. שתי עובדות אלו מצביעות על הצורך בהשקעה גדולה של סיביות בכדי לקודד ביעילות את השארית הקפסטרלית הנ"ל.

בעיה נוספת שהתעוררה היא שהוספת מקדמי השארית למודל ה-LPC כפי שמבצעים במערכת PCR מקלקלת את איכות הדיבור המשוחזר. כלומר במצב זה אלגוריתם PCR בתנאי רעש אינו נראה משופר על פני אלגוריתם LPC רגיל.



ציוור (6.9): השארית הקפטטרלית המתקבלת עבור מודל PCR עבור אות דבור נקי

Fig. (6.9): Cepstral residual received in PCR model for clean speech

שתי התופעות שתוארו בסעיפים לעיל נובעות ממבנהו של הספקטרום המנוקה ומצביעים על היתרון במימוש האלגוריתם תוך מעבר דרך האות הזמני. בסעיף הבא תואר דרך "לשיפור" מבנהו של הספקטרום המנוקה תוך שימוש מתאים בסף K_0 של אלגוריתם ההפחתה הספקטרלית בכדי להתגבר על התופעות שתוארו לעיל.

6.6 בחירת הסף באלגוריתם ההפחתה הספקטרלית

במקום ליצור החלקה על ידי המעבר דרך האות הזמני, דרך יקרה מבחינת סיבוכיות המערכת, נבדקה אפשרות להעלמת הבעיות תוך שימוש בסף המגדיר ספקטרום מנוקה בעל תחום דינמי נמוך ובעל קפיצות קטנות יותר בנקודות האי רציפות.

מתוך התבוננות בציורים (6.7) ו-(6.8) הועלה הרעיון להגביה את הסף של אלגוריתם ההפחתה הספקטרלית ככל האפשר. העלאת הסף, כאשר אלגוריתם ההפחתה הספקטרלית משמש ככלי לניקוי דיבור בלבד, אינה יעילה מכיוון שפעולה זו מגדילה את כמות הרעש הנותרת אחרי תהליך הניקוי: אולם כאשר הניקוי מתבצע כפעולה מקדימה עבור אלגוריתם ה-LPC השפעת הגדלת הסף נמוכה מאוד על כמות רעש הרקע הנותר.

הגבהת הסף בצורה ישירה אינה יעילה מכיוון שמסתכנים באיבוד מידע ספקטרלי עבור קטעים בעלי אנרגיה נמוכה. הפתרון הוא לבחור בסף המשתנה מקטע לקטע ואשר יהיה גבוה ככל האפשר עבור אותו קטע. שתי שיטות לבחירת סף כזה נבדקו:

1. סף הנגזר מהעוטפת הספקטרלית המתאימה למסנן ה-LPC של קטע האנליזה הקודם.
2. סף הנגזר מהממוצע של הספקטרום הרועש בקטע האנליזה הנוכחי.

6.6.1 סף הנגזר מהעוטפת הספקטרלית של הקטע הקודם.

בשיטה זו הוחלף הסף K_0 המופיע באלגוריתם ההפחתה הספקטרלית, ראה נוסחה (4.12), בסף המשתנה מקטע אנליזה אחד למשנהו ומתא תדר אחד למשנהו. את מחליף הספקטרום $K_m(w)$ המוגדר להלן:

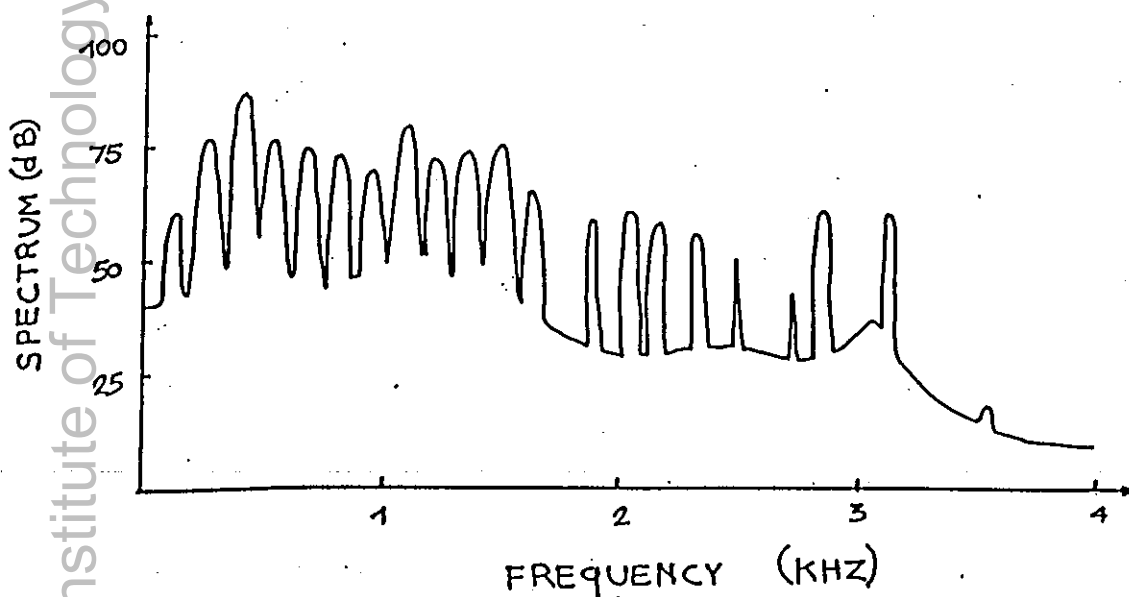
$$K_m(w) = H_{m-1}(w) / 100 \quad (6.2)$$

כאשר: m משמש כמונה קטעים
 $H_{m-1}(w)$ הוא הספקטרום המתאים למסנן LPC של קטע האנליזה הקודם

המוטיבציה לשליטה זו היא ההנחה שמבנה הספקטרום בין קטעי דיבור סמוכים אשר חופפים בחצי קטע אינו משתנה בצורה דרסטית.

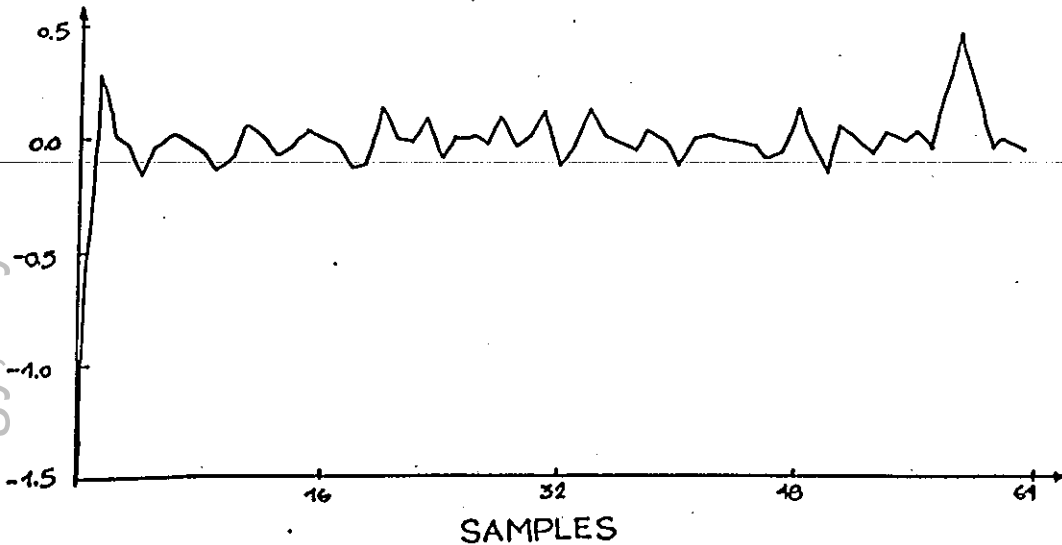
בסלימולציות של שיטה זו הסתבר שאיכות הדיבור של מודל ה-LPC משתפרת בנוסף התברר שמקדמי השארית הקפסטרלית המקודדים במודל PCR חזרו למצב שבו הם משפרים את איכות המערכת.

ציור (6.10) מתאר את הספקטרום המנוקה שנוצר בשיטה זו. ציור (6.11) מתאר את אות השארית הקפסטרלית בשיטה זו.



ציור (6.10): ספקטרום מנוקה על ידי אלגוריתם ה-SPS כאשר הסף K_0 נגזר מהעוטפת הספקטרלית של הקטע הקודם כפי שזו מתוארת על ידי מסנן ה-LPC.

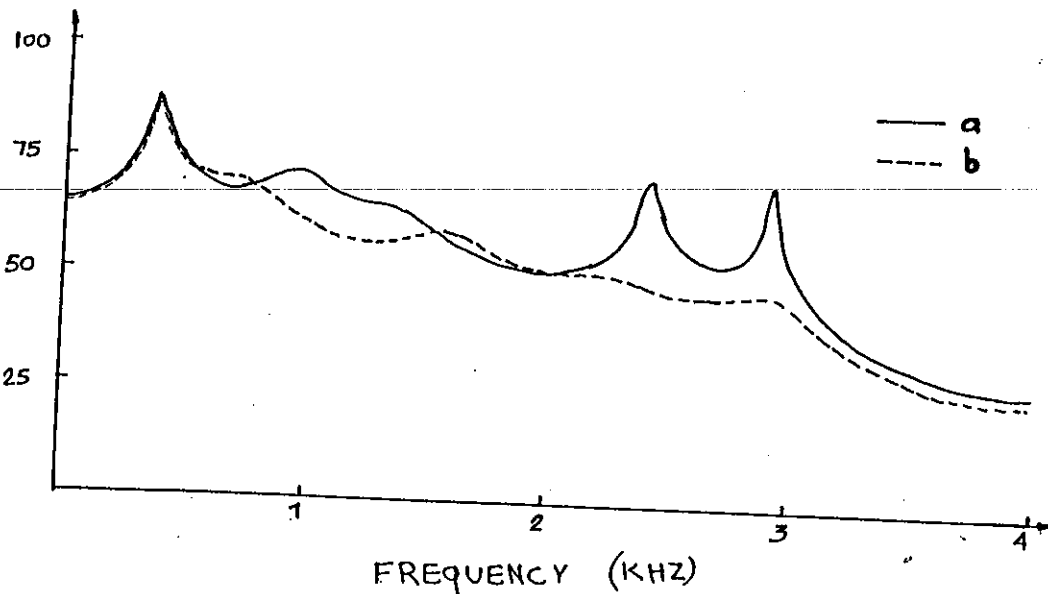
Fig. (6.10): Cleaned spectrum by SPS algorithm, the threshold is related to the spectral envelope as this is estimated by filter.



ציור (6.11): השארית הקפטטרלית המתקבלת במקדד PCR כאשר מופעל אלגוריתם ה- SPS עם הסף K_0 אשר נגזר מהעוטפת הספקטרלית של הקטע הקודם כפי שזו מתוארת על ידי מסנן ה-LPC.

Fig. (6.11): Cepstral residual that is received in PCR Vocoder using SPS algorithm, with the threshold related to the spectral envelope as this is estimated by filter.

מעיון בציורים (6.10) ו-(6.11) מסתבר ששיטה זו פתרה את בעיית התחום הדינמי הגדול של הספקטרום המנוקה ואת בעיית "השתוללותו" של אות השארית הקפטטרלית. אולם נוצרה בעיה חדשה, אות הדיבור המשוחזר נשמע מעוות. התכונות בציור (6.12) מסבירה את הסיבה לעיוות זה. הסיבה היא היוצרות של שיאי תהודה מלאכותיים בעיקר בתדרים גבוהים שאינם שיכים לאות הדיבור באותו קטע אנליזה.



ציור (6.12): תגובת התדר של מסנן LPC: a. כפי שהתקבלה אחרי הפעלת SPS עם סף הקשור לעוטפת הספקטרלית של הקטע הקודם. b. המתאים לאות הנקי.

Fig. (6.12): Spectrum of the LPC model: a. after using SPS with threshold related to spectral envelope of the previous frame. b. for clean speech.

בבדיקת מקורם של אותם שיאי תהודה מיותרים מסתבר שאלו שיאי ספקטרום הנגררים מספקטרום LPC של קטעים קודמים ואשר המקדם $1/100$ אינו מצליח לבטלם. תופעה זו נובעת מתכונתו של אלגוריתם ה-LPC להדגיש לחדר ולהגבילה שיאים חדים. בכדי להעלים שיאים אילו נראה שיש להנמיך את הסף על ידי הקטנת המקדם $1/100$ דבר המחזיר את הבעיות הקודמות.

6.6.2 סף המתחם לממוצע הספקטרום בכל קטע.

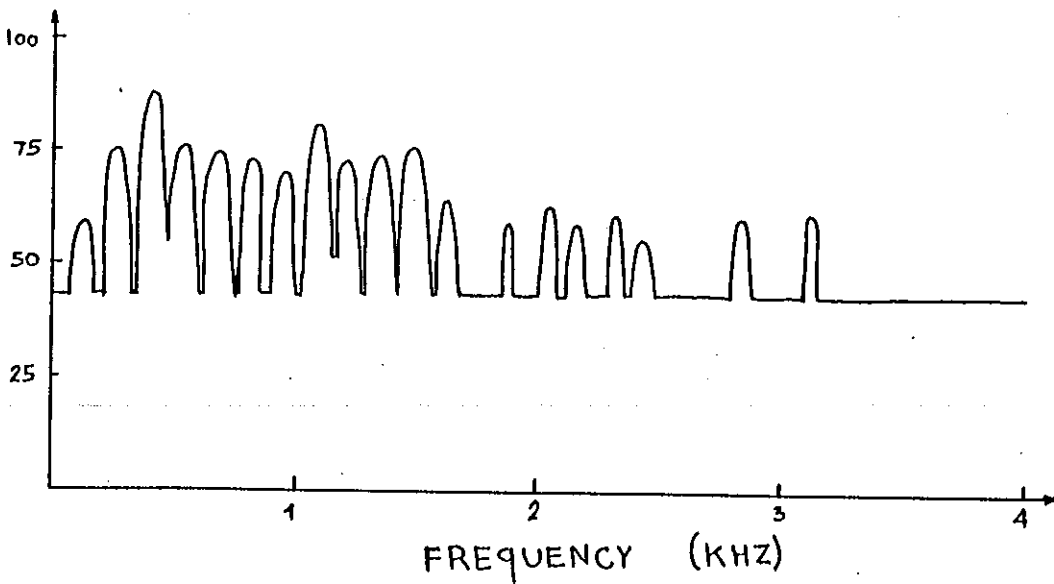
בשיטה זו הוחלף הסף K_0 בסף K_m המשתנה מקטע לקטע אך קבוע על פני כל התדרים בכל קטע. סף זה נגזר מהממוצע של הספקטרום בקטע כאשר זה מבוטא לוגריתמית ב-dB:

$$T=0.01 \left(\sum_{i=0}^{255} [S(i)] \right)_{dB} \quad (6.3)a$$

$$K_m = 10^{0.1T} \quad (6.3)b$$

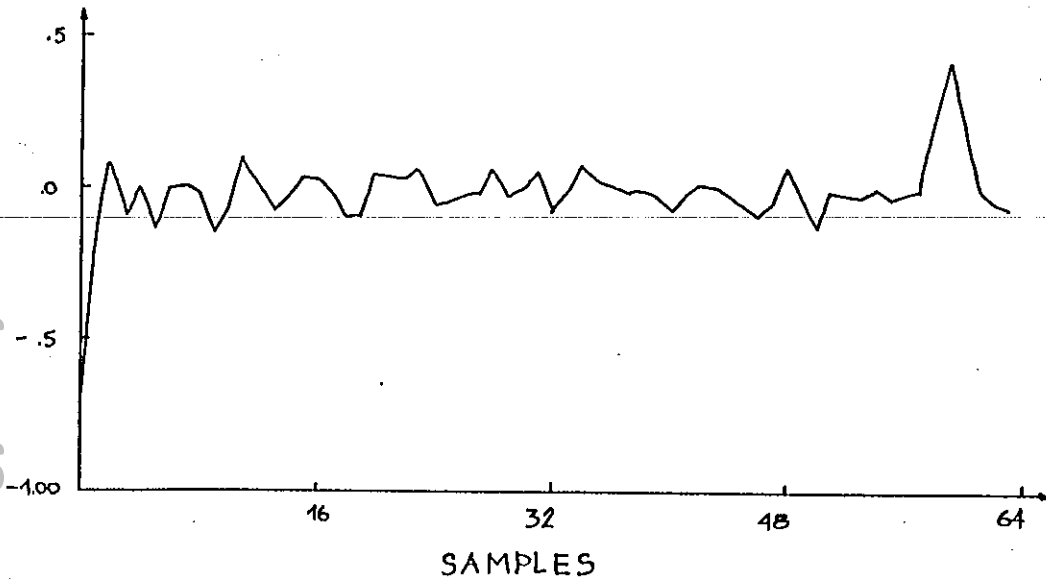
כאשר $S(w)$ הוא הספקטרום של האות הרועש, T משתנה זמני.

ציור (6.13) מתאר את הספקטרום המנוקה בשיטה זו. ציור (6.14) מתאר את השארית הקפסטרלית הנוצרת בשיטה זו.



ציור (6.13): ספקטרום מנוקה על ידי אלגוריתם ה-SPS כאשר הסף K_0 נגזר מהמוצע של הספקטרום בקטע הנוכחי ב- dB.

Fig. (6.13): Cleaned spectrum by SPS algorithm, the threshold is related to the averaged of the spectrum in db, in the same frame.



ציור (6.14): השארית הקפסטרלית המתקבלת במקדד PCR כאשר מופעל אלגוריתם ה- SPS עם הסף K_0 אשר נגזר מהמוצע של הספקטרום של הקטע הנוכחי.

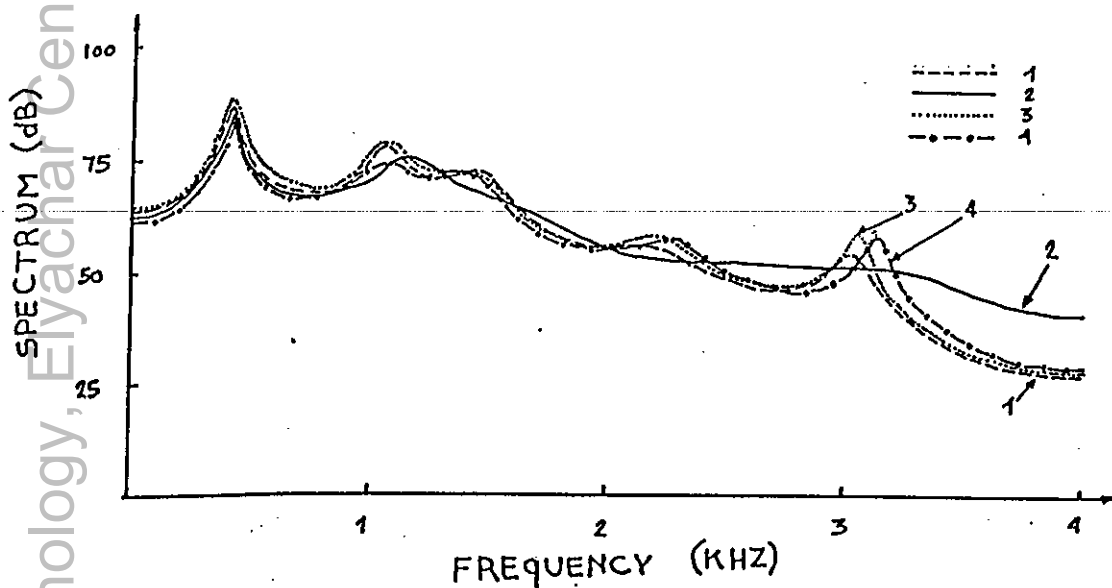
Fig. (6.14): Cepstral residual that is received in PCR Vocoder using SPS algorithm, with the threshold related to the averaged of the spectrum in the same frame.

ציור (6.14)

ציור (6.13) מתאר את הספקטרום המנוקה בשיטה זו ואת התחום הדינמי המוקטן. ציור (6.14) מתאר את אות השארית הקפסטרלית עבור סף זה.

שימוש בשיטה זו העלים כמעט לגמרי את התופעות שהוזכרו בסעיפים 4 ו-5 בפרק הנוכחי. שיטה זו מאפשרת שימוש במקדד PCR על הספקטרום המנוקה ללא מעבר דרך האות הזמני עם כל היתרונות שבכך.

ציור (6.15) מתאר את הספקטרום של מסנני LPC שהתקבלו עבור כל שיטת הסף שהוזכרו לעיל, בהשוואה לספקטרום ה- LPC על אות דיבור נקי ועל האות הרועש. כל התוצאות מוצגות עבור דיבור רועש בעל $SNR=10dB$ ו- LPC בעל 10 מקדמים.



ציוור (6.15): תגובת התדר של מסנן ה-LPC עבור: 1. אות נקי. 2. K_0 נגזר מהמוצע של הספקטרום בקטע. 3. K_0 נגזר מהעוטפת הספקטרלית של הקטע הקודם. 4. K_0 קבוע.

Fig. (6.15): Spectrum of the LPC model: 1. clean speech. 2. K_0 related to the averaged of the spectrum. 3. K_0 related to the spectral envelope of the previous frame. 4. constant K_0 .

מהתבוננות בציוור (6.15) מסתבר שהשיטה שנבחרה (סף משתנה לפי נוסחה (6.3)), נוטה להרחיב את השיאים בתדרים הגבוהים ביחס לעוטפת האות הנקי, בעוד שהשיטות האחרות נוטות להדגיש שיאים אלה ואף ליצור שיאים חדשים שאינם מופיעים באות הנקי. מסקנה המתלווה לביתוח זה היא שתוספת פורמנטים לספקטרום הדליבור וחידודם (בתחום התדרים הגבוה) גורמת לעיוות גדול יותר בדליבור המשוחזר מאשר מריחת פורמנטים קימים.

6.7 תוצאות הסימולציות

מן הראוי היה בסעיף זה להציג את תוצאות הסימולציה של המקדד בעזרת שני המדדים שהוזכרו לאורך חיבור זה, כמיצגים את ביצועי המקדד. שני המדדים המדוברים הם מדד אי ההתאמה הספקטרלית E_c שהוזכר בפרק השלישי נוסחאות (3.22) ו-(3.23) ובדיקות שמיעה סוביקטיביות.

כאשר נבדקו תוצאות הסימולציה של המקדד ביחס למדד אי ההתאמה הספקטרלית, E_c , הסתבר שמדד זה אינו מיצג נכונה את איכות הדיבור במקרה זה של קידוד האות הרועש. בעוד שבשיטה הנבחרת איכות הדיבור היתה המשופרת ביותר, הרי המדד הנמוך ביותר היה כאשר אין מבוצע אלגוריתם הניקוי, כלומר כאשר אלגוריתם ה-PCR מופעל על האות הרועש. הסיבה לתופעה זו היא העובדה שמדד אי ההתאמה הספקטרלית משקלל את אי ההתאמה הספקטרלית בכל התדרים באותה מידה, בעוד שההתאמה הספקטרלית באזורי הפורמנטים, חשובה לאיכות הדיבור יותר מאשר תחומי תדר אחרים. אלגוריתם ההפחתה הספקטרלית משערך טוב למדי את הפורמנטים אך לוקה בשיערוך אזורי התדר בעלי הספק נמוך, לכן מדד ההתאמה הוא גרוע יחסית לזה עבור האות הרועש אולם איכות הדיבור המשוחזר טובה יותר.

לאור עובדות אלו מסתפקים בהשוואה בין השיטות השונות על ידי האזנה לדיבור המשוחזר בלבד אשר מראה שהשיטה שנבחרה היא אומנם הטובה בין אלו שנבדקו.

הערה: כל תוכניות המחשב העוסקות בסימולציות שהוזכרו לעיל נמצאים וניתן לקבלם במעבדה לעבודות אותות, בפקולטה לאלקטרוניקה קרית הטכניון חיפה.

פרק 7: סיכום ומסקנות.

במסגרת חיבור זה הוצעו ונבחנו שני מקדדים לאות דיבור רועש. מקדדים אלו נבדקו בסימולציות על גבי מחשב, ביחסי אות לרעש שונים כאשר המטרה היא להגיע לאיכות דיבור סבירה עבור יחס אות לרעש של 10dB.

המערכת הראשונה מתבססת על אלגוריתם של Yegnanarayana [27] להפרדה בין קטבים ואפסים במישור נגזרת ספקטרום הפאזה (DPS) ושימוש בקטבים למימוש מסנן הסינטזה. כאשר נבדקה המערכת הנ"ל בבדיקות שמיעה סוביקטיבית הסתבר שבצועיה גרועים והדיבור המשוחרר מעוות לחלוטין, למרות שהמערכת הוכיחה תוצאות סבירות בכל הקשור למדד אי ההתאמה הספקטרלי, עבור יחסי אות לרעש של 10dB ומטה ובצועיה נראו כעולים על מקדד חיזוי לינארי מקובל.

במערכת השניה מומש מקדד חיזוי לינארי עם קידוד השארית הקפסטרלית עבור אות דיבור רועש, כאשר מקדים אותו תהליך של ניקוי מרעש המבוסס על אלגוריתם ההפחתה הספקטרלית. אלגוריתם זה שונה במקצת מהמקובל (ראה פרק 4) בכך שהספ בו משתנה אדפטיבית מקטע אנליזה אחד למישנהו, וזאת כדי להקטין את התחום הדינמי של הספקטרום המנוקה ואת הקפיצות בנקודות אי הרציפות וכך בעקיפין, לקבל שארית קפסטרלית הנוחה לקידוד. תוך כדי מימוש המקדד התגלו בעיות אי יציבות של מסנן ה-LPC והוצע פיתרון לפיו מקפיעים את סדר המודל לסדר הגדול ביותר שעדיין יציב וזאת בשימוש בתכונות אלגוריתם Durbin במסגרת אנליזה ה-LPC. ביצועי מקדד זה נבדקו בסימולציה בעזרת מחשב עם קידוד השארית (ללא בחינת בעית הקוונטיזציה שלהם). תוצאות הסימולציה מראות שיפור בביצועי מקדד זה על פני מקדד LPC רגיל. גם בעת בחינת מקדד זה הסתבר שמדד אי ההתאמה הספקטרלית אינו מיצג נכונה את ביצועי המקדד מבחינת איכות הדבור המשוחרר. המסקנה כמובן היא שבבחינת ביצועי מקדד דיבור יש להשתדל לבצע בדיקות שמיעה.

לקראת סיום עבודת המחקר שתוארה בחיבור זה, פורסמה עבודת דוקטורט שנעשתה בפקולטה להנדסת חשמל בטכניון [28], אשר מציגה שיטה לניקוי דיבור רועש תוך שימוש במשעך השגיאה הריבועית הממוצעת המינימלית של האמפליטודה הספקטרלית לזמן קצר. בעבודה זו התקבל שהאלגוריתם הנ"ל עולה בביצועיו על אלגוריתם ההפחתה הספקטרלית בכל הקשור לשיחזור אות הדיבור הזמני. שיטה זו

נראית אטרקטיבית ויש מקום לבדוק את האפשרות להשתמש בה במקום אלגוריתם ה-
SPS במקדד הדיבור שתואר לעיל.

נספח א: חישוב נגזרת ספקטרום הפאזה (DPS)
עבור מסנן מסדר ראשון ומסנן מסדר שני.

A.1 מבוא.

בנספח זה תובא הוכחה לנוסחאות (5.4) ו (5.5) המציגות את נגזרת ספקטרום הפאזה (DPS) עבור מסננים מסדר ראשון ומסדר שני. כדי להקל על ההוכחה תובא זו עבור מסנן אפסים בלבד כאשר יש לזכור שעבור מסנן בעל קטבים השוני הוא רק בסימן של התוצאה.

A.2 DPS עבור מסנן מסדר ראשון.

יהי $H(w)$ חגובת החדר של מסנן אפסים בלבד מסדר ראשון המוגדר להלן:

$$H(w) = 1 + jw/\delta \quad (A.1)$$

ותהי $\theta(w)$ הפאזה של $H(w)$.
אזי מ (A.1) מתקבל:

$$\theta(w) = \text{tg}^{-1}(w/\delta) \quad (A.2)$$

ובשימוש בנוסחא:

$$\frac{d}{dx} \text{tg}^{-1}(f(x)) = \frac{1}{1+(f(x))^2} \cdot \frac{df(x)}{dx} \quad (A.3)$$

מתקבל:

$$\theta'(w) = \frac{d}{dx} (\text{tg}^{-1}(w/\delta)) = \frac{1}{1+(w/\delta)^2} \cdot \frac{1}{\delta} \quad (A.4)$$

כלומר:

$$\frac{d\theta(w)}{d(w)} = \theta'(w) = \frac{\delta}{w^2 + \delta^2} \quad (\text{A.5})$$

DPS A.3 - עבודת מסנן מסדר שני.

יהי $H(w)$ מסנן אפסים בלבד מסדר שני המוגדר להלן:

$$H(w) = w_0^2 + 2ww_0qj - w^2 \quad (\text{A.6})$$

אזי מנוסחא (A.6) מתקבל:

$$\theta'(w) = \text{tg}^{-1} \left(\frac{2ww_0q}{w_0^2 - w^2} \right) \quad (\text{A.7})$$

ושימוש בנוסחא (A.3) מתקבל:

$$\theta''(w) = \frac{2w_0q(w_0^2 + w^2)}{(w_0^2 - w^2)^2 + 2w^2(w_0q)^2} \quad (\text{A.8})$$

כלומר:

$$\theta''(w) = \frac{2(w^2 + \alpha^2 + \beta^2)}{(\alpha^2 + \beta^2 - w^2)^2 + 4\alpha^2 w^2} \quad (\text{A.9})$$

כאשר:

$$w_0^2 = \alpha^2 + \beta^2 \quad (\text{A.10}a)$$

$$\alpha = qw_0 \quad (\text{A.10}b)$$

REFERENCES

- [1] J.L.Flanagan et al, "Speech Coding", IEEE Trans. on Communication, Vol. COM-27, No. 4, April 1979.
- [2] J.Makhoul "Linear Prediction: A Tutorial Review", Proceeding of the IEEE, Vol. 63, No. 8, April 1975.
- [3] Alan V.Oppenheim and Ronald W.Schafer, "Digital Signal Processing" Prentice-Hall, Inc. Engelwood Cliffs, New Jersey.
- [4] J.D.Markel "The SIFT algorithm for fundamental frequency estimators" IEEE Trans. Audio Electroacoust., Vol.AU-20, Dec. 1972.
- [5] A.M.Noll "Cepstrum Pitch determination", J. Acoust. Soc. Amer. Vol. 41, Feb. 1967.
- [6] L.R.Rabiner et al, "A comparative performance study of several Pitch detection algorithm", IEEE Trans. Acoust. Speech, Signal processing, Vol. ASSP-27, No.5 pp 399-418, October 1976.
- [7] J.Makhoul et al, "A mixed-source model for speech compression and synthesis", Proc. IEEE Int. Conf. ASSP, pp 163-166, 1979
- [8] C.K.Un and D.T.Magill "The residual excited linear prediction Vocoder with transmission rate below 9.6 Kbit/s" IEEE Trans. Communication, Vol. COM-23, Dec. 1975/
- [9] M.D.Dankenberg and D.Y.Wong "Development of a 4.8-9.6 Kbps RELP Vocoder" Proc. IEEE Int. Conf. ASSP, pp 554-557, 1979.
- [10] L.L.Burg and R.Yariagadda "An efficient coding of the prediction residual" Proc. IEEE Int. Conf. ASSP, pp 538-541, 1979.
- [11] J.Makhoul and M.Berouti "High-frequency regeneration in speech coding system" Proc. IEEE Int. Conf. ASSP, pp 428-431, 1979.
- [12] D.Y.Wong "On understanding the quality of LPC speech" Proc. IEEE Int. Conf. ASSP, pp 725-729, 1980.
- [13] S.T.Kay "The effect of noise on autoregressive spectral estimator" IEEE Trans. Acoust. Speech, Signal processing, Vol. ASSP-27 No. 3, June 1980.

- [14] B.S.Atal and N.David "On synthesizing natural sounding speech by linear prediction", Proc. IEEE Int. Conf. ASSP, pp. 44-47 1979.
- [15] D.Malah "Cepstral residual vocoder for improved quality speech transmission at 4.8 kbps", Proc. IEEE Int. Conf. ASSP pp. 622-625 1982.
- [16] D.Malah "Efficient spectral matching of the LPC residual signal", Proc. IEEE Int. Conf. ASSP, pp. 1288-1291 1981
- [17] L.R.Rabiner et al, "LPC Prediction error analysis of its variation with position of the analysis frame", IEEE Trans. Acoustics, Speech and Signal Proc. Vol. ASSP-25, No.5 pp 434-442, October 1977.
- [18] S.Mitra and C.R.Davis "Improvements in the classical model for better speech quality", Proc. IEEE Int. Conf. ASSP pp. 23-26 1980
- [19] J.D.Markel and A.H.Gray "Linear Prediction of speech", New York:Springer-verlag 1976.
- [20] S.F.Boll "Suppression of noise in speech using the SABAR method", Proc. IEEE Int. Conf. ASSP pp. 606-609 1978.
- [21] - "Suppression of acoustic noise in speech using spectral subtraction", IEEE Trans. on Acous., Speech, Signal Processing Vol. ASSP-27 No. 2 April 1979.
- [22] - "A Spectral Subtraction algorithm for suppression of acoustic noise in speech", Proc. IEEE Int. Conf. ASSP pp. 200-203 1979.
- [23] J.S.Lim and A.V.Oppenheim "Enhancement and bandwidth compression of noisy speech", Proceeding of the IEEE, Vol. 67 No. 12, Dec. 1979.
- [24] R.D.Preuss "Low complexity robust LPC of speech signal", Report No. MTP212 MITRE Corp. Bedford, Mass. Feb. 1982.
- [25] S.M.Kay "Noise Compensation for autoregressive spectral estimates", IEEE Trans. Acoust. Speech, Signal Processing Vol. ASSP-28 No. 3, June 1980.

- [26] S.M.Kay and S.L.Marple "Spectrum analysis - a modern perspective", Proceeding of the IEEE, Vol. 69 No. 11, Nov. 1981.
- [27] B.Yanganrayana "Speech analysis by pole-zero decomposition of the short time spectrum", Signal Processing Vol. 3 No. 1 January 1981.
- [28] Y.Eprhaim "Enhancement of noisy speech", D.Sc thesis, Technion Israel Institute of technology, June 1984.

NOISY SPEECH CODING USING LINEAR
PREDICTION AND CEPSTRAL RESIDUAL

Research Thesis

Submitted in partial fulfillment of the requirements

for the degree of Master of Science

in

Electrical Engineering

By

KOPOLOVICH ZVI

Submitted to the Senate of the Technion-Israel Institute of Technology

Tishray 5745

Haifa

September ,1984

This research was carried out in the Signal Processing
Laboratory of the Faculty of Electrical Engineering
under the supervision of Prof. David Malah.

I am indebted to Prof. David Malah for his most
helpful guidance and for his patience throughout
the various stages of this research.

I would like to thank Mr. Yoram Or-Chen, the
laboratory engineer, for his assistance and
sympathetic attitude.

I am also grateful to Ms. Zippy Portnoy for her
support during the difficult moments experienced
with the computer.

CONTENTS

	<u>Page</u>
ABSTRACT	1
SIMBOLS LIST	3
CHAPTER 1 : Introduction	7
CHAPTER 2 : Linear Predictive Coding of speech	11
2.1 Introduction	11
2.2 Speech production model	13
2.3 Algorithm analysis and parameters computation	15
2.3.1 Prediction parameters estimation	15
2.3.2 V/UV decision and Pitch estimation	24
2.4 Residual signal in LPC	31
2.5 Shortcomings of LP coding of speech	32
2.6 LP coding of speech in white noise	33
CHAPTER 3 : LPC with cepstral residual coding	34
3.1 Introduction	34
3.2 The cepstrum	36
3.2.1 What is the cestrum ?	36
3.2.2 The properties of the cepstrum and the complex cepstrum	36
3.3 Analysis model	39
3.3.1 Pre-emphysys and windowing	39
3.3.2 LPC parameter evaluation and cepstrl residual computation	40
3.3.3 Averege cepstral residual computation	42
3.3.4 Zeros computation	43
3.3.5 Remaining cepstral residual computation	44
3.3.6 V/UV decision and Pitch estimation	45

CONTENTS (CONTINUED)

	<u>Page</u>
3.3.7 The complete system	45
3.4 Synthesis system	48
3.5 Quality measure E_c	50
3.6 Cepstral residual coding system simulating ...	52
3.6.1 Simulations results	52
3.6.2 Conclusions	56
CHAPTER 4 : LPC of speech in white noise	58
4.1 Introduction	58
4.2 The noisy speech signal	59
4.2.1 Noisy speech signal model	59
4.2.2 The influence of the noise on the LPC model ..	61
4.3 Way to reduce the influence of the noise on the LPC vocoder	62
4.3.1 Increase in the model order	62
4.3.2 Spectral Subtraction	63
4.3.3 Autocorrelation Subtraction	65
4.3.4 High order Yule Walker Equations	66
4.3.5 Pole-Zero estimation	67
4.4 Summary	68
CHAPTER 5 : Speech coder with pole-zero decomposition in the Derivative of Phase spectrum domain (DPS)	69
5.1 Introduction	69
5.2 Pole-zero decomposition in the DPS domain	70
5.2.1 The Problem ?	70
5.2.2 DPS properties	70
5.2.3 The connection between DPS and the cepstrum ..	72
5.2.4 Pole-zero decomposition algorithm	74
5.3 "YANG" vocoder with clean speech	76

CONTENTS (CONTINUED)

	<u>Page</u>
5.3.1 Analysis system ("YANG")	76
5.3.2 Synthesis system ("YANG")	77
5.3.3 Simulations results	78
5.3.4 Conclusions	83
5.4 "YANG" vocoder with noisy speech	84
5.5 Summary	87
 CHAPTER 6 : Noisy speech coder with Spectral Subtraction and cepstral residual coding	 88
6.1 Introduction	88
6.2 Analysis system	89
6.2.1 General structure of the system	89
6.2.2 Analysis system parts	90
6.3 Instability problem of the LPC filter	93
6.4 LPC algorithm directly on the clean spectrum .	96
6.5 The cepstral residual structure	98
6.6 Choosing the threshold in the in the spectral subtraction algorithm	 100
6.6.1 Threshold that is related to the spectral envelope of the previous analysis frame	 100
6.6.2 Threshold that is related to the average of the spectrum in the frame	 103
6.7 Simulations results	107
 CHAPTER 7 : Summary and conclusions	 108
 APPENDIX A The derivative of Phase Spectrum for first order filter and for second order filter	 110
A.1 Introduction	110
A.2 DPS for first order filter	110

CONTENTS (CONTINUED)

	<u>Page</u>
A.3 DPS for second order filter	110
REFERENCES LIST	112

- | -

ABSTRACT

Linear Predictive Coding (LPC) is an efficient and well known method for low bit rate speech coding (e.g. 2400 bps). This method is based on the classic speech production model which is a Linear time-varying filter, excited by one of the following input signals: periodic impulses with varying period according to the pitch or white noise. The filter is assumed to be an all-pole filter. This assumption was found to give good results and is particularly suitable for applying linear prediction.

The LPC synthesized speech is not natural sounding. To regain its naturalness, one can use the residual signal (which would result in exact reconstruction but requires many bits for encoding), instead or as an addition to the classical BUZZ-HISS excitation. One efficient way of using this residual is to encode its spectral envelope in the cepstral domain, and to use the cepstral coefficients to shape the spectrum of the BUZZ-HISS excitation signal before it enters the synthesis filter.

When using the LPC algorithm on noisy speech, a drastic degradation in synthesized speech quality appears. This degradation is due to the fact that the noisy speech model is no longer an all-pole model, thus the LPC algorithm does not estimate its spectral envelope sufficiently well. Therefore, to apply the LPC algorithm on noisy speech the effect of noise must be reduced.

This work describes two methods for applying LPC to noisy speech.

The first method is based on the assumption that noisy speech has a pole-zero model, where the poles are due to the speech and are the same as those for the clean speech, whereas the zeros are due to the noise. Pole-zero decomposition is done using the Derivative of the

Phase Spectrum (DPS). The poles coefficients are then used in the synthesis filter. In simulations of this system it was found that the decomposition algorithm does not provide a good estimate of the spectral envelope and in particular of the formants.

The second method uses LPC with cepstral residual coding preceded by Spectral Subtraction (SPS) which is a well known speech enhancement method. In simulating this system, two major problems were encountered: the first was instability of the synthesis filter, to which a solution is presented in this work. The second problem was that the usual SPS algorithm method was found to be unsuitable for the system, and a modification of the method was required.

To test the performance of the systems described above, a spectral matching measure was used. This measure was found to correlate well with speech quality while dealing with clean speech, but inadequate when dealing with noisy speech. Due to this fact, subjective informal listening tests were used to examine the quality of the two speech coding systems. Using these tests, it was recognized that the first method produces very distorted speech, although the spectral matching measure indicated that this method is efficient. Using the same listening tests, the second coding system was found to provide an improvement in speech quality, over the ordinary LPC Vocoder.