



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.*

מימוש מערכת לזיהוי דיבור  
באמצעות מיני-מחשב

חבור על פרויקט

לשם מילוי חלקי של הדרישות לקבלת התואר  
מגיסטר למדעים  
בהנדסת חשמל

מ א ת

יצחק אנגל

הוגש לסנט הטכניון - מכון טכנולוגי לישראל  
טבת תשל"ז  
ינואר 1977



המחקר נעשה בהנחיית די"ר דוד מלאך  
בפקולטה להנדסת חשמל, במעבדה לעיבוד אותות,  
הטכניון-מכון טכנולוגי לישראל.

תודתי העמוקה נתונה לדי"ר דוד מלאך  
על הנחייתו המצויינת ועל עזרתו  
המסורה במהלך המחקר.

תוכן הענייניםעמוד

1	תקציר
3	רשימת סמלים וקיצורים
6	פרק 1 - מבוא
10	פרק 2 - השלבים בפעולת מערכת לזיהוי מילים בודדות
10	2.1 שלב עיבוד הקדם
11	2.2 שלב מיצוי התכונות
13	2.2.1 יצוגים פרמטריים הנגזרים מהספקטרום של אות הדיבור
14	2.2.2 יצוגים פרמטריים הנגזרים מהתאור הזמני של אות הדיבור
20	2.3 שלב מיון ההגייה
21	2.3.1 שיטות מיון
23	2.3.2 התאמת צירי הזמן של ההגיות
27	2.3.3 חישוב מידת הדמיון בין שתי הגיות
31	פרק 3 - שיקולים בפיתוח מערכת לזיהוי מילים בודדות הממומשת באמצעות מיני-מחשב
31	3.1 מבוא
32	3.2 השיקולים בבחירת מקדמי הקורלציה החלקית כיצוג פרמטרי עבור ההגייה
39	3.3 שיקולים בפיתוח תהליך המיון
41	פרק 4 - דחיסת ההגייה בזמן
41	4.1 תאור האלגוריתם
47	4.2 שיקולים בבחירת פרמטרים
56	פרק 5 - חיזוי לינארי של אותות דיבור
56	5.1 מודל המסנן כולו קטבים
59	5.2 שיטת האוטוקורלציה (השיטה הסטציונרית)
60	5.3 שיטת הקוואריאנס (השיטה הלא סטציונרית)
61	5.4 השוואה בין שיטת האוטוקורלציה לבין שיטת הקוואריאנס
63	5.5 הקשר בין ה-LPC לבין מקדמי ה-PARCOR
67	פרק 6 - פרטי מימוש המערכת הכוללת דחיסה בזמן
67	6.1 דגימת ההגייה וקביעת גבולותיה

תוכן הענינים (המשך)

<u>עמוד</u>	
70	6.2 מימוש האלגוריתם לדחיסה בזמן של ההגייה
75	6.3 חישוב מקדמי הקורלציה החלקית
78	6.4 תהליך המיון
80	פרק 7 - פתוח מערכת השוואתית ללא דחיסה
80	7.1 מבוא
80	7.2 התאמת צירי הזמן ומיון באמצעות תכנות דינמי
86	7.3 הפעלת תכנות דינמי במערכת ההשוואתית
88	פרק 8 - נסויים בזיהוי הגיות ותוצאותיהם
88	8.1 הכנת שני בסיסי נתונים
89	8.2 ארגון הניסויים והתוצאות
92	פרק 9 - סכום ומסקנות
92	9.1 השוואה בין שתי המערכות מבחינת כמות החישובים
93	9.2 השוואה בין שתי המערכות מבחינת תוצאות הזיהוי
94	9.3 מסקנות
97	נספח א' - תכנית הזיהוי עבור המערכת הכוללת דחיסה
114	נספח ב' - התכניות עבור המערכת ההשוואתית ללא דחיסה
129	רשימת מקורות
I	תקציר (אנגלית)

## ת ק צ י ר

בחיבור זה מתואר מחקר לפיתוח מערכת לזיהוי מילים בודדות ומימושה בתכנה באמצעות מיני מחשב NOVA 2.

במערכת הממומשת בתכנה בלבד, תהליך הזיהוי נמשך זמן רב יותר מאשר במערכת הנעזרת בחומרה אנלוגית ו/או ספרתית.

לכן, כאשר נדרש ממערכת לבצע זיהוי נכון בזמן קצר ככל האפשר, כפי שנדרש מהמערכת שפותחה, הגבלת המימוש לתכנה בלבד מכבידה על קיום הדרישה.

בעבודה זו מוצעת גישה חדשה שמטרתה לקצר באופן ניכר את זמן תהליך הזיהוי במערכות הממומשות בתכנה בלבד.

לפי גישה זו, משתמשים באלגוריתם יעיל לשינוי ציר הזמן של אותות דיבור שפותח לאחרונה, לצורך ביצוע דחיסה בזמן של ההגייה לאורך שהינו אחיד עבור כל ההגיות, לפני תהליך מיצוי הפרמטרים. עקב דחיסה זו שאינה משנה את תחום התדרים של אות הדיבור, מכילה כל הגייה דחוסה מספר דגימות זהה, הקטן יותר ממספר הדגימות שמכילה הגייה טבעית. אם מימד וקטור הפרמטרים המייצג את ההגייה, יחסי לארכה, אזי לאחר הדחיסה יקטן המימד שלו. מכיוון שהזמן הנחסך כתוצאה מחישוב וקטור פרמטרים קצר יותר הוא גדול בהרבה מהזמן הנדרש לצורך ביצוע הדחיסה באמצעות האלגוריתם הנ"ל, יתקבל עקב הדחיסה חסכון ניכר בזמן חישוב. גם בשלב המיון יתקצר זמן החישוב וזאת משתי סיבות. האחת, וקטורי הפרמטרים הם בעלי מימד קטן יותר מהמקובל, כך שפעולות המיון דורשות זמן קצר יותר. הסיבה השנייה - עקב תהליך הדחיסה, וקטורי הייחוס ווקטור ההגייה הנבדקת יהיו בעלי אורך אחיד. לכן, ניתן לחשב באופן ישיר את מידת הדמיון בין וקטור ההגייה הנבדקת לבין כל אחד מוקטורי הייחוס, ואין ההכרח, הקיים כאשר הוקטורים המשווים הם בעלי אורך שווה, לבצע תאום בין צירי הזמן שלהם על מנת שאפשר יהיה לחשב את מידת הדמיון ביניהם. תהליך התאום הזה דורש זמן חישוב רב יחסית לחישוב הישיר הנ"ל, במיוחד כאשר מממשים אותו באמצעות תכנות דינמי בכדי להבטיח הצלחה מכסימלית בזיהוי.

כיצוג פרמטרי עבור ההגייה נבחרו מקדמי הקורלציה החלקית המתקבלים עקב תהליך חיזוי לינארי של אות הדיבור. מקדמים אלה הוכיחו את יעילותם בזיהוי

הגיות במערכת קיימת, ונוסף לכך זמן הזיהוי עבורם קטן יותר מזמן הזיהוי הנדרש כאשר משתמשים, במערכת הממומשת בתכנה, ביצוגים פרמטריים אחרים שהוכיחו הצלחה גבוהה בזיהוי.

בתהליך המיון מחושבת מידת הדמיון בין וקטורי הפרמטרים המשווים, באמצעות פונקציית המרחק של צ'ביצ'יף. זו שיטה חסכונית ביותר מבחינת זמן החישוב, ובניסויי זיהוי שנערכו במהלך פתוח המערכת, התברר שבחישוב מידת הדמיון בעזרת פונקציית מרחק אוקלידי לא התקבלו תוצאות טובות יותר מאשר עבור פונקציית המרחק של צ'ביצ'יף.

בכדי להעריך את יעילות הגישה החדשה - דחיסת ההגייה בזמן, פותחה מערכת נוספת לזיהוי מילים בודדות שאיננה כוללת דחיסה. מערכת זו זהה למערכת הכוללת דחיסה, פרט לכך שמכיוון שלא הופעלה דחיסה מתקבלים במערכת הנוספת וקטורי פרמטרים באורכים שונים, ויש צורך לתאם בין צירי הזמן של שני וקטורים מושווים. תאום זה מבוצע בעזרת תכנות דינמי כדי להבטיח, כאמור לעיל, הצלחה מכסימלית בזיהוי.

מבחינת כמות החישובים, המערכת ללא דחיסה דורשת לצורך זיהוי הגייה, פי 3 יותר חישובים מהמערכת הכוללת דחיסה. בכדי להשוות בין המערכות מבחינת הצלחה בזיהוי, נערכו בשתי המערכות ניסויי זיהוי של אותן ההגיות. המילון שנבחר עבור הניסויים הוא עשר הספרות העבריות מ- "אפס" עד "תשע". יש לציין שתהליך הזיהוי אינו תלוי במילון שנבחר, אולם הוא תלוי בדובר. חלק מההגיות נאמרו ע"י שלושה דוברים בסביבה שקטה, והחלק האחר - ע"י שניים מתוך השלושה, בסביבה רועשת שיחס האות לרעש שהתקיים בה היה 20 dB.

עבור הסביבה הרועשת, השיגו שתי המערכות שפותחו תוצאה זהה בזיהוי - הצלחה של 97.9% ב- 240 ניסויים. לעומת זאת, עבור הסביבה השקטה התקבלה עבור המערכת הכוללת דחיסה, הצלחה של 99.1% ב- 360 ניסויים. תוצאת זיהוי זו טובה יותר במקצת מאשר התוצאה שהתקבלה עבור המערכת ללא דחיסה - 98.3% ב- 360 ניסויים. כלומר, הדחיסה לא פגעה בהצלחת הזיהוי, למרות שהיו גם הגיות שנדחסו פי 3. ניתן להסביר זאת בכך שהאלגוריתם המבצע את הדחיסה מתבסס על שקלול מתאים של מספר קטעי זמן של ההגייה המקורית לצורך יצירת קטע זמן אחד של ההגייה הדחוסה, ואין הוא משמיט קטעי הגייה באופן שרירותי.

רשימת סמלים וקיצורים

- $a_k$  - מקדם חיזוי לינארי ה-  $k$ -י.
- $b_p(n)$  - השגיאה בחיזוי לינארי לאחור באמצעות  $p$  מקדמים.
- $C$  - גורם הדחיטה בזמן.
- $d(X,Y)$  - פונקצית המרחק בין הוקטורים  $X$  ו-  $Y$ .
- $e(n)$  - השגיאה בחיזוי לינארי (קדימה).
- $E_p$  - השגיאה הריבועית הכללית המינימלית עקב חיזוי לינארי באמצעות  $p$  מקדמים.
- $f$  - תדר.
- $f_p(n)$  - השגיאה בחיזוי לינארי קדימה באמצעות  $p$  מקדמים.
- $f_s$  - תדר הדגימה.
- $g(x,n)$  - איבר מהסדרה המתקבלת משקלול דגימות הדיבור בדגימות התגובה להלם של מסנן.
- $G$  - הגבר.
- $G(k,n)$  - איבר מהסדרה המתקבלת כתוצאה מהתמרת פוריה הדיסקרטית של סדרת  $g(x,n)$ .
- $h(t)$  - תגובה להלם של מסנן.
- $H(z)$  - פונקצית תמסורת.
- $\overline{K}_i$  - מקדם הקורלציה החלקית ה-  $i$ -י.
- $L_{c_0}$  - מספר הדגימות בהגייה דחוטה.
- $L_j$  - מספר הדגימות בהגייה הנבדקת ה-  $j$ -ית לפני דחיטה.
- $L_p$  - אורך המסלול האופטימלי.
- $L_s$  - מספר קטעי הזמן לצורך מיצוי פרמטרים מההגייה הנבדקת לאחור דחיטה.
- $m$  - מספר הדגימות של ההגייה המקורית המשוקללות לצורך קבלת דגימה אחת של ההגייה הדחוטה.
- $N$  - מספר הדגימות בסדרה בתחום הזמן השווה למספר הפסים בתחום התדר, או מספר הדגימות שמכיל קטע זמן ממנו מחושבים הפרמטרים.



רשימת סמלים וקיצורים (המשך)

- $N_c$  מספר מידות הדמיון החלקיות שיש לחשב בתחום הקיום של המסלולים האפשריים (בתכנות דינמי).
- $N_k$  מספר הדגימות בצעד שבו מקודם החלון בתהליך מיצוי הפרמטרים.
- $N_p$  מספר הפסים שמכיל תחום התדר (הדו-צדדי) של אות הדיבור.
- $N_w$  מספר הדגימות שמכיל חלון.
- $p$  סדר המסנן כולו קטבים - מספר הפרמטרים המחושבים מקטע זמן.
- $R(j)$  מקדם האוטוקורלציה עבור ההשהיה  $j$  (תהליך סטציונרי).
- $s_{i,j}$  מידת הדמיון בין קטע הזמן ה-  $i$  לקטע הזמן ה-  $j$ .
- $s(n)$  הדגימה ה-  $n$ -ית של אות הדיבור.
- $S(\omega_k, nT)$  התמרת פוריה הדיסקרטית לזמן קצר של  $s(n)$ , המחושבת בתדירות הזוויתית  $\omega_k$ .
- $S$  מידת הדמיון בין שתי הגיות.
- $t$  זמן.
- $T$  מחזור הדגימה.
- $w(n)$  הדגימה ה-  $n$ -ית של חלון.
- $x_i$  הרכיב ה-  $i$  של וקטור הפרמטרים  $X$ .
- $X$  וקטור פרמטרים.
- $y_i$  הרכיב ה-  $i$  של וקטור הפרמטרים  $Y$ .
- $y^q(nCT)$  דגימה המתקבלת כתוצאה מדחיסת ההגייה בזמן בגורם דחיסה  $C$ ,  
( $q = 1/C$ )
- $Y$  וקטור פרמטרים.
- $z$  משתנה התמרת  $Z$ .
- $\eta_c, \eta_o$  השגיאה הריבועית הכללית המינימלית המנורמלת, בחיזוי לינארי לפי שיטת הקוואריאנס ( $\eta_c$ ) ולפי שיטת האוטוקורלציה ( $\eta_o$ ).
- $\phi_{i,k}$  מקדם הקוואריאנס (האוטוקורלציה) עבור ההשהיות  $i$  ו-  $k$  (תהליך לא סטציונרי).
- $\omega$  תדירות זוויתית.

רשימת סמלים וקיצורים (המשך)

Analogue to Digital Converter	A/D	ממיר אנלוגי לספרתי
Discrete Fourier Transform	DFT	התמרת פוריה הדיסקרטית
Dynamic Programming	DP	תכנות דינמי
Finite Impulse Response	FIR	תגובה לדגם יחידה בעלת משך סופי
Fast Fourier Transform	FFT	
Isolated-Word Recognition System	IWRS	מערכת לזיהוי מילים בודדות
Linear Prediction Coefficients	LPC	מקדמי חיזוי לינארי
Linear Prediction Residual	LPR	שארית החיזוי הלינארי
Linear Time Scaling	LTS	
Partial Correlation	PARCOR	קורלציה חלקית
Time Scale Compression	TSC	דחיסה בזמן

פרק 1: מ ב א

הדיבור נחשב לאמצעי התקשורת הנוח ביותר בין בני האדם. לפיכך, טבעי שבני אדם ינסו לפתח יכולת תקשורת בדיבור בין בני אדם לבין המחשב-מכונה התופשת מקום חשוב בתחומים רבים של הפעילות האנושית. בכדי לקיים תקשורת בדיבור בין המחשב לבין האדם, על המחשב להיות בעל שתי תכונות: כושר הבנת הדיבור ויכולת להגיב בדיבור. המחקר סביב הקניית כושר הבנת דיבור למחשב נעשה במסגרת הנושא זיהוי דיבור באמצעות מחשב (Speech recognition by computer), והמטרה העיקרית שלו היא לפתח טכניקות ומערכות עבור זיהוי דיבור ע"י המחשב.

Reddy (1) מאפיין שש מערכות שונות לזיהוי דיבור בהתאם לרמת הקשיים העומדים בפני כל מערכת. עד כה פותחו רק שלוש המערכות הפחות מסובכות בעוד שהשלוש הנותרות הן עדיין בגדר חזון לעתיד. המערכת הפחות מסובכת מכולן היא מערכת לזיהוי מילים בודדות (Isolated word recognition system), והיא היחידה שהגיעה לייצור תעשייתי.

המערכת לזיהוי מילים בודדות נבדלת משאר המערכות בכך שהיא נדרשת לזהות הגייה - מילה מדוברת בודדת, כאשר לפני ואחרי ההגייה הכרחי שיהיה קטע זמן שאינו מכיל דיבור. ולעומתה שאר המערכות מטפלות בזיהוי דיבור מקושר (connected), כלומר דיבור רצוף (continuous) שבו אין הכרח לרווח זמן ללא דיבור בין הגייה אחת לעוקבת אחריה.

התכונות האקוסטיות של הגייה הקשורה במחרוזת הגיות מסוג זה, שונות מתכונותיה האקוסטיות של הגייה בודדת. יותר מכך, התכונות האקוסטיות של ההגייה הקשורה משתנות בתלות במספר גורמים שאינם קיימים במקרה של הגייה בודדת. כמו למשל, מקום ההגייה במחרוזת וההקשר והמשמעות שבהם היא נהגתה (בהנחה שרצף ההגיות מהווה משפט). התלות הזאת בגורמים לא אקוסטיים, דורשת ממערכות המטפלות בדיבור מקושר, להעזר לצורך הזיהוי במקורות מידע אחרים פרט למקור האקוסטי, בעיקר מקורות לשוניים, בעוד שהמערכת לזיהוי מילים בודדות מקבלת את כל המידע הנדרש לזיהוי רק מהמקור האקוסטי. משום כך היא נחשבת לפחות מסובכת מהמערכות האחרות.

נמקד כעת במערכות לזיהוי מילים בודדות, מכיוון שבמימוש מערכת מסוג זה נדון בהמשך החיבור.

מערכת לזיהוי מילים בודדות מסוגלת לזהות מספר מוגבל של מילים. קבוצת המילים הזאת נקראת מילון המערכת. גודל המילון נע בדרך כלל בין עשר לכמה מאות מילים. ככל שהמילון גדול יותר הזמן הנדרש לזיהוי גדל. אולם לגבי אחוז הטעות בזיהוי, הגדלת המילון עלולה לגרום את הגדלתו, אך הדבר תלוי בגורם נוסף, והוא העמימות (ambiguity) של המילון. כלומר, אם המילון החדש יכול יותר מילים הדומות זו לזו מבחינה אקוסטית, אזי אחוז הטעות בזיהוי יגדל. דוגמא לכך ראה ב- [1], [2].

קיימות מערכות ששינוי תוכן המילון שלהן גורר את שינוי שיטת הזיהוי - האלוגריתמים הממומשים במערכת, על מנת שיתאפשר זיהוי המילים החדשות. מערכות אלו נכנה בשם מערכות תלויות במילון. לעומתן קיימות מערכות שאין צורך לשנות את שיטת הזיהוי שלהן בכל פעם שמוכנס שינוי בתוכן המילון שלהן. אותן נכנה בשם מערכות בלתי תלויות במילון. לפיכך, למערכת בלתי תלויה במילון ניתן להתאים מילון חדש ביתר קלות מאשר למערכת מהסוג הראשון.

אפיון חשוב אחר שקיים עבור מערכות לזיהוי מילים בודדות, הוא האם המערכת תלויה בדובר - מאומנת לדובר אחד מסוים. כלומר, עבור דובר מסוים תתקבל הצלחה בזיהוי גדולה יותר מאשר עבור כל דובר אחר, או שהמערכת בלתי תלויה בדובר, ואז עבור כל דובר שישתמש במערכת תתקבל הצלחה סבירה בזיהוי. מערכת זו היא מסובכת הרבה יותר מבחינת שיטת הזיהוי ביחס למערכת הראשונה ורמת הצלחתה בזיהוי נמוכה מהרמה של המערכת התלויה בדובר. אך יתרונה בכך שאין צורך לאמן מחדש את המערכת עבור כל דובר חדש הרוצה להשתמש בה. אמון המערכת הוא תהליך הכנת האינפורמציה שבעזרתה המערכת מזהה את ההגייה הנבדקת, תהליך זה גוזל זמן, במיוחד עבור מילון גדול, ועובדה זו עושה לעיתים את השימוש במערכת התלויה בדובר לבלתי כדאי. חסרון זה מתגבר אם יש צורך לאמן מחדש את המערכת בתכיפות עבור אותו דובר עקב שינויים שחלים משך הזמן בתכונות האקוסטיות של הגיותיו. לפיכך, הקטנת זמן האימון ככל האפשר, תגרום להחלשת החסרון הנ"ל.

מערכת בלתי תלויה בדובר היא זו המתוארות ב- [3]. המילון של מערכת זו הוא עשר הספרות, ושיטת הזיהוי מבוססת על תכונותיהן, כלומר המערכת תלויה במילון ובכך ישנה מגבלה מסוימת. כנראה שזה התשלום עבור התכונה של אי תלות בדובר.

בין שני הסוגים הנ"ל קיים מקום לסוג ביניים שבו המערכת מאומנת למספר דוברים - אלה המשתמשים בה. לדוגמא [4].

בפרק הבא מוצגות מספר מערכות קיימות המהוות רק חלק מכלל המערכות שפותחו עד כה. מערכות אלה הובאו בכדי להדגים מספר גישות למימוש השלבים השונים שבפעולת המערכת לזיהוי מילים בודדות, ובמידת האפשר נעשתה השוואה בין גישות שונות. חלק מהמערכות המוצגות ומערכות רבות נוספות מתוארות ב-[5].

כעת נציג בטבלה מס' 1.1 את הביצועים של חלק מהמערכות שפותחו בזמן האחרון. הנתונים לקוחים מתוך Fig. 8 ב-[5]. טבלה זו יש בה כדי להראות את רמת הביצוע הנוכחית של מערכות לזיהוי מילים בודדות.

טבלה מס' 1.1: הביצועים של מערכות לזיהוי מילים בודדות שפותחו לאחרונה.

Table 1.1: Performance of recently developed isolated word recognition systems.

החוקר Investigator	המילון Vocabulary	מספר הדוברים No. of Talkers	% השגיאה בקרוב Approx. Error Rate	הערות comments
Martin	10 הספרות באנגלית	דובר אחד	0	תלויה בדובר התוצאה התקבלה עבור דובר מאומן שאמר 1000 הגיות
Martin	34 מילים (המספרים 1÷34) באנגלית	12 דוברים	1.5	תלויה בדובר. רעש רקע של 85dB התוצאה עבור 9149 הגיות
Itakura [2]	200 מילים ביפנית	דובר אחד	2.7	תלויה בדובר התוצאה עבור 2000 הגיות
	36 אותיות א-ב וספרות באנגלית	דובר אחד	11.4	תלויה בדובר התוצאה עבור 720 הגיות.
White [6]	91 מילים רבות הברות באנגלית	דובר אחד	0.5	תלויה בדובר התוצאה עבור 1820 הגיות
	36 אותיות א-ב וספרות באנגלית	דובר אחד	2	תלויה בדובר התוצאה עבור 720 הגיות
Scott	10 הספרות באנגלית	30 דוברים	2	בלתי תלויה בדובר התוצאה עבור 9300 הגיות
Sambur & Rabiner [3]	10 הספרות באנגלית	10 דוברים	2.7	בלתי תלויה בדובר התוצאה עבור 980 הגיות

בחיבור זה מתואר מחקר לפיתוח מערכת לזיהוי מילים בודדות ומימושה בתכנה באמצעות מיני-מחשב NOVA 2.

במערכת שמימושה מוגבל לתוכנה, תהליך הזיהוי נמשך זמן רב יותר מאשר במערכת הממומשת בחלקה באמצעות חומרה אנלוגית ו/או סיפרתית. במטרה לקצר זמן זה, נבדק במסגרת המחקר, השימוש באלגוריתם יעיל לשינוי ציר הזמן של אותות דיבור [7], כאמצעי לדחיסת כל הגייה למשך זמן אחיד, קצר יותר. דחיסה זו גורמת לקיצור הזמן הנדרש לזיהוי, אולם היה צורך לבדוק באיזו מידה היא פוגעת בדיוק הזיהוי. לצרכי בדיקה זו פותחה ומומשה מערכת נוספת לזיהוי מילים בודדות שאיננה כוללת דחיסה. בשתי המערכות נערכו ניסויי זיהוי של אותן ההגיות, כך שניתן היה להשוות בין המערכות מבחינת ההצלחה בזיהוי. נוסף לכך, חושב פי כמה קטנה כמות החשובים לצורך זיהוי הגייה במערכת הכוללת דחיסה לעומת המערכת ללא דחיסה.

השיקולים בפתוח מערכת לזיהוי מילים בודדות הממומשת בתכנה מובאים בפרק 3.

פרק 4 דן באלגוריתם לדחיסת ההגייה בזמן.

בפרק 5 נערך דיון בחיזוי לינארי של אותות דיבור - הבסיס ליצוג הפרמטרי של ההגייה במערכת שפותחה.

פרטי המימוש של המערכת הכוללת דחיסה מובאים בפרק 6.

פרק 7 דן בפיתוח המערכת ללא דחיסה.

הניסויים בזיהוי הגיות שנערכו בשתי המערכות ותוצאותיהם מתוארים בפרק 8.

בפרק 9 מובאת ההשוואה בין שתי המערכות מבחינת כמות החשובים ומבחינת תוצאות הזיהוי. המסקנות מהשוואה זו המסכמות את עבודת המחקר מופיעות בסיום הפרק.

## פרק 2: השלבים בפעולת מערכת לזיהוי מילים בודדות.

בתהליך של זיהוי הגייה במערכת לזיהוי מילים בודדות, ניתן להגדיר שלשה שלבי פעולה עיקריים:

1. עיבוד קדם (Preprocessing).
2. מיצוי תכונות (Feature extraction).
3. מיון ההגיה (Utterance classification).

תהליך הזיהוי מתחיל כאשר הגייה נאמרת ע"י דובר אל מיקרופון או כאשר הגייה מושמעת מסרט הקלטה אנלוגי. בשני המקרים, אות חשמלי אנלוגי לגלי הקול המרכיבים את ההגיה הנבדקת נמסר לשלב עיבוד הקדם. בשלב זה נקבעים קצוות ההגיה ונעשה עיבוד, בעיקר לצורכי הסרת רעש ועיצוב תגובת התדר של אות הדיבור לצורכי השלב הבא. ברוב המערכות נעשית בשלב זה גם המרה לאות ספרתי. מכאן עובר האות המעובד לשלב מיצוי תכונות, שבו ממצים מתוך האות את התכונות או את הפרמטרים המייצגים את ההגיה שלה שייך האות. השלבים שתוארו עד כה ממומשים באמצעות מחשב או באמצעות חומרה מתאימה, לעומת זאת השלב הבא - שלב מיון ההגיה ממומש תמיד באמצעות מחשב. שלב זה מקבל את התכונות המייצגות את ההגיה הנבדקת, מידע נוסף הוא מקבל מזכרון המערכת. מידע זה מתאר את התכונות של כל המילים השייכות למילון המערכת. על סמך השוואה בין תכונות ההגיה הנבדקת לבין התכונות של מילות המילון מתבצע המיון של ההגיה - שיוכה לאחת ממילות המילון. על סמך המיון מתקבלת החלטה שההגיה הנבדקת היא מילה מסוימת אחת מתוך מילות המילון. בהחלטה זו שבה זיהתה המערכת את ההגיה הנבדקת (בצורה נכונה או לא), מסתיים תהליך הזיהוי.

בסעיפים הבאים של פרק זה ינתן תאור מפורט יותר של המתבצע בכל שלב בלוויית דוגמאות ממערכות קיימות.

### 2.1 שלב עיבוד הקדם

שלב זה מקבל ככניסה אות חשמלי אנלוגי לגלי הקול של ההגיה. האות האנלוגי מתקבל ממיקרופון אם ההגיה נאמרה ישירות אל מיקרופון, או מסרט הקלטה אנלוגי עליו היא הוקלטה קודם לכן.

בחלק מהמערכות [8], [9], [3], [2], [10], [6], [11], מתבצעת דגימה של האות האנלוגי באמצעות ממיר אנלוגי לספרתי, לאחר שהאות הועבר דרך מסנן מעביר נמוכים. תדירות הקטעון של המסנן בדרך כלל בין 3KHz ל- 5KHz ובהתאם לכך תדירויות הדגימה המקובלות הן בין 6.667KHz לבין 10KHz. הסיבה לכך שאות הדיבור מוגבל עד ל- 5KHz היא שכמעט כל האינפורמציה הספקטרלית של הדיבור נמצאת בתחום התדרים שמתחת לגבול זה.

לאחר תהליך הדגימה מיוצגת ההגייה בצורה ספרתית ומכאן והלאה כל עיבוד שיעשה יהיה עיבוד ספרתי.

לעומת זאת ישנן מערכות [12], [4], [13], [6], שבשלב זה עדיין לא מתבצעת דגימה, אלא מעבירים את האות האנלוגי במקביל דרך בנק של מסננים, כאשר כל מסנן מעביר תחום תדרים מסוים של ספקטרום האות.

פעולה חשובה הנלווית בשלב זה לפעולות הנ"ל היא קביעת הקצוות של ההגייה, כלומר היכן מתחיל קטע הדיבור בתוך רעש הרקע והיכן הוא מסתיים. לאחר זאת כל עיבוד יעשה רק על אותו קטע זמן שנקבע שהוא מכיל אותות דיבור.

בדרך כלל, מקובל לבצע בשלב עיבוד הקדם גם פעולות של הסרת רעש, החלקה, ופעולות סינון נוספות לעיצוב הספקטרום של אות הדיבור (preemphasis). הפעולה האחרונה באה לשפר את ביצועי השלב הבא.

## שלב מיצוי התכונות

2.2

פעולות השלב הראשון נותנות לאות האנלוגי של ההגייה יצוג פרמטרי ראשוני - דגימות כיצוג ספרתי, או תפוקות של בנק מסננים. אולם בדרך כלל יצוג זה מכיל אינפורמציה יתירה (redundant) ולא רלבנטית, שסילוקה לא יפגע בדיוק הזיהוי. לפיכך, תפקידו של שלב מיצוי התכונות הוא לחשב מתוך היצוג הפרמטרי הראשוני הנ"ל יצוג פרמטרי חדש שיכיל במידת האפשר את כל האינפורמציה הרלבנטית לזיהוי ההגייה, אך עם זאת ללא יתירות. הפעולה הזאת נקראת מיצוי תכונות ואת הפרמטרים החדשים מקובל לכנות בשם תכונות (Features).

חשיבות היצוג הפרמטרי החדש היא בכך שהוא מייצג את ההגייה באמצעות מספר קטן יותר של סיביות אינפורמציה מאשר הייצוג הקודם, ובזה מופחתים קיבול



הזכרון הנדרש לאחסון האינפורמציה וכן כמות החישובים הנדרשים בשלבי הניתוח שבהמשך. לכן, ניתן לראות את המתבצע בצעד זה כדחיסת אינפורמציה.

פרט לדרישה שיכיל אינפורמציה רלבנטית בלבד, ללא יתירות ובצמצום ככל האפשר, קיימת עבור היצוג הפרמטרי החדש דרישה חשובה במיוחד כאשר יצוג זה משמש לצורך זיהוי ההגייה, והיא שהפרמטרים שיתקבלו יאפיינו היטב את ההגייה אותה הם מייצגים, כך שאפשר יהיה להפריד נכונה בין ההגיות השונות במרחב הפרמטרים המייצגים אותן, כלומר למיין בצורה נכונה את ההגייה הנבדקת.

מידת ההצלחה של השלב הבא - שלב המיון - בזיהוי נכון של ההגייה, תלויה במידה בה ממלאים הפרמטרים, המחושבים בשלב מיצוי התכונות, אחר שתי הדרישות הנ"ל. אופי התלות הזאת הוא שגם שיטת מיון מתוחכמת לא תוכל להצליח כאשר הפרמטרים אינם ממלאים דרישות אלה, וככל שהפרמטרים יותר אופטימליים מבחינת הדרישות הנ"ל כן דרושה שיטת מיון מסובכת פחות בכדי להשיג דרגת דיוק רצויה בזיהוי.

לפיכך, שלב מיצוי התכונות הוא שלב המפתח בפעולת מערכת לזיהוי מילים בודדות, אך עד כה טרם פותחה תאוריה כללית אשר בעזרתה אפשר יהיה לבחור את הפרמטרים האופטימליים. לכן תכנון שלב המיצוי נעשה בשיטות יוריסטיות והערכת יעילותם של הפרמטרים נעשית על סמך תוצאות של נסויים רבים בזיהוי הגיות.

נקודת המוצא בבחירת פרמטרים ליצוג ההגייה היא זאת: אוזן האדם פועלת כמנתח ספקטרום הספק [14], מכאן ניתן להסיק שבכדי לאפיין היטב את ההגייה, יש לייצגה באמצעות פרמטרים המכילים בצורה מפורשת, או סתומה, אינפורמציה ספקטרלית. לכן ברוב המערכות לזיהוי מילים בודדות מייצגים את ההגיות באמצעות פרמטרים כנ"ל.

נסקור כעת מספר יצוגים פרמטריים המקובלים במערכות לזיהוי מילים בודדות. לצורך סקירה זו נחלק יצוגים אלה לשני סוגים. לראשון שייכים יצוגים פרמטריים הנגזרים באופן מפורש מהספקטרום של אות הדיבור, ולשני, יצוגים הנגזרים מהתאור הזמני של אות הדיבור, אך גם הם מכילים באופן סתום אינפורמציה ספקטרלית על אות הדיבור.

יצוגים פרמטריים הנגזרים מהספקטרום של אות הדיבור.

2.2.1

במערכת של Clark המתוארת ב- [8], הפרמטרים הם רכיבי ההגייה לפי בסיס אורתונורמלי אופטימלי. הבסיס הזה מחושב מתוך התמרות פוריה הדיסקרטיות של קטעי הגיות השייכות לספרות "ו" עד "יו".

במערכת של Clapper [12] מועבר אות הדיבור האנלוגי דרך בנק של שמונה מסננים, כל מסנן מעביר תחום תדרים מסוים מתוך התחום של 10KHz-100Hz. כל אחת מיציאות המסננים מועברת דרך גלאי עוטפת לממזה תכונות שמספר היציאות ממנו מתאים למספר המסננים. שיא מקומי של עוטפת המופיע באחת הכניסות של הממזה גורם לכך שהממזה יעורר את קו היציאה המתאים לכניסה שבה הופיע השיא, ויחד עם זאת ימנע הממזה הופעת ערור על הקוים השכנים המתאימים לכניסות שקבלו עוטפת בעלות אמפליטודה נמוכה יותר. לפיכך, הופעת ערור על קו מסוים מתוך ממזה התכונות, מראה על קיום פורמנט בתחום התדרים המתאים לאותו קו. פורמנט - שיא מקומי של אנרגיה האות בתחום תדרים כלשהו מתוך הספקטרום של אות הדיבור לפיכך, היצוג הפרמטרי שב- [12] הוא הפורמנטים.

במערכת של Pols המתוארת ב- [4] מועבר אות הדיבור האנלוגי דרך 17 מסננים של  $1/3$  אוקטבה. יציאות המסננים מוגברות באופן לוגריתמי, עוברות גילוי עוטפת ונדגמות כל 15 msec. כלומר כל 15 msec מתקבל וקטור בן 17 ממדים, כל מימד מתאר תחום תדרים מסוים. מבצעים תהליך זה על כל ההגיות של המילון ומחשבים תוך כדי ביצוע התהליך את מטריצת השונות (covariance matrix) של 17 הרכיבים שבוקטור הנ"ל. עבור מטריצה זו מבצעים ניתוח מתמטי הנקרא ניתוח של רכיבים עיקריים, ומקבלים מתוך 17 הממדים 3 ממדים חדשים. מכאן כל דגימה הנ"ל של ההגייה ניתנת ליצוג במרחב של 3 ממדים, המתאים למרחב הנפרש ע"י שלושת הפורמנטים העיקריים.

במערכת של Ichikawa המתוארת ב- [9] מבצעים הערכה על הצלחתם של יצוגים פרמטריים שונים, כל אחד בנפרד, בזיהוי ספרות מדוברות. מתוכם שניים מחושבים ישירות מהספקטרום של אות הדיבור: הלוגריתם של ספקטרום ההספק לאחר החלקה והקפסטרום (cepstrum). אות הדיבור מוגבל לתחום התדרים 150-4200 Hz ונדגם ב- 10 KHz. מיצוי הפרמטרים מבוצע כל 10 msec על סמך 256 דגימות. עבור היצוג הראשון, הספקטרום שחושב ועובד נדגם כל 200 Hz ולכן מתקבל וקטור בעל 25 רכיבים. עבור היצוג השני, הקפסטרום, מחשבים וקטור בן 6 רכיבים.

במערכת המתוארת ב- [13] מועבר אות הדיבור האנלוגי דרך 24 מסננים מעבירי תחום, המכסים את תחום התדרים 110-7000 Hz. יציאות המסננים נדגמות כל 10 msec ומקודדות בצורה הבאה; אם השיפוע של אמפליטודת היציאה חיובי, מקבלת היציאה את הערך 1, אחרת - 0. לפיכך, כל הגייה מיוצגת ע"י סדרת וקטורים, כל וקטור בן 24 רכיבים בינריים המתארים בקוד את הספקטרום של קטע זמן שארכו 10 msec.

White and Neely [6] עורכים השוואה בין מספר יצוגים שונים עבור ההגייה. אחד מהם שייך לסוג היצוגים הראשון. יצוג זה מתקבל בדרך הבאה: אות הדיבור האנלוגי מועבר דרך 20 מסננים מעבירי תחום של 1/3 אוקטבה, המכסים תחום של 100-10,000 Hz, על יציאת כל מסנן נעשית החלקה, החסרת רעש, העברה לסקלה לוגריתמית ולבסוף דגימה בקצב של 100 Hz (דגימה כל 10 msec). לפיכך, כל 10 msec מתקבל וקטור פרמטרים בן 20 ממדים, המתאר את הספקטרום של קטע הזמן הנדון. יצוג פרמטרי אחר עבור ההגייה מתקבל כאשר במקום 20 מסננים משתמשים ב- 6 מסננים של אוקטבה, ולכן וקטור הפרמטרים יהיה בן 6 ממדים.

במערכת של Velichko המתוארת ב- [10], האות הספרתי שהתקבל עקב דגימת אות הדיבור בקצב של 20 KHz, מועבר דרך מערכת של 5 מסננים ספרתיים (מסדר שני). התדירויות המרכזיות של המסננים הן: 225, 450, 900, 1800 ו- 7200 Hz וגורם הטיב שלהם (Q) היה 2.45. כל הגייה נחלקה לקטעי זמן שווים, כ"א באורך של 1.4 msec. עבור כל קטע זמן חושבו,  $E_0$  - הספק האות שהספקטרום שלו מכסה את כל תחומי התדרים של המסננים הנ"ל, וכן הספקי האותות ביציאות המסננים:  $E_i, i=1, \dots, 5$ . הספקים אלה שמשו לקבלת וקטור הפרמטרים:  $\{\ln(E_0/E_i)\}_{i=1, \dots, 5}$ . שייצג את קטע הזמן.

בבחירת ביטוי לוגריתמי מסתמכים ב- [10] על חוק Weber-Fechner, לפיו השנוי המינימלי בעצמה הניתן להבחנה ע"י איבר חישה (האוזן), יחסי לערך העוצמה סביבה נעשה השינוי, ומכאן גודל התחושה יחסי ללוגריתם של עצמת הערוור.

### 2.2.2 יצוגים פרמטריים הנגזרים מהתאור הזמני של אות הדיבור.

לצד הייצוגים הפרמטריים של הסוג הראשון מטפל Ichikawa [9] ביצוגים המשתייכים לסוג השני, כלומר יצוגים המחושבים בתחום הזמן.

הראשון מתוכם הוא פונקציית האוטוקורלציה  $R(i)$  שהוגדרה כך:

$$R(i) = \begin{cases} \sum_{j=0}^{N-1-i} x(j)x(j+i) & 1 \leq i \leq 4 \\ 0 & \text{אחרת} \end{cases} \quad (2.1)$$

כאשר  $\{x(j)\}$ ,  $j=0, \dots, N-1$  הן  $N$  דגימות דיבור לאחר ששוקללו באמצעות חלון Hamming באורך  $N$ . ב- [9] נבחר  $N=256$ , ומכיוון שקצב הדגימה היה 10 KHz, אזי משך קטע הזמן שממנו מוצו הפרמטרים הנייל היה 25.6 msec, כאשר כל 10 msec של אות הדיבור חושב וקטור פרמטרים חדש. הקשר של יצוג פרמטרי זה אל ספקטרום אות הדיבור הוא ברור, מכיוון שהתמרת פורייה ההפוכה של ספקטרום ההספק נותנת את פונקציית האוטוקורלציה.

היצוג הפרמטרי הבא הוא מקדמי החיזוי הלינארי (LPC - Linear Predictive Coefficients). מקדמים אלה מתקבלים בדרך הבאה [15]:

א. מניחים שדגימת אות הדיבור  $s(n)$  היא תפוקה של מסנן כולו קטבים (all-pole model) עבור כניסה  $u(n)$ . פונקציית התמסורת של מסנן זה היא:

$$H(z) = \frac{S(z)}{U(z)} = \frac{G}{1 + \sum_{k=1}^p a_k z^{-k}} \quad (2.2)$$

לפיכך, בתחום הזמן קיים הקשר הבא:

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

כלומר,  $s(n)$  מתקבל מתוך צרוף לינארי של  $p$  דגימות העבר המשוקללות באמצעות  $p$  מקדמים  $\{a_k\}$ ,  $k=1, \dots, p$ , ועקב כניסה  $u(n)$  המוגברת בגורם  $G$ .

ב. בהנחה שהכניסה  $u(n)$  אינה ידועה, ניתן לחזות רק בקרוב את  $s(n)$  בעזרת סכום של דגימות העבר המשוקללות.  
 $\tilde{s}(n)$  יהיה הקרוב ל-  $s(n)$ , כאשר

$$\tilde{s}(n) = - \sum_{k=1}^p a_k s(n-k) \quad (2.4)$$

השגיאה בין  $s(n)$  - הדגימה האמיתית לבין  $\tilde{s}(n)$  היא

$$e(n) = s(n) - \tilde{s}(n) = s(n) + \sum_{k=1}^p a_k s(n-k) \quad (2.5)$$

כאשר מסכמים את ריבועי השגיאות עבור סידרת דגימות, מתקבלת השגיאה הריבועית הכללית הנקראת גם שארית החיזוי הלינארי (LPR - Linear Prediction Residual):

$$\text{LPR} (s(n), a_k) = \sum_n e^2(n) \quad (2.6)$$

טווח הסיכום ב- (2.6) וההדגרה של  $s(n)$  בתחום זה, הם בעלי חשיבות ונדון בהם בהמשך בפרק 5. מקדמי החיזוי הלינארי (LPC) הם  $k=1, \dots, p, \{a_k\}$  המגשימים את הערך המינימלי של הביטוי הנתון ב- (2.6). תאור מפורט של תהליך המינימיזציה הזה ינתן בפרק 5. על סמך המודל שהוצג ב- א', המעטפת הספקטרלית של אות הדיבור בקטע זמן מסוים ניתנת ע"י  $|H(z)|^2_{z=e^{j\omega T}}$ , הוא מחזור הדגימה היוצרת את  $s(n)$ , כאשר  $k=1, \dots, p, \{a_k\}$  (ה- LPC) חושבו מתוך סידרת הדגימות של קטע הזמן הנדון.

היצוג הפרמטרי האחרון ב- [9] היה מקדמי הקורלציה החלקית - (RARCOR-Partial Correlation Coefficients). המכונים גם בשם מקדמי החזרה - [Reflection Coefficients] [15].

נסמן ב-  $f_p(n)$  את השגיאה  $e(n)$  שב- (2.5) ונקרא לה השגיאה בחיזוי קדימה (forward prediction), כלומר, זו השגיאה בחיזוי  $s(n)$  על סמך  $p$  הדגימות המקדימות אותה בזמן:  $s(n-p), \dots, s(n-1)$ . באופן דומה מגדירים שגיאה בחיזוי לאחור (backward prediction)  $b_p(n)$ , כשגיאה בחיזוי  $s(n-1-p)$ , על סמך  $p$  הדגימות הנ"ל הבאות אחריה בזמן, באמצעות אותם מקדמי החיזוי הלינארי הנותנים את  $f_p(n)$ . לכן

$$b_p(n) = s(n-1-p) + \sum_{k=1}^p a_k s(n-1-p+k) \quad (2.7)$$

מקדם קורלציה חלקית  $K_{p+1}$ , מוגדר כערך השלילי של הקורלציה הסטטיסטית בין  $f_p(n)$  לבין  $b_p(n)$ , כלומר:

$$K_{p+1} = - \frac{E(f_p(n)b_p(n))}{\sqrt{E(f_p^2(n))E(b_p^2(n))}} \quad (2.8)$$

כאשר  $E(\ )$  מבטא תוחלת. הסבר מפורט על מקדם הקורלציה החלקית ותכונותיו יובא בפרק 5.

ב- [9] מופיעה טבלה המשווה את הביצועים של היצוגים הפרמטריים השונים שנבדקו שם. מלבד ההצלחה בזיהוי מופיעים בטבלה גם הזמן הנדרש לזיהוי וכן מימד וקטור הפרמטרים עבור כל יצוג פרמטרי. מבין היצוגים של הסוג הראשון (הנגזרים באופן מפורש מספקטרום אות הדיבור), ה- cepstrum הוא היעיל ביותר. עבור שני היצוגים של הסוג הראשון התקבלה הצלחה של 100% בזיהוי עשר הספרות היפניות, כאשר כל ספרה נאמרה 9 פעמים ע"י קריין מקצועי. אולם, זמן הזיהוי עבור ה- cepstrum היה קטן פי שניים מאשר עבור היצוג האחר - הלוגריתם המוחלק של ספקטרום ההספק. נוסף לכך, מימד וקטור הפרמטרים של הראשון היה 6 לעומת 25 של השני. לעובדה זו יש משמעות מבחינת כמות הזכרון הדרושה עבור המערכת, פרט להשפעתה הישירה על הזמן הנדרש לזיהוי.

מבין היצוגים הפרמטריים של הסוג השני (המחושבים בתחום הזמן) שנבדקו ב- [9], ה- LPC השיג את התוצאה הגרועה ביותר בזיהוי - 77%, לעומת מקדמי

ה- PARCOR שהשיגו את התוצאה הטובה ביותר בזיהוי - 98%. יש לציין שזמן הזיהוי עבור היצוגים הפרמטריים של הסוג השני אינו ארוך יותר מאשר מחצית זמן הזיהוי הנדרש עבור ה- cepstrum, לכן הושם דגש מיוחד על שיפור הביצועים של מקדמי ה- PARCOR, ובעזרת שקלול המקדמים לפי האמפליטודה של קטע אות הדיבור אותו הם מיצגים, התקבלה הצלחה של 100% בזיהוי. מימד וקטור מקדמי ה- PARCOR בבדיקה זו היה 6, כזה של וקטור ה- cepstrum. המסקנה הנובעת מתוצאות הניסויים שב- [9] היא שמקדמי ה- PARCOR הם היעילים ביותר מבין כל היצוגים שנבדקו שם.

המערכת של Sambur and Rabiner המתוארת ב- [3] היא מערכת לזיהוי ספרות באופן בלתי תלוי בדובר. למערכת זו נבחרו יצוגים פרמטריים שבעזרתם התבצע הזיהוי, על סמך שלוש הדרישות הבאות:

1. היצוג הפרמטרי יהיה ניתן למדידה בפשטות ובאופן ברור.
2. יאפיין חלק גדול מקולות הדיבור.
3. היצוג הפרמטרי יהיה ניתן לפרוש בצורה נוחה, באופן בלתי תלוי בדובר. הדרישה השלישית באה עקב יעודה של המערכת - להיות בלתי תלויה בדובר.

לפיכך, נבחרו היצוגים הפרמטריים הבאים:

א. קצב חציית האפס (ZCR - Zero Crossing Rate) המוגדר כמספר חציות האפס בקטע זמן קבוע בסדר גודל של 10 msec. (קצב דגימת אות הדיבור הוא 10 kHz).

ב. אנרגיה המוגדרת כסכום ריבועי הדגימות של אות הדיבור בקטע זמן.

ג. השגיאה המנורמלת המתקבלת בעזרת 2 מקדמי LPC המיצגים את קטע הזמן.

ד. תדירויות הקטבים המתקבלות בעזרת 2 מקדמי LPC המיצגים את קטע הזמן.

לפי [3], השימוש בערכי סף קבועים מראש עבור מיצוי הפרמטרים גורם לבעיות בזיהוי, מכיוון שערכי סף אלה ברובם תלויים בדובר או בזמן. לכן בכדי להתגבר על מכשול זה נעשה נרמול עצמי של כל יצוג פרמטרי. כלומר, רוב ערכי הסף המשמעותיים ביותר הקשורים במדידת ערכם של הפרמטרים, מתקבל מתוך מדידות המבוצעות באופן ישיר על אות הדיבור הנבדק.

במערכת של Itakura המתוארת ב- [2] אין משתמשים ישירות ב-LPC כיצוג פרמטרי וזאת עקב אי הצלחתם לעומת יצוגים אחרים כפי שהראה Ichikawa [9]. Itakura משער שאי הצלחה זו, בחלקה נובעת מכך, שמרחב התכונות הנפרש ע"י ה-LPC הוא מורכב יתר על המידה בכדי שאפשר יהיה להציע עבורו מידת מרחק בין האלמנטים שבו, שתהיה פשוטה ויעילה.

הגישה של Itakura [2] לפתרון הבעיה של בחירת פרמטרים ומידת מרחק יעילים, היא מנקודת מבט סטטיסטית. הוא מראה שהלוגריתם של יחס הסבירות, שהוא הקריטריון הטוב ביותר לבדיקת ההשערה של השתייכות הגייה לאלמנט במרחב LPC המהווה וקטור יחוס, ניתן לצמצום ללוגריתם של היחס בין ה-LPR המתקבל עבור חיזוי לינארי של הגייה נבדקת באמצעות ה-LPC של וקטור ייחוס, לבין ה-LPR המינימלי של ההגייה הנבדקת. ואז ניתן להשתמש בביטוי המתקבל כמד מרחק יעיל.

כפי שנאמר קודם, נבדקים במערכת של White and Neely המתוארת ב- [6] מספר יצוגים. הראשון שכבר נסקר - בנק של מסננים משתייך לאלה הנגזרים מתחום התדר באופן מפורש. השני משתייך לסוג היצוגים הפרמטריים המחושבים בתחום הזמן והוא קשור ל-LPC. ב- [6] השימוש ב-LPC מבוסס על הגישה של Itakura המוצעת ב- [2] ביחס ל-LPR.

במערכת של White and Neely [6] מופיעה דרגת ביניים הנקראת דרגת הפחתת נתונים (data reduction), מיד לאחר שלב מיצוי הפרמטרים. בדרגה זו מבוצע התהליך הבא. עבור כל קטע זמן של 10 msec של דיבור, מחושב המרחק בין קטע הדיבור הנבדק לבין כל אחד מתוך 20 קולות דיבור שונים המשמשים כקולות ייחוס. לכל אחד מהם יש שם (label) והקטע הנבדק מקבל את שמו של קול הייחוס הקרוב ביותר אליו. לאחר שמבצעים תהליך זה עבור כל קטעי ההגייה הנבדקת, מתקבלת מחרוזת של שמות המייצגת את ההגייה. לכן נקראת שיטת ייצוג זו "character-string encoding" - C.S.Code. כאשר מיוצגת ההגייה ב-C.S.Code ניתן לסלק את וקטור הפרמטרים שיצג את ההגייה ובכך לחסוך בזכרון ובזמן חישוב באופן ניכר. אך יחד עם זאת, נגרמות בשיטה זו שגיאות במיון ההגיות.

ב- [6] מסכמים את ההשפעות של מספר גורמים על הצלחת המערכת בזיהוי ההגיות. בטבלה 2.1 מתוארת השפעת היצוג הפרמטרי על תוצאות הזיהוי. בטבלה מדורגים היצוגים הפרמטריים השונים לפי קצב סיביות (מספר הסיביות הנדרש עבור יצוג פרמטרי נתון כדי ליצג שניה של דיבור) יורד. לירידה זו מתאימה ירידה בהצלחה בזיהוי ההגיות של מילון הכולל את הספרות והאותיות. זמן הזיהוי



גם הוא יורד עם ירידת קצב הסיביות של היצוגים הפרמטריים. מהתוצאות הדומות שהתקבלו עבור הפרמטרים המופקים מבנק של מסננים ועבור הפרמטרים המבוטסים על LPC, ניתן להסיק שהם אקוילונטיים מבחינת הצלחתם בזיהוי.

טבלה 2.1: הביצועים של היצוגים הפרמטריים השונים שב- [6].

Table 2.1: Parametric representations performance in [6].

היצוג הפרמטרי Parameter representation	תוצאת הזיהוי Recognition Result (%)	זמן זיהוי הגייה Recognition time per utterance (sec)	קצב הסיביות Data rate (bits/sec)
בנק של 20 מסננים	98	30	12000
LPC	97	20	4200
בנק של 6 מסננים	96	15	3600
C.S. Code	91	2	500

המערכת המתוארת ב- [11], היצוג הפרמטרי עבור ההגייה מבוטס על LPR בדומה למה שפותח ע"י Itakura [2]. כן מראים ב- [11], בצורה חדשה, כיצד שיטת יצוג זו מתבססת למעשה על חישוב ההתאמה הספקטרלית של צורות גל.

בסעיף זה נסקרו יצוגים פרמטריים שונים עבור ההגיות. היצוגים סווגו לשני סוגים. אלה המחושבים בתחום הזמן ואלה הממוצים בתחום התדר. אולם, כפי שהוצג בסקירה, גם האחרונים מיצגים למעשה את הספקטרום של אות הדיבור. כל אחד מהפרמטרים שהוזכרו משמש כבסיס הנתונים עבור שלב מיון ההגייה, השלב הבא של תהליך הזיהוי במערכות לזיהוי מילים בודדות.

### שלב מיון ההגייה

2.3

החלק הראשון של סעיף זה מגדיר את שלב מיון ההגייה ומתאר שתיים משיטות המיון המקובלות. החלקים הבאים אחריו עוסקים בשני נושאים מרכזיים של שלב המיון: התאמת צירי הזמן של ההגיות וחישוב מידת הדמיון בין שתי ההגיות.

שלב זה מקבל כקלט את וקטור הפרמטרים המייצג את ההגייה הנבדקת. הפלט של שלב המיון הוא החלטה הקובעת לאיזו ממילות המילון של המערכת שייכת ההגייה הנבדקת.

אם השלב הקודם היה קשור באופיו של האות הנדון, הרי השלב הנוכחי - שלב המיון, ברוב המקרים, אינו "מכיר" כלל את האות המקורי והוא מטפל אך ורק בוקטורים המייצגים את האות המקורי. לפיכך, שיטות מיון המקובלות בנושאי זיהוי צורות, עשויות להתאים גם למיון של הגיות. נדון כאן בשתיים מהן: התאמת הדפסים (templates matching) ומיון בעזרת עץ החלטה.

השיטה הראשונה - התאמת הדפסים, היא המקובלת ביותר במערכות לזיהוי מילים בודדות. היא הונהגה בכל המערכות המוזכרות בפרק זה פרט למערכת של Sambur and Rabiner המתוארת ב-[3]. בשיטה זו בודקים את מידת ההתאמה (או הקירבה) של מייצג ההגייה הנבדקת (המייצג הוא "ההדפס" של ההגייה) למייצגים של כל אחת ממילות המילון של המערכת. כלומר מחשבים את מידת הדמיון בין וקטור הפרמטרים המייצג את ההגייה הנבדקת לבין כל אחד מוקטורי הפרמטרים המייצגים את מילות המילון של המערכת (וקטורי הייחוס), ומחליטים שההגייה הנבדקת שייכת למילה שעם וקטור הפרמטרים המייצג אותה התקבלה מידת הדמיון הגבוהה ביותר. מידת הדמיון בין שני וקטורי פרמטרים מתקבלת כאשר מציבים אותם בפונקציה הנקראת פונקציית דמיון. הדיון בפונקציות דמיון ינתן בסעיף 2.3.3.

וקטור פרמטרים האגור בזכרון המערכת והמשמש כוקטור ייחוס, מתקבל בדרך כלל באחת משתי דרכים. לפי הראשונה, מחשבים אותו מתוך הגייה בודדת (הגיית הייחוס), בעזרת אותם אלגוריתמים המשמשים לחישוב וקטור הפרמטרים של ההגייה הנבדקת. לפי הדרך השנייה מחושב וקטור הייחוס מתוך מיצוע של מספר וקטורי פרמטרים המייצגים הגיות שונות של אותה מילה. בדרך זו מנסים לשפר את תוצאות הזיהוי המתקבלות כאשר משתמשים בדרך הראשונה. וזאת על סמך ההנחה שההבדלים בין הגיות שונות של אותה מילה הם ניכרים, אפילו אם הן נאמרו ע"י דובר יחיד, כך שאין לצפות שהגייה בודדת תוכל לשמש כהגיית יחוס מתאימה לכל הגייה של אותה מילה. לעומת זאת וקטור פרמטרים המתקבל ממיצוע של מספר וקטורי פרמטרים המייצגים מספר הגיות של אותה מילה, עשוי לתאר טוב יותר את השינויים האפשריים בתכונות האקוסטיות של המילה, ולכן יוכל לשמש כוקטור יחוס יעיל

יותר מוקטור פרמטרים של הגייה בודדת. אולם, כל מיצוע מסוג זה דורש בדיקה, האם שיתוף וריאציות רבות מידי במיצוע אינו גורם להקטנת ההבדלים בין וקטורים ממוצעים של מילים שונות. תוצאות זיהוי התומכות בדרך השניה מופיעות ב- [8] וב- [11], אולם קיימות גם מערכות [2], [6], [9], [10] היוצרות וקטור יחוס בדרך הראשונה, שהיא בודאי מהירה יותר מאשר השניה. כל המערכות האלה תלויות בדובר ולכן חישוב וקטור יחוס מהגייה בודדת של הדובר אינו גורם לפגיעה בהצלחת הזיהוי. לעומת זאת במערכת המאומנת למספר דוברים אין כלל מקום לחישוב וקטור היחוס בדרך הראשונה. במערכת זו וקטור היחוס צריך להכיל מידע המייצג את ההגייה כפי שהיא נהגית בפיהם של הדוברים השונים שעבורם מאומנת המערכת, ולכן וקטור היחוס מחושב כאן ע"י מיצוע של וקטורי הפרמטרים המייצגים הגיות שונות של אותה המילה, כאשר כל דובר תורם הגייה אחת לפחות. במערכת המתוארת ב- [4] משתתפים 20 דוברים ביצירת 20 וקטורי היחוס ליצוג 20 מילים, כאשר כל דובר תורם הגייה אחת עבור כל מילה.

לשיטה של התאמת הדפסים כפי שהיא מתוארת לעיל ישנו חסרון המתבטא בכך שהגדלת מילון המערכת גוררת, כאשר מפעילים שיטה זו, הגדלת הזמן הנדרש לזיהוי הגייה, עקב התוספת למספר בדיקות ההתאמה הדרושות בשלב המיון. לכן עבור מילון גדול (המכיל יותר מ- 100 מילים) מקובל לבצע את השיטה הנייל בדרך של החלטות עקיבות (sequential decisions) [2]. לפי דרך זו מתחילים לחשב את מידות הדמיון בין חלק אחד של וקטור ההגייה הנבדקת, החלק המייצג את תחילת ההגייה, לבין החלק המתאים לו בזמן מכל אחד מוקטורי היחוס. וקטור יחוס אשר מידת הדמיון שחושבה עבורו קטנה מסף שנקבע מראש, נדחה. בצעד הבא מחשבים את מידות הדמיון בין קטע של הוקטור הנבדק המורכב מהחלק הקודם בצרף חלק הבא אחריו בזמן, לבין הקטע המתאים לו בזמן מכל אחד מוקטורי היחוס שלא נדחו בצעד הקודם. בדרך זו מתקדמים, כאשר ערך הסף מתעדכן בכל צעד, עד אשר בצעד האחרון מחשבים את מידת הדמיון בין הוקטור הנבדק בשלמותו לבין כל אחד מוקטורי היחוס שנותרו, כלומר אלו שלא נדחו בצעדים הקודמים. כעת המיון נעשה לפי מידת הדמיון המקסימלית. אם לצעד האחרון לא נשאר וקטור ייחוס, מחליטים שההגייה הנבדקת אינה שייכת למילון.

התהליך הזה של החלטות עקיבות מתבסס על ההנחה שמידת הדמיון בין קטעים מתאימים של הוקטורים המושוים, ביחס לערך סף מתאים, דייה להצביע על הסיכוי שיש לוקטור מסוים להיות הדומה ביותר. לכן, סף שאינו מתאים יגרום לשגיאות בזיהוי ומשום כך יש חשיבות רבה בתהליך זה, לחישוב נכון של הסף.

מיון בדרך של החלטות עקיבות חוסך לפיכך בזמן חישוב עקב ההפחתה במספר ההשוואות, אך יחד עם זאת גדל בו הסיכוי לשגות בזיהוי ההגייה.

שיטת המיון השנייה שצוינה בתחילת הפרק היתה מיון בעזרת עץ החלטה. שיטה זו מודגמת יפה במערכת של Sambur and Rabiner המתוארת ב- [3]. מערכת זו היא בלתי תלויה בדובר והמילון שלה מכיל את עשר הספרות באנגלית.

בשלב המיון עולים ב- [3] על עץ החלטה אשר התנועה בו מענף מסוים לענף אחר נעשית בהתאם לערכי הפרמטרים שחושבו מתוך ההגייה הנבדקת. תנועה זו אל ענף מסוים (אחד מתוך שניים אפשריים), פרושה דחיית המילים שבכמה תכונות לא התאימו להגייה הנבדקת וצמצום הבדיקה בהמשך, למספר מילים קטן יותר הקשורות לענף מסוים אליו מגיעים. בדרך זו מגיעים בסוף התהליך לענף שקשורה אליו מילה בודדת ומחליטים שזו המילה המתאימה ביותר להגייה הנבדקת. בשיטה זו של עץ החלטה קיימת סכנה שעקב אי דיוקים בחישוב פרמטר מסוים (במערכת הני"ל מחשבים ארבעה), מגיעים לענף המוביל להחלטה שגויה. לכן מקובל, וכך נוהגים גם ב- [3], לשלב בתהליך מיון מסוג זה, אמצעי ביקורת המתריאים על כך שעליה על ענף מסוים יוצרת סבירות גבוהה מדי לקבלת שגיאה בזיהוי, ואז יש צורך לחזור לאחור ולבחור בענף אחר.

לסיכום, שיטת מיון בעזרת עץ החלטה היא מסובכת הרבה יותר מאשר השיטה של התאמת הדפסים, אך היא אחד מהגורמים המאפשרים למערכת המתוארת ב- [3] לשמור על אי-תלות בדובר ויחד עם זאת להשיג הצלחה בזיהוי שכמעט ואינה נופלת מזו המתקבלת במערכת המאומנת לדובר. אולם, כל זה מושג במחיר של התלות החזקה במילון, מכיוון, שעץ החלטה בנוי על סמך התכונות האקוסטיות של מילות המילון, וכל שינוי בתוכן המילון יחייב שינויים רציניים בעץ החלטה או אף את בנייתו מחדש.

### 2.3.2 התאמת צירי הזמן של ההגיות

נושא התאמת צירי הזמן (time alignment) של ההגיות קיים במערכות בהן מימד וקטור הפרמטרים המייצג את ההגייה נמצא ביחס ישר למשך ההגייה. במערכות אלו כל קבוצת רכיבים של הוקטור מייצגת קטע זמן אחר של ההגייה, כאשר קטעי הזמן הם בעלי אורך אחיד. מכיוון שבדרך כלל לכל מילה, ואף לשתיה הגיות של אותה מילה הנאמרת עי"י אותו דובר, משך הגייה שונה, אזי וקטורי

הפרמטרים המיצגים את ההגיות יהיו גם הם בעלי אורך שונה. לכן, על מנת שאפשר יהיה לחשב את מידת הדמיון בין זוג וקטורים, יש צורך לבצע התאמה או נירמול של צירי הזמן של שני הוקטורים המשווים. לעומת זאת, קיימות מערכות [3] בהן כל ההגיות מיוצגות ע"י וקטורי פרמטרים בעלי אורך אחיד שבהם כל רכיב מיצג באופן כמותי תכונה מסויימת של ההגיה. לכן עבור מערכות אלה בעית ההבדלים במשכי הזמן של ההגיות אינה קיימת בשלב המיון, למעשה היא מתחסלת בשלב מיצוי התכונות.

בסעיף זה נתייחס למערכות בהן נדרש לבצע התאמת צירי הזמן של ההגיות. ניתן להבחין בשלוש גישות קיימות לבצוע התאמה זו.

1. בזמן מיצוי הפרמטרים של ההגיה מתבצע שינוי משך הזמן של ההגיה למשך זמן שנקבע מראש כמשך זמן אחיד עבור כל ההגיות של כל מילות המילון. בהתאם לגישה זו מפעיל Clapper [12] שיטה, אותה הוא מכנה nonlinear time base. שיטה זו ממומשת כמו כל המערכת, בחומרה בלבד. במערכת זו כל מילה של המילון מיוצגת באמצעות מטריצת זכרון בעלת 48 תאים, כאשר כל תא נושא ערך בינארי. במטריצה 8 שורות כאשר כל שורה מתאימה לתחום תדרים מסוים, ו- 6 עמודות - כל עמודה מתאימה לקטע זמן אחר של ההגיה. מטריצת הזכרון הזאת נטענת בשלב מיצוי הפרמטרים, עמודה אחר עמודה בתאום עם מהלך ההגיה בזמן. פרק הזמן שבו נמצאת עמודה בשלב של טעינה אינו אחיד לגבי כל העמודות, אלא הוא גדל באופן לינארי מהעמודה הראשונה המתאימה לקטע הזמן ההתחלתי של ההגיה, עד לעמודה האחרונה המתאימה לקטע האחרון של ההגיה. לפיכך, בסיס הזמן של המטריצה שלפיו נטענות 6 העמודות, אינו לינארי. בנוסף לכך, קיים במערכת זו מנגנון מיוחד הדואג לכך שהגיה ארוכה במיוחד תכווץ כך שתתואר באמצעות שש עמודות\* מימוש המערכת כמתואר ב- [12], מתבסס על ההנחה שלקראת סוף הגיית המילה, פרק הזמן בו תכונה אקוסטית מסוימת מופיעה, אינו יציב והוא שונה מהגיה להגיה של אותה המילה. לעומת זאת, בתחילת ההגיה קיימת עקביות גדולה יותר בעיתוי של הופעת התכונות, ועקביות זו הולכת ופוחתת באופן לינארי לקראת סוף ההגיה. לפיכך, שיטת המימוש של המערכת הנ"ל מונעת "מריחה"

\* עבור הגיה ארוכה במיוחד, פועל המנגנון רק כאשר הוא מזהה קול חוככי (fricative) בתחילת ההגיה. לפי [12] משך הזמן של קול מסוג זה כאשר הוא מופיע בתחילת ההגיה, הינו ארוך ונתון לשינויים. לכן, כאשר מתגלה קול חוככי בתחילת ההגיה, מוחזקת העמודה הראשונה במצב של טעינה זמן רב יותר מהרגיל, וכך מתבצע גם עבור העמודות שאחריה. בדרך זו נדחסת הגיה כזאת ב- 6 עמודות זמן.

הגיה קצרה במיוחד מיוצגת ע"י מספר עמודות היחסי למשכה, ואין מבצעים עבורה מתיחה לצורך יצוגה באמצעות כל ששת העמודות.

שלא לצורך של מידע על פני מספר עמודות, בעיקר לקראת סיום ההגייה. כלומר, מתבצעת דחיסת אינפורמציה במידת מה.

2. לאחר מיצוי הפרמטרים אך לפני מיון ההגייה מבצעים תאום או נירמול ציר הזמן של וקטור הפרמטרים המייצג את ההגייה הנבדקת, כך שארכו ישווה לאורך האחיד המקובל עבור כל אחד מוקטורי היחוס. (בתהליך נרמול זהה מביאים את וקטורי היחוס לאותו אורך אחיד).

המערכת של Pols המתוארת ב- [4] פועלת בהתאם לגישה זו. במערכת זו מחשבים עבור כל 15 msec של הגייה וקטור תכונות. אחת המסקנות המובאות ב- [4] על סמך ניסויים שנערכו היא, שייצוג כל אחת מההגיות באמצעות 35 וקטורים כאלה אינה פוגעת בהצלחת הזיהוי ביחס למערכת אחרת המתוארת ב- [4], שבה לכל הגיית יחוס היה מספר וקטורים יחסי לארכה, ובשלב המיון לפני כל השוואה, היה צורך לשנות את מספר הוקטורים של ההגייה הנבדקת כך שישווה למספר הוקטורים של הגיית היחוס עמה היא משווית. לפיכך, בכדי לקבל 35 וקטורים עבור כל הגייה, משמיטה המערכת וקטורים מסוימים, או חוזרת על וקטור אחד מספר פעמים.

מערכת אחרת שבה נבדקת גישה זו עבור פתרון הבעיה של קיום הבדלים באורכי ההגיות, היא המערכת של White and Nelly המתוארת ב- [6]. השיטה שם נקראת בשם Linear time scaling (with shift) נכנה אותה LTS. בשיטה זו וקטורי הפרמטרים המייצגים את ההגיות נמתחים או נדחסים באופן ליניארי כך שכולם יגיעו לאותו אורך. לפיכך, וקטור הפרמטרים המייצג את ההגייה הנבדקת מנורמל באופן ליניארי כך שארכו ישווה לאורך האחיד של כל אחד מוקטורי היחוס, ואז ניתן יהיה לחשב את המרחק (או הדמיון) בין הוקטור הנבדק לבין וקטור יחוס. בכדי לקזז שגיאות הנובעות מקביעה לא מדויקת של קצוות ההגייה בשלב עיבוד הקדם, וקטור ההגייה הנבדקת מוזז באופן ליניארי (linear shift), ימינה ושמאלה יחסית לוקטור היחוס, ובכל פעם מחושב המרחק ביניהם. מניחים שהמרחק הקטן ביותר המתקבל, נובע מההתאמה הטובה ביותר של צירי הזמן של שני הוקטורים המשווים, והוא זה שנלקח כמייצג את מידת הדמיון בין ההגייה הנבדקת לבין הגיית היחוס אליה היא משווית.

תארנו שתי מערכות הפועלות בהתאם לגישה מס' 2, לפיה יש להביא את וקטורי ההגיות לאורך אחיד. אולם, קיים בין שתיהן הבדל יסודי. במערכת של Pols המתוארת ב- [4] מעדיפים לא להשתמש בנתונים הנובעים מאינטרפולציה,

ולכן אין מעבדים שם מחדש ואין משנים את וקטור הפרמטרים שהתקבל בשלב מיצוי התכונות, אלא משמיטים ממנו חלקים או חוזרים על חלקים מספר פעמים וכך מושג האורך האחיד. לעומת זאת במערכת המתוארת ב- [6], בעזרת פעולות לינאריות מיצרים מתוך וקטור הפרמטרים שהתקבל בשלב מיצוי התכונות, וקטור פרמטרים חדש בעל אורך רצוי. יש לציין שבשיטה זו קיימת סכנה של ערבוב בין צורות דיבור שונות, למשל דיבור קולי (voiced) הנוצר עקב הרעדת מיתרי הקול, עם דיבור לא קולי (unvoiced) הנוצר ללא ערוך מיתרי הקול. ערבוב זה עלול להכניס שגיאות בזיהוי הגיות. סכנה זו מנסים כנראה למנוע במערכת המתוארת ב- [4].

3. לפי גישה זו מחשבים את מידת הדמיון בין וקטורי הפרמטרים כפי שהתקבלו בשלב המיצוי. כלומר, בניגוד לגישה הקודמת אין מביאים את וקטורי הפרמטרים לאורך אחיד, ולכן תוך כדי חישוב מידת הדמיון מופעלים אלגוריתמים לתאום בין צירי הזמן של הוקטורים המשווים. אלגוריתמים אלה מבוססים על תכנות דינמי (DP) Dynamic Programming, המבצע את תאום צירי הזמן באמצעות תהליך לא לינארי. נתאר תהליך זה בקצרה. כפי שנאמר בתחילת סעיף זה, וקטור הפרמטרים המיצג את ההגייה, מורכב מקבוצות רכיבים בעלות אורך אחיד, כאשר כל קבוצה מיצגת קטע זמן של ההגייה, ומיקומה הסידורי של קבוצה בוקטור הוא כמיקומו של קטע הזמן, אותו היא מייצגת, במהלך ההגייה. נסמן ב-  $a_{ij}$  את מידת הדמיון בין קבוצת רכיבים המיצגת את קטע הזמן ה-  $i$  של הגייה אחת, לבין קבוצת הרכיבים המיצגת את קטע הזמן ה-  $j$  של הגייה אחרת. האלגוריתם לתאום צירי הזמן מחפש בעזרת תכנות דינמי, בשתי ההגיות המשוות זו לזו, את זוגות קטעי הזמן  $\{i, j\}$  הנותנים לביטוי (2.9) ערך מקסימלי.

$$S_a = \sum_{i,j} a_{ij} \quad (2.9)$$

$S_a$  המקסימלי המתקבל עקב תהליך זה, מהווה את מידת הדמיון בין שתי ההגיות.

תאור מפורט יותר של תהליך תאום צירי הזמן בעזרת תכנות דינמי (DP) יופיע בפרק 7. כאן נציין רק את המסקנות שהתקבלו מניסויים שבדקו את היעילות של שיטת ה-DP.

White and Nelly [6] קבלו של- DP היתה הצלחה גדולה יותר מאשר ל- LTS בזיהוי הגיות בעלות הברות רבות, אך לעומת זאת, לא היה ל- DP כל יתרון על שיטת LTS בזיהוי הגיות בעלות הברה אחת. ניתן להסביר תופעה זאת בכך, שהתאום הלא לינארי של צירי הזמן שבשיטת ה- DP, מצליח להתגבר על אי התאמות בין קטעים פנימיים, השכיחות יותר בהגיות רבות הברות, בעוד שהתאום הלינארי שב- LTS מתקשה להתמודד איתן. גם ב- [13] מדווח על כך שנרמול (תאום) לא לינארי בעזרת תכנות דינמי יעיל יותר מאשר שיטות הנרמול הלינארי.

לפיכך, שיטת ה- DP יעילה במיוחד עבור זיהוי הגיות בעלות הברות רבות, אולם יש לה חסרון בולט - שיטה זו דורשת כמות חישובים גדולה הרבה יותר יחסית לשיטות הפועלות לפי שתי הגישות הראשונות.

חישוב מידת הדמיון בין שתי הגיות

2.3.3

את מידת הדמיון בין שתי הגיות ניתן לחשב בעזרת פונקציות דמיון שונות, כאשר במקרים רבים אפיו של היצוג הפרמטרי קובע את פונקציית הדמיון שתופעל.

White [14] מזכיר שלשה סוגים של פונקציות דמיון: פונקציות קורלציה, פונקציות מסנן ופונקציות של מרחק גיאומטרי. בסעיף זה נגדיר כל אחד מסוגים אלה ונסקור מערכות המשתמשות בו.

1. פונקציות קורלציה. תחילה הגדרה כללית:

יהיו X ו- Y וקטורי פרמטרים בעלי מימד N ויהיו  $x_i, y_i$  רכיביהם בהתאמה. פונקציית הקורלציה בין X ל- Y נתונה ב- (2.10).

$$S(X, Y) = \frac{\sum_{i=1}^N x_i y_i}{\left( \sum_{i=1}^N x_i^2 \right)^{1/2} \left( \sum_{i=1}^N y_i^2 \right)^{1/2}} \quad (2.10)$$

ערך הפונקציה S(X, Y) המתקבל לפי (2.10) והנקרא גם מקדם הקורלציה, מהווה את מידת הדמיון בין שתי ההגיות המיוצגות ע"י הוקטורים X ו- Y. במערכות של Clark [8] ו- Ichikawa [9] משתמשים בפונקציית הקורלציה. במערכת האחרונה נבדקים מספר יצוגים פרמטריים, ובפונקציית הקורלציה



משתמשים עבור היצוגים שאינם מנורמלים - אלה שבערכם המוחלט אינם קטנים המיד מ-1. גם Clapper [12] מבצע במערכת שלו חישוב של פונקציית הקורלציה, אך החישוב ממומש בחומרה ולא בתוכנה כמקובל בשתי המערכות הנ"ל.

2. פונקציות מסנן. פונקציות אלה נקראות בשם זה, מכיוון שהן מתבססות על ה-LPC - מקדמי מסנן כולו קטבים (בעל קטבים בלבד), שהונח כמודל עבור מנגנון יצור הדיבור (ראה 2.2.2).

Gray and Markel [16] דנים בארבע פונקציות מסנן (הן נקראות שם מידות מרחק). Likelihood ratio, cepstral distance, rms log spectral distance ו-cosh measure. מכיוון שכאמור לעיל כולן מבוססות על ה-LPC, הרי שהן מתארות למעשה את מידת הדמיון בין הספקטרום של ההגייה אחת לבין הספקטרום של ההגייה האחרת המושווית עם הראשונה. אולם, בעוד שעבור הפונקציה הראשונה מחשבים את הספקטרום על סמך ה-LPC, עבור האחרות אין עושים זאת, אלא מחשבים אותן בתחום הזמן באמצעות טרנספורמציות שמבצעים על ה-LPC. בצורה זו כל אחת מהפונקציות הנ"ל קובעת למעשה את היצוג הפרמטרי של ההגייה.

גדון כעת בכל אחת מפונקציות המסנן המתוארות ב-[16].

rms log spectral distance מוגדר בביטוי (2.11).

$$\sum_{i=1}^N \left( \ln |H_1(i)|^2 - \ln |H_2(i)|^2 \right)^2 \quad (2.11)$$

כאשר  $H_1(i)$ ,  $H_2(i)$  הן פונקציות התמסורת של שני המסננים (המכילים קטבים בלבד) המתארים קטעי זמן מתאימים בשתי ההגיות המשוות.

$i$  הוא מציין של רכיב ספקטרי מתוך  $N$  הרכיבים המחושבים על סמך ה-LPC שמוצו מקטע הזמן הנדון.

ה-Cepstral distance מבוססת על מקדמי ה-cepstrum המחושבים באופן רקורסיבי מתוך ה-LPC.

פתוח הקשר ביניהם ותכנית החישוב מופיעים ב-[16].

ה- likelihood ratio היא למעשה מידת דמיון המבוטסת על ה- LPR. מידה שהיא הלוגריתם של ה- likelihood ratio פותחה ע"י Itakura ב- [2], (ראה 2.2.2).

ה- cosh measure מתקבלת מתוך ממוצע אריתמטי או גיאומטרי של שני likelihood ratios, כאשר היחס האחד מתאים למצב בו וקטור LPC מסוים הוא וקטור ייחוס האחר הוא הוקטור הנבדק, והיחס השני מתאים למצב שבו הפוכים תפקידי הוקטורים לעומת המצב האחר.

מארבעת הפונקציות הנ"ל, רק ה- log likelihood ratio המבוטסת על ה- LPR נוצלה לצורך חישוב דמיון בין הגיות במערכות לזיהוי מילים בודדות. מלבד Itakura [2] משתמשים בה גם במערכת של White and Neely [6] ושל Coker [11].

3. פונקציות של מרחק גיאומטרי. כאן נדון בשלוש פונקציות המרחק הבאות: מרחק אוקלידי, מרחק ציביציף ומרחק האמינג. תחילה נגדיר אותן. יהיו X ו- Y וקטורי פרמטרים בעלי מימד N, ויהיו  $x_i, y_i$  רכיביהם בהתאמה. מרחק אוקלידי:

$$d(X,Y) = \left( \sum_{i=1}^N (x_i - y_i)^2 \right)^{\frac{1}{2}} \quad (2.12)$$

מרחק ציביציף:

$$d(X,Y) = \sum_{i=1}^N |x_i - y_i| \quad (2.13)$$

מרחק האמינג:

$$d(X,Y) = \sum_{i=1}^N x_i \oplus y_i \quad (2.14)$$

כאשר ב- (2.14) הם מספרים ביצוג בינרי (סיביות) ו-  $\oplus$  מסמן פעולת exclusive or (XOR).

נסקרו מערכות המשתמשות במרחקים גיאומטריים לצורך חישוב הדמיון בין 2 הגיות ותחילה אלה המשתמשות במרחק אוקלידי.

במערכת של Pol's [4] נגזרת מידת המרחק האוקלידי מתוך הלוגריתם של ההסתברות שהפרמטרים של ההגיה הנבדקת ישתייכו לפילוג של הגיית יחוס מסוימת. כאשר מניחים פילוג נורמלי ואי תלות של רכיבי וקטור הפרמטרים זה בזה, מתקבלת כתוצאה מהלוגריתם הני"ל, פונקצית מרחק אוקלידי, שבסכום המופיע בה, כל מחובר מחולק בפיזור של הרכיב מוקטור הפרמטרים של הגיית היחוס, המופיע במחובר.

במערכת של Ichikawa [9], מחשבים מתוך פונקצית המרחק האוקלידי את מידת הדמיון בין 2 הגיות, כאשר מיצגים אותן באמצעות פרמטרים מנורמלים ביחס ל- 1 - מקדמי ה- PARCOR.

גם במערכת של Velichko [10] מחשבים מתוך פונקצית המרחק האוקלידי את מידת הדמיון בין קטע של הגיה אחת לקטע של הגיה אחרת. מידת הדמיון הזאת היא:  $a = \alpha^2 / (\alpha^2 + \rho^2)$ . כאשר  $\rho^2$  הוא ריבוע המרחק האוקלידי בין וקטורי הפרמטרים, ו-  $\alpha$  הוא קבוע שעל סמך נסויים נקבע ל-  $\sqrt{2}$ . טווח ההשתנות של  $a$  הוא בין אפס עבור מרחק אינסופי לבין 1 עבור זהות.

בפונקצית המרחק של צ'ביצ'יף משתמשים במערכת של White and Neely [6]. עושים זאת כאשר וקטור הפרמטרים מתקבל מתוך היציאות של בנק מסננים. במערכת זו ניסו את השימוש במרחק אוקלידי במקום מרחק צ'ביצ'יף, והתקבלו אז תוצאות זהות טובות יותר במקצת. אולם חישוב מרחק אוקלידי דורש יותר זמן ולכן השתמשו במרחק צ'ביצ'יף.

פונקצית המרחק של האמניג שונה מכל פונקציות המרחק הגיאומטרי שהוזכרו עד כה, בכך שהמשתנים שלה מקבלים אך ורק ערכים בינריים. לעומת זאת הערך שמקבלת הפונקציה הזאת אינו בינרי. הטווח שלה הוא בין אפס עבור זהות לבין  $N$  עבור המרחק המכסימלי. כאשר  $N$  הוא המימד של כל אחד מוקטורי הפרמטרים המשווים.

לפיכך, במערכת המתוארת ב- [13] שבה וקטור הפרמטרים הוא וקטור בינרי, משתמשים בפונקצית המרחק של האמניג לצורך חישוב מידת הדמיון בין שתי הגיות.

לסיכום, בטעיף זה הוצגו פונקציות דמיון המשתייכות לשלשה סוגים, ונסקרו מערכות המשתמשות בהן. ניתן להסיק מאופן הבחירה של יוצרי המערכות בפונקציות הני"ל, שבכל מקרה, שואפים לפונקצית דמיון שתתאים ליצוג הפרמטרי שבו משתמשים, וזאת בכדי להגביר את ההצלחה בזיהוי וכן מנסים במידת האפשר להשתמש בפונקציות דמיון הדורשות מינימום חישובים.

פרק 3: שיקולים בפתוח מערכת לזיהוי מילים בודדות הממומשת באמצעות מיני-מחשב.

מבוא

3.1

מטרת המחקר המתוארת בחיבור זה הייתה לפתח מערכת יעילה לזיהוי מילים בודדות, בעלת זמן זיהוי קצר ככל האפשר. לצד המטרה הוצבה גם המגבלה שמימוש המערכת יעשה באמצעות מיני מחשב - בתכנה, ללא אמצעי עזר בחמרה אנלוגית ו/או ספרתית המאפשרים כיום ביצוע זיהוי בזמן אמיתי. עקב מגבלה זאת הוחלט לבדוק את האפשרות לדחוס את ההגיות בזמן לאורך אחיד, באמצעות אלגוריתם יעיל לשינויים בציר הזמן של אותות דיבור [7] שפותח לאחרונה. הדחיסה שנבדקה, מתבצעת על דגימות ההגיה כבר בשלב עיבוד הקדם, לפני מיצוי הפרמטרים, כך שבכל השלבים הבאים בתהליך הזיהוי, מיוצגת ההגיה ע"י כמות מצומצמת יותר של נתונים, וכמות זו היא אחידה עבור כל ההגיות\*.

השאיפה לקיצור זמן החישוב, גרמה לכך שעבור שלבים אחרים של תהליך הזיהוי, נבחרו מתוך השיטות הקיימות אלו המבטיחות ביצוע הנדרש בזמן קצר ככל האפשר. דיון מפורט בשיקולים הקשורים בבחירתן ינתן בסעיפים הבאים של פרק זה.

מכיוון שנוסתה כאן גישה חדשה היה צורך להעריך את יעילותה, וזאת ע"י השוואת ביצועיה לביצועים של מערכת הפועלת לפי גישות מקובלות שיעילותן כבר הוכחה. לשם כך פותחה ומומשה מערכת נוספת שבה לא מתבצעת דחיסה בזמן כנ"ל, ולכן היא מטפלת בהגיות בעלות משך זמן טבעי. במקרה כזה מקובל במערכות קיימות לממש את שלב המיון באמצעות תיכנות דינמי, עקב רמת הצלחתו הגבוהה בזיהוי (ראה 2.3.2) וכך נעשה גם כאן. הדיון במערכת זו מופיע בפרק 7.

בשאר הפרטים של תהליך הזיהוי דומה המערכת הזאת למערכת הכוללת דחיסה, כך שהשוואת הביצועים של שתי המערכות נותנת הערכה נכונה על יעילות הגישה החדשה שפותחה - דחיסת הגיות בזמן.

המערכת החדשה שפותחה אינה תלויה במילון. הדבר מתבטא בכך ששיטת הזיהוי שלה אינה מבוססת על התכונות האקוסטיות של מילות המילון אשר השתתפו בניסויי הזיהוי - עשר הספרות העבריות מאפס עד תשע. המערכת החדשה מאומנת לדובר, כלומר הגיות היחוס עבור הזיהוי נאמרו ע"י הדובר שאת הגיותיו מנסה המערכת לזהות.

\*נעשה אמנם צעד בכיוון זה במערכת של Coker המתוארת ב-[11]. הגישה שם הייתה להשמיט קטעי דיבור יתרים (redundant), על סמך מידת הדמיון שלהם לקטעים השכנים. אולם, חישוב מידת הדמיון דורש ב-[11] מיצוי של מספר פרמטרים עבור כל קטע (אמנם, פחות מאשר לצורכי מיון), וכן לא נפתרה בגישה זו, הבעיה של הפרשי הזמן בין הגיות, במיוחד של מילים שונות.

3.2 השיקולים בבחירת מקדמי הקורלציה החלקית כיצוג פרמטרי עבור ההגייה.

בסעיף 2.2. הדן בשלב מיצוי התכונות, תוארו שתי דרישות הקיימות עבור יצוג פרמטרי המשמש לצורך זיהוי הגיות. הדרישה האחת שיכיל אינפורמציה רלבנטית בלבד, ללא יתרות ובצמצום ככל האפשר, והשניה שאינפורמציה זו תאפשר להפריד נכונה בין ההגיות השונות, כלומר למיין בצורה נכונה את ההגייה הנבדקת. לשתי דרישות אלה נוסף כאן את השלישית - זמן חישוב הפרמטרים שייצגו את ההגייה, יהיה קצר ככל האפשר. דרישה נוספת זו נובעת מהדרישה שהועמדה בפני המערכת - לבצע זיהוי יעיל בזמן קצר ככל האפשר, והיא מקבלת חשיבות יתר כאשר קיימת המגבלה, שהוזכרה בסעיף הקודם, שהמימוש יעשה בתכנה בלבד.

בסעיף 2.2 נסקרים יצוגים פרמטריים שונים שנוסו במערכות קיימות. התוצאות שהתקבלו בניסויים אלה, ושצוינו בסעיף 2.2, מהוות בסיס שעליו אפשר להסתמך כאשר באים לבחור יצוג פרמטרי. כאן נתרכז בתוצאות של שתי עבודות שנעשו בנושא של השוואת הביצועים של יצוגים פרמטריים שונים - עבודתו של Ichikawa [9] ועבודתם של White and Neely [6]. עבודות אלה מטפלות כמעט בכל היצוגים המקובלים, וחשיבותן בכך שהן בודקות יצוגים שונים עבור אותן מילות מילון. תוצאות זיהוי תלויות בתוכן המילוי במידה מסוימת (ראה מסקנותיהם של White and Neely [6]). עבור מילון שחלק ממילותיו דומות זו לזו מבחינה אקוסטית, אחוז ההצלחה בזיהוי יהיה בדרך כלל נמוך יותר מאשר עבור מילון שאין לו התכונה הזאת. לכן השוואת הביצועים של יצוגים פרמטריים שונים עבור מילונים שונים, לא תתרום להערכה נכונה על יעילותו של היצוג בזיהוי הגיות. לא נדון בגישתם של Sambur and Rabiner [3] - יצוג רב פרמטרי עבור ההגייה, מכיוון שנראה שיצוג זה יגזול זמן חישוב רב בשלבי המיצוי והמיון, יחסית ליצוגים שנדון בהם.

Ichikawa [9] בדק את היצוגים הפרמטריים הבאים: הלוגריתם של ספקטרום ההספק לאחר החלקה, הקפסטרום, פונקצית האוטוקורלציה, ה-LPC ומקדמי ה-PARCOR (מקדמי הקורלציה החלקית). לגבי תאור הבדיקה ותוצאותיה ראה סעיף 2.2. המסקנה הנובעת מתוצאות הבדיקה היא שמבין כל היצוגים שנבדקו מקדמי ה-PARCOR הם היעילים ביותר, הם השיגו 100% הצלחה בזיהוי בזמן הקצר ביותר ובעזרת כמות נתונים מינימלית - 6 מימדים לוקטור המייצג קטע זמן של הגייה. (5) מקדמי PARCOR ופרמטר נוסף המהווה גורם שקלול עבור המקדמים של הקטע, והמבוסס על האמפליטודה של הקטע).

White and Neely [6] בדקו את היצוגים הפרמטריים המתקבלים מבנק של 20 מסננים ומבנק של 6 מסננים, ולעומתם את היצוג הפרמטרי המבוטס על ה-LPC בהתאם לגישת Itakura [2] עבור ה-LPR, וכן בדקו את יעילותו של יצוג חסכוני במיוחד מבחינת סיביות לשניה - ה-C.S.Code. תנאי הבדיקה ותוצאותיה מופיעים ב-2.2, ובטבלה 2.1 מרוכזים הביצועים של היצוגים שנבדקו, בהתאם לטבלה זו התקבל שהיצוג המבוטס על ה-LPC משיג תוצאת זיהוי הנחותה רק ב-1% ביחס ליצוג המבוטס על בנק של 20 מסננים, אולם זמן הזיהוי של הראשון הוא 2/3 מזמן הזיהוי של האחרון. חסרון זה של היצוג המבוטס על בנק של 20 מסננים נובע מהעובדה שיצוג זה דורש קצב סיביות גבוה 12000 bits/sec, לעומת 4200 bits/sec הנדרשים עבור היצוג המבוטס על ה-LPC, עובדה חשובה זו תועלה בהמשך. לפי הטבלה, ניתן גם לראות שהיצוג המבוטס על בנק של 6 מסננים דורש זמן זיהוי הקצר יותר ב-25% ביחס לזמן שדורש היצוג המבוטס על ה-LPC, אך זה במחיר של 1% נחיתות בהצלחת הזיהוי. היצוג התמציתי במיוחד - ה-C.S.code, דורש זמן זיהוי קצר ביותר אך הצלחתו בזיהוי נמוכה מאד - 91%.

מתוך עיון בתוצאות שתי העבודות הנ"ל, ניתן להחליט שהיצוגים הפרמטריים היעילים ביותר שמתוכם נבחר את היצוג המתאים ביותר למערכת שעלינו לפתח, הם: מקדמי ה-PARCOR, היצוג המבוטס על ה-LPC בהתאם לגישת Itakura עבור ה-LPR שנכנה אותו בשם יצוג ה-LPR, והיצוג המבוטס על בנק של מסננים. אולם מכיוון שמימוש המערכת הוגבל לתכנה, אין אפשרות להשתמש בבנק של מסננים בחומרה ולכן אם מעונינים לבדוק יצוג המבוטס על בנק של מסננים עבור המערכת, יש לבדוק את המימוש שלו בתכנה. נשווה כעת, מבחינת זמן חישוב הנדרש לזיהוי, בין היצוגים המבוטסים על ה-LPC - מקדמי ה-PARCOR, ויצוג ה-LPR, לבין היצוג המתקבל עקב מימוש בתכנה של בנק מסננים.

נדון תחילה בזמן החישוב הנדרש למיצוי כל אחד מהיצוגים המשווים. עבור היצוגים המבוטסים על ה-LPC מספר הכפלים הנדרש לחישוב המקדמים ליצוג הגייה, בדרך המהירה ביותר, וכאשר משתמשים בחלון שאינו מלבני, (ראה סעיף 5.2) הוא כ- $L_s N_w (p+2)$ , כאשר  $p$  מספר המקדמים,  $N_w$  מספר הדגימות בקטע הזמן ממנו מחשבים את  $p$  המקדמים, ו- $L_s$  הוא מספר קטעי הזמן מהם ממצים את המקדמים. (אם לדייק, לביטוי זה יש להוסיף איברים מסדר  $p^2$  ו- $p$  שאינם פונקציה של  $N$ , ומכיוון ש- $N \gg p$  הם אינם משמעותיים וניתן להזניחם).

מספר החיבורים הנדרש לחישוב המקדמים הנ"ל הוא  $L_s N_w (p+1)$ .

לגבי היצוג המבוטס על מימוש בנק מסננים בתכנה, נתאר בקצרה את המימוש.

בסעיף 4.1, יתואר מימוש של בנק מסננים באמצעות התמרת פורייה עבור קטע זמן קצר (short-time Fourier transform), המנוסחת בצורה של התמרת פוריה הדיסקרטית - DFT, כך שניתן להשתמש באלגוריתם המהיר ה-FFT לצורך מימוש בנק המסננים. בהתאם לפיתוח המופיע בסעיף 4.1, יציאתו של כל אחד מהמסננים בבנק מתוארת עידי הביטוי (4.11) כאשר מציבים בו  $q=1$ . (ערך זה מתאים למצב בו מועבר אות דיבור דרך בנק מסננים ללא ביצוע הזזה של הקוים הספקטרלים שלו, ראה סעיף 4.1). לפיכך, לאחר הצבה זו נקבל

$$y_k(nT) = 2\text{Re}\{G(k,n)\} \quad (3.1)$$

כאשר  $y_k(nT)$  היא היציאה מהמסנן ה- $k$  ברגע  $nT$ ,  $T$  הוא מחזור הדגימה של ההגייה הנבדקת), ו- $G(k,n)$  מוגדרת כך:

$$G(k,n) \triangleq \sum_{r=0}^{N-1} g(r,n) W_N^{kr}, \quad k = 0,1,\dots,N-1 \quad (3.2)$$

כאשר

$$W_N \triangleq \exp [j2\pi/N] \quad (3.3)$$

$N/2$  הוא מספר המסננים בבנק, וכאשר הסדרה  $g(r,n)$ , שה- DFT שלה נותן את  $G(k,n)$  לפי (3.2), היא

$$g(r,n) = \sum_{i=0}^{m-1} s(nT-rT-iNT) h(rT+iNT), \quad r=0,1,\dots,N-1 \quad (3.4)$$

$h(nT)$  שבביטוי (3.4) הוא התגובה להלם של מסנן מעביר נמוכים המשמש כאב טיפוס לכל המסננים האחרים בבנק, ומשכה הוא  $mNT$ .  $s(nT)$  שבביטוי (3.4) הוא אות ההגייה הנבדקת.

המימוש המתואר כאן הוא קרוב לבנק מסננים אידיאליים, הקרוב ילך וישתפר ככל ש-  $m$  יהיה גדול יותר.

מספר הכפלים לחישוב היציאות של המסננים בזמן  $t = nT$  לפי (3.1) הוא,  $mN$  לקבלת  $g(x, n)$  עבור  $x=0, 1, \dots, N-1$ , לפי (3.4), ועוד  $N \log N$  לקבלת  $G(k, n)$  עבור  $k=0, 1, \dots, N-1$ , לפי (3.2). מכיוון שיציאת המסנן משתנה עם הזמן באופן איטי יחסית לאות הכניסה, אין צורך לדגום אותה בכל  $T$  וניתן להסתפק בדגימה בקצב איטי יותר, אך מחזור הדגימה המינימלי, לפי משפט הדגימה, הוא  $NT$  (כי רוחב הסרט של יציאה המסנן הוא  $2\pi/NT$ ). לפיכך, עבור הגייה המכילה  $L_{c_0}$  דגימות, מספר הכפלים לחישוב יציאות המסננים, כאשר מחזור דגימתן הוא  $NT$ , יהיה  $(L_{c_0}/N) N(m+\log N)$ , כלומר  $(m+\log N)L_{c_0}$ , וזהו גם מספר החיבורים הנדרש למען אותו חישוב.

כעת בכדי לבצע השוואה כמותית בין זמני החישוב הנדרשים למיצוי כל אחד מהפרמטרים המשווים, נציב ערכים מספריים מקובלים עבור כל אחד מהמשתנים המופיעים בביטויים שקבלנו.

עבור היצוגים המבוטסים על ה- LPC קבלנו שמספר הכפלים הנדרש לחישוב הפרמטרים עבור ההגייה הוא כ-  $L_S N_W(p+2)$ . נציב בביטוי זה את הערכים המקובלים במערכת הנדונה בחיבור זה. לפיכך נציב  $p=6$  שהוא ערך מקובל על סמך העבודות של [2] Itakura ו- [9] Ichikawa, וכן נציב  $N_W=300$  ו-  $L_S=19$ .  $L_S=19$  מתקבל עקב העובדה שכל ההגייה לאחר דחיסה מכילה  $L_{c_0}=3000$  דגימות, וכל 150 דגימות מחשבים פרמטרים מקטע זמן המכיל  $N_W=300$  דגימות. השיקולים הקשורים בקביעת ערכים אלה יובאו בסעיף 6.3. לאחר ההצבה נקבל שדרושים 45,600 כפלים. נציב ערכים אלה גם בביטוי עבור מסי החיבורים הנדרש:  $L_S N_W(p+1)$  ונקבל שדרושים 39,900 חיבורים.

עבור היצוג המתקבל מתוך מימוש בנק מסננים בתכנה, דרושים כאמור  $L_{c_0}(m+\log N)$  כפלים וכך גם לגבי החיבורים. אם נציב  $N=32$  נקבל בנק המכיל 16 מסננים, כמות המבטיחה תוצאות טובות לפי White and Neely [6]. בכדי להשיג קרוב טוב של המימוש בתכנה לבנק מסננים/אידיאליים, נבחר  $m=10$ , כלומר משך הזמן עבור  $h(nT)$  יהיה  $320T$ ,  $L_{c_0}=3000$  כנ"ל. נציב כל אלה בביטוי הנ"ל, ונקבל שדרושים 45000 כפלים, וכאמור לעיל, מספר החיבורים הנדרש גם הוא 45000.

לפיכך, קבלנו שזמן מיצוי הפרמטרים כמעט זהה עבור כל היצוגים.



נעבור כעת לדון בשלב הבא של תהליך הזיהוי - שלב המיון, שבו יקבע מי היצוג החסכוני ביותר מבחינת זמן חישוב.

בשלב זה, עיקר זמן החישוב מוקדש לחישוב מידות הדמיון בין וקטור הפרמטרים המיצג את ההגיייה הנבדקת, לבין כל אחד מוקטורי הפרמטרים המיצגים את מילות המילון - וקטורי הייחוס. במערכת הנדונה המפעילה דחיסה, וקטורי הייחוס והוקטורים המיצגים הגיות נבדקות הם בעלי אותו אורך, לכן אפשר לחשב את מידת הדמיון הנ"ל מתוך סכום הערכים המוחלטים של ההפרשים בין הרכיבים של שני הוקטורים המשווים - על פי ההגדרה של מרחק צ'ביצ'יף בביטוי (2.13). השיקולים הקשורים בבחירת פונקציית דמיון לצורך חישוב מידת הדמיון ינתנו בסעיף הבא. חישוב מידת הדמיון בדרך זו דורש רק חיבורים (חיבורים נחשבים כחיבורים). כמות החיבורים הנדרשת לצורך חישוב מידת הדמיון בין שתי הגיות היא  $9L_R$ . כאשר  $L_R$  הוא אורכו של וקטור הייחוס. הגורם 9 נובע מכך שעבור ערך מוחלט של הפרש בין 2 רכיבים דרושים 2 חיבורים,  $L_R$  חיבורים נוספים דרושים לביצוע הסיכום, ועל פעולות אלה חוזרים 3 פעמים. התחלת וקטור הפרמטרים של ההגיייה הנבדקת שארכו  $L_T$ , המקיים  $L_R + 2\Delta_S = L_T$ , מוזזת יחסית להתחלת וקטור הייחוס ב-  $q\Delta_S$  רכיבים, כאשר בכל פעם  $q$  מקבל ערך אחר משלושת הערכים הבאים: 0, 1, 2. הסיבה לפעולה האחרונה תובא בסעיף הבא. חישוב מידת הדמיון בדרך הזאת יעשה עבור היצוג המבוסס על מימוש בתכנה של בנק מסננים ועבור מקדמי ה- PARCOR. עבור יצוג ה- LPR קיימת פונקציית זמרחק המתאימה ליצוג זה [2], וחישוב מידת הזמרחק בהתאם לפונקציה זו יעשה בהמשך. כעת נטפל בשני היצוגים הראשונים.

עבור מקדמי ה- PARCOR, נציב את הערכים המקובלים במערכת שפותחה:  $L_T = pL_S = 114$ ,  $\Delta_S = p = 6$ ,  $L_R = 102$  ומכאן נובע שמספר החיבורים הנדרש עבור היצוג הזה היא  $9L_R = 918$ . כך לגבי השוואה בין 2 הגיות, ואם במילון 10 מילים (כפי שקיים במערכת שפותחה) אזי מספר החיבורים לצורך השוואה בין מייצגי הגייה נבדקת לבין מייצגי מילות המילון הוא 9180. כאשר היצוג הפרמטרי מבוסס על מימוש בתכנה של בנק מסננים, עם הנתונים שנקחו בחישובים עבור זמן המיצוי:  $N = 32$  (16 מסננים), נקבל  $L_T = \{[(3000-320)/32]+1\} \cdot 16 = 1344$ . נציב  $\Delta_S = 80$  המתאים להצבה של  $\Delta_S = 6$  עבור מקדמי ה- PARCOR.  $\Delta_S = 80$  רכיבים ביצוג של בנק מסננים מייצגים 160 דגימות של ההגיייה הדחוסה, ו-  $\Delta_S = 6$  מקדמי ה- PARCOR מייצגים 150). לפיכך נקבל  $L_R = L_T - 2\Delta_S = 1184$ . ומכאן נובע שמספר החיבורים לצורך השוואה בין מייצגי ההגיייה הנבדקת לבין מייצגי מילות המילון הוא

עבור היצוג הפרמטרי המבוטס על מימוש בתכנה של בנק מסננים.  $90L_R = 106,560$

לבסוף, נדון ביצוג ה-LPR. יצוג זה קובע את פונקציית המרחק שמשמשים בה לחישוב מידת המרחק בשלב המיון. לפי [2] פונקציית המרחק הזאת מוגדרת כך,

$$d(k,j) = \log \left[ \sum_{i=0}^p (a_i^k)^2 \right] + \log \left[ \left( 2 + \sum_{i=1}^p b_i^{k,j} r_i^j \right) / \left( \sum_{i=0}^p a_i^j r_i^j \right) \right] \quad (3.5)$$

כאשר  $d(k,j)$  מביעה את המרחק בין קטע זמן של הגיית היחוס ה- $k$  לבין קטע זמן של ההגייה הנבדקת ה- $j$ .  $\{r_i^j\}$ ,  $0 \leq i \leq p$ , הם מקדמי האוטוקורלציה לפי (5.12) המנורמלים ב- $R(0)$ , של קטע הזמן מההגייה הנבדקת,  $\{a_i^j\}$ ,  $0 \leq i \leq p$ , הם ה-LPC עבור אותו קטע הזמן.

הם מקדמי האוטוקורלציה של ה-LPC המתקבלים מקטע הזמן של הגיית הייחוס, כלומר

$$b_i^k = \left( 2 \sum_{j=0}^{p-i} a_j^k a_{j+i}^k \right) / \left[ \sum_{j=0}^p (a_j^k)^2 \right], \quad 0 \leq i \leq p \quad (3.6)$$

כאשר  $\{a_j^k\}$ ,  $0 \leq j \leq p$ , הם ה-LPC המתקבלים מקטע הזמן של הגיית היחוס ה- $k$ . לפיכך, לשם קבלת וקטור היחוס ה- $k$  יש לחשב את הביטויים

$$\log \left[ \sum_{i=0}^p (a_i^k)^2 \right] - \log \left[ \sum_{i=0}^p (b_i^k)^2 \right], \quad 0 \leq i \leq p, \{b_i^k\}$$

נחשב את מספר הפעולות (כפלים וחילבורים) שיש לבצע לצורך חישוב מידת הדמיון בין שתי הגיות באמצעות (3.5). במערכת הכוללת דחיטה, בכל הגייה יש אותו מספר של קטעי זמן, לכן המרחק בין שתי הגיות ינתן כסכום המרחקים בין קטעי הזמן המתאימים של שתי ההגיות. לפיכך, לחישוב (3.5) דרושים:  $p+2$  כפלים ו- $p+1$  חילבורים, כאשר חילוק נחשב ככפל וה- $\log$  מקורב ב- [2]

ע"י כפל וחיבור. הביטוי שמופיע במכנה של (3.5) מחושב כמוצר לוואי עוד בשלב המיצוי (הוא אינו מחושב מחדש עבור כל הגיית יחוס חדשה). גם עבור יצוג זה מבצעים את חישוב המרחק בין שתי ההגיות, שלש פעמים. מספר קטעי הזמן של הגייה הנבדקת הוא  $L_S$  ושל הגיית הייחוס הוא  $L_S - 2$  ובכל פעם מוזנת התחלת ההגייה הנבדקת ב-  $i$  קטעי זמן ביחס להתחלת הגיית היחוס, כאשר  $i$  מקבל כל פעם ערך אחר מהערכים 0, 1, 2. לפיכך, עבור יצוג זה דרושים לחישוב מידת המרחק בין שתי הגיות  $3(L_S - 2)(p + 2)$  כפלים ו-  $3(L_S - 2)(p + 3)$  חיבורים. ובהצבת הערכים המקובלים במערכת שפותחה,  $p = 6$ ,  $L_S = 19$ , נקבל שמספר הכפלים הנדרש הוא 408. ומספר החיבורים הוא 408. עבור מילון המכיל 10 מילים נקבל שדרושים 4080 כפלים ו- 34080 חיבורים. כעת לצורך השוואה, נניח שכפל אחד שקול לשלשה חיבורים מבחינת זמן חישוב בנקודה צפה, ונקבל שעבור יצוג ה- LPR דרושים 16,320 חיבורים.

נדרג כעת את היצוגים שהשוונו לפי מספר חיבורים עולה: מקדמי ה- PARCOR - 9180, יצוג ה- LPR - 16,320, והיצוג המבוסס על מימוש בתכנה של בנק מסננים - 106,560\*. המסקנה הנובעת מתוצאות אלה היא שמקדמי ה- PARCOR מהווים את היצוג הפרמטרי החסכוני ביותר מבחינת זמן חישוב. מבחינת הצלחה בזיהוי כל אחד מהיצוגים האלה הוכיח את עצמו במערכת שלו, כלומר, מקדמי ה- PARCOR השיגו 100% הצלחה בזיהוי במערכת של Ichikawa [9], ויצוג ה- LPR ובנק של מסננים השיגו 97% ו- 98% בהתאמה במערכת של White and Neely [6]. לכן, כאשר לוקחים בחשבון את הדרישה המרכזית שהועמדה בפני המערכת - זמן זיהוי קצר ככל האפשר, מקדמי ה- PARCOR (מקדמי הקורלציה החלקית) הם הנבחרים כיצוג פרמטרי עבור ההגייה

---

\* יש לשים לב לעובדה החשובה שהיצוג הפרמטרי המבוסס על מימוש בתכנה של בנק מסננים מייצג כל אחת ממילות המילון בעזרת וקטור בעל 1184 רכיבים לעומת 102 רכיבים בלבד בוקטור של מקדמי ה- PARCOR. לכן היצוג הראשון דורש הרבה יותר זכרון מאשר האחרון (פי 11 בערך). זה חסרון אשר יש לו משמעות מיוחדת עבור מיני מחשב עם זכרון בעל קיבול קטן, כמו למשל NOVA 2 שהזכרון שלו מכיל 32K מילים בלבד, כי בתנאים אלה אין אפשרות לאחסן בזכרון את וקטורי היחוס של היצוג הראשון, עוד לפני תחילת תהליך הזיהוי, כפי שניתן לעשות עבור האחרון, אלא נאלצים לעשות זאת בעיצומו של התהליך. כלומר; בזמן התהליך, תכנית הזיהוי צריכה לקרוא את וקטורי היחוס מה-DISK אל הזכרון, ופעולה זו גוזלת זמן רב, כך שזמן הזיהוי כאשר משתמשים ביצוג הראשון עולה בהרבה על זמן הזיהוי כאשר משתמשים במקדמי ה- PARCOR.

### 3.3 שיקולים בפיתוח תהליך המיון

הנושא הראשון בפיתוח תהליך המיון הוא קביעת שיטת המיון. מכיוון שנקבע שהמערכת הנדונה תהיה מאומנת לדובר, הוחלט לבחור בשיטה של התאמת הדפסים שהיא פשוטה וחסכונית יותר בזמן חישוב מאשר השיטה של מיון בעזרת עץ החלטה (חאור השיטות האלה מופיע בסעיף 2.3.1), ונתנה תוצאות טובות במערכות קיימות המאומנות לדובר. לדוגמא במערכות של Itakura [2], White and Neely [6], Ichikawa [9], ו-Coker [11].

אין הכרח לבצע התאמת צירי הזמן (ראה סעיף 2.3.2) עבור המערכת הכוללת דחיסה. זאת מכיוון שלאחר הדחיסה בזמן של ההגייה, בשלב עיבוד הקדם, מתקבל עבור כל הגייה נבדקת, בשלב מיצוי תכונות, וקטור פרמטרים באורך האחיד לגבי כל ההגיות, כך שוקטור ההגייה הנבדקת שווה בארכו לכל אחד מוקטורי הייחוס. לפיכך, דחיסת ההגייה בזמן תרמה מלבד לחסכון ניכר בזמן חישוב, כפי שיוכח בהמשך החיבור, גם לפתרון בעית השוני במשכי הזמן של ההגיות.

כעת יש להחליט על פונקציית דמיון מתאימה לצורך חישוב מידת הדמיון בין וקטור ההגייה הנבדקת לבין וקטורי הייחוס. נבדוק, לשם כך, את מספר החישובים הנדרשים עבור מספר פונקציות דמיון.

עבור פונקציית הקורלציה המוגדרת בביטוי (2.10) דרושים  $3L_R$  כפלים ו- $3L_R$  חיבורים לחישוב מידת הדמיון בין שתי הגיות.  $L_R$  הוא כמוגדר בסעיף 3.2, ארכו של וקטור הייחוס. הגורם 3 נובע מכך שחישוב ערך הפונקציה מבוצע 3 פעמים בדרך המתוארת בסעיף 3.2 עבור מקדמי ה-PARCOR. לשם השוואה, נניח שכפל אחד אקוילנטי מבחינת זמן חישוב בנקודה צפה ל-3 חיבורים, ונקבל שדרושים  $12L_R$  חיבורים לחישוב מידת הדמיון בין שתי הגיות באמצעות פונקציית הקורלציה.

עבור פונקציית המרחק האוקלידי המוגדרת בביטוי (2.12), דרושים  $3L_R$  כפלים ו- $6L_R$  חיבורים לחישוב מידת המרחק בין שתי הגיות. גם כאן מבוצע חישוב ערך הפונקציה 3 פעמים כמתואר בסעיף 3.2 בהקשר למקדמי ה-PARCOR. נמיר גם כאן כפל אחד ב-3 חיבורים ונקבל שדרושים  $15L_R$  חיבורים לחישוב מידת המרחק בין שתי הגיות.

כעת נציב  $L_R = 102$  המקובל במערכת הנדונה, בכל הביטויים הנ"ל, ונכפול ב- 10 בכדי לקבל את מספר החיבורים הדרושים לחישוב מידת הדמיון עבור 10 וקטורי היחוס של המילון. עתה נדרג את פונקציות הדמיון הנ"ל וכן את פונקצית המרחק של ציבצ'יף לפי מספר חיבורים עולה. (הנתונים עבור הפונקציה האחרונה מופיעים בטעיף 3.2): מרחק ציבצ'יף - 9180, פונקצית הקורלציה - 12,240, מרחק אוקלידי - 15300.

המסקנה הנובעת מתוצאות אלה הוא שמרחק ציבצ'יף הוא החסכוני ביותר מבחינת זמן חישוב. לגבי יעילות פונקציה זו למטרות זיהוי, נמסר ב- [6] שהיא נתנה תוצאות פחות טובות מאשר פונקצית מרחק אוקלידי. אולם על סמך ניסויי זיהוי במערכת שפותחה, עבור שתי פונקציות אלה, התברר שמבחינת הצלחה בזיהוי אין לפונקצית המרחק האוקלידי כל יתרון. לבסוף, נבחרה לכן, פונקצית המרחק של ציבצ'יף - החסכונית ביותר מבחינת זמן חישוב.

בטעיף 3.2 תוארה כבר שיטת חישוב הדמיון בין שני וקטורים שנכנה אותה שיטת ההזזה. חישוב מידת הדמיון נעשה לפי שיטה זו שלוש פעמים, כאשר בכל פעם נמצאת המחלת וקטור הפרמטרים של הגיית היחוס (הרכיב הראשון שלו), מול רכיב אחר של וקטור הנבדקת. בפעם הראשונה מול הרכיב הראשון, בפעם השניה מול הרכיב השביעי ובאחרונה מול ה- 13. כלומר בכל פעם הזזה ב- 6 רכיבים המיצגים קטע זמן אחד. בכך מנסים להתגבר על אי דיוקים בקביעת הקצוות של ההגיות, העלולים לשבש את הזיהוי, ולכן משלוש מידות הדמיון המתקבלות לוקחים את המכסימלית לצורך המיון. שיטה זו נוסחה ב- [6].

יש לציין שניתן היה להפעיל כאן את שיטת המיון המבוססת על תכנות דינמי (ראה סעיף 2.3.2 ופרק 7), בכדי לחשב את מידת הדמיון הדרושה. שיטה זו אמורה להשיג תאום טוב יותר בין חלקים פנימיים של שתי ההגיות המושוות ולכן לתת תוצאות טובות יותר. בניסויים שנערכו במערכת שפותחה (הכוללת דחיסה), התברר ששיטת המיון המבוססת על תכנות דינמי לא השיגה כל יתרון מבחינת הצלחה בזיהוי של הגיות דחוסות, ביחס לשיטת ההזזה שתוארה קודם. ניתן להסביר זאת בכך, שבמילון עבורו נערכו הניסויים (עשר הספרות העבריות מאפט עד תשע) אין מילה עם יותר משתי הברות, ולפי המסקנה של White and Neely [6], שיטת המיון המבוססת על תכנות דינמי מתבלטת מבחינת הצלחה בזיהוי, רק כאשר מילות המילון הן רבות הברות. לכן במקרה שלנו היא לא השיגה כל יתרון. על סמך התוצאה הזאת ומכיוון ששיטת המיון בעזרת תכנות דינמי דורשת יותר חישובים מאשר שיטת ההזזה, הועדף להפעיל בשלב המיון, במערכת הכוללת דחיסה, את שיטת ההזזה.

פרק 4: דחיסת ההגיייה בזמן

תאור האלגוריתם 4.1

בצעד השני של שלב עיבוד הקדם, מתבצעת דחיסה בזמן של ההגיייה לאורך האחיד עבור כל ההגיות. דחיסה זו מבוצעת באמצעות אלגוריתם יעיל לשינוי בתחום הזמן של ציר הזמן של אותות דיבור [7]. נביא בסעיף זה תאור קצר של פתוח האלגוריתם (תאור מפורט מופיע ב- [7]). אמנם האלגוריתם לביצוע דחיסה בזמן (TSC-time scale compression) מבוצע בתחום הזמן, אך הפיתוח שלו מקורו בתחום התדר [7]. הרעיון הוא לבצע חלוקה בתדר של אות הדיבור בגורם C (ללא שינוי בציר הזמן) ואחר כך עבור האות שתקבל לבצע מכפלה בתדר בגורם C ע"י השמעתו במהירות הגבוהה פי C מאשר מהירות ההקלטה של האות המקורי. התוצאה היא אות דיבור בעל ספקטרום כשל האות המקורי, אך משכו קצר פי C יחסית לאות המקורי. C הוא רציונלי, לכן  $C \approx a/b$  כאשר  $a, b$  שלמים,  $a > b$ , כי  $C > 1$ .

הבעיה היא לממש את החלוקה בתדר של אות הדיבור<sup>†</sup>. ניתן לקרב אותה כאשר מסתמכים על התכונה שאות הדיבור הקולי הוא קואזי-מחזורי. לפי קרוב זה מעבירים את אות הדיבור דרך בנק של L מסננים מעבירי תחום זהים, כל מסנן סביב תדר מרכזי מסוים כך שכולם יכסו את תחום הספקטרום של אות הדיבור. את היציאה של כל מסנן מזיזים בתדר כלפי מטה לפי הגורם C הנ"ל, ולבסוף מסכמים את כל היציאות. קרוב זה לחלוקה בתדר יהיה מוצלח יותר ככל שהמסננים יקיימו בצורה נאמנה יותר את הדרישות הבאות - יהיו צמודים זה לזה אך לא חופפים, ומספיק צרים על מנת שיעבירו לא יותר מאשר קו ספקטרי בודד של אות הדיבור. קיום הדרישה האחרונה מותנה במידה בה ההנחה שאות הדיבור הקולי הוא אות קואזי-מחזורי, מתאימה למציאות\*. נתאר דרך למימוש הקרוב הנ"ל.

אם המסננים הנ"ל מתקבלים כתוצאה מהזזה בתדר של מסנן מעביר נמוכים כאב-טיפוס (prototype lowpass filter), אזי כל אחד מהם ניתן לתאור ע"י:

$$h_k(nT') = 2h_0(nT') \cos(\omega_k nT'), \quad k=1, \dots, L \quad (4.1)$$

<sup>†</sup> מבלי לשנות את משכו.

\* עבור דיבור לא קולי לא קיימת הבעייה של טיב הקרוב הנ"ל, מכיוון שדיבור זה נושא אופי של רעש - אינו מחזורי, ולכן גם אם ההזזה בתדר אינה מצליחה לקרב את החלוקה בתדר, הדבר אינו פוגם בדיבור המתקבל.

כאשר  $h_0(nT') = h(nT')$  הוא המסנן המשמש כאב-טיפוס ממנו מתקבלים האחרים ע"י הזזה אל התדרים  $\pm\omega_k = \pm k\Delta\omega$ . נבחר כך שבנק המסננים יכסה את תחום התדרים  $[-\pi/T', \pi/T']$ . כאשר  $T' \triangleq T/b$ . הנבחר כך שכמעט כל האינפורמציה הספקטרלית של אות הדיבור תמצא מתחת לתדירות  $1/2T$ .

כעת נניח שידועים ערכי הדגימות של אות הדיבור  $s(t)$  כאשר הוא דגום במחזור  $T$  ובמחזור  $T'$ , וכן ש-  $h(t)$  נתון כפונקציה רציפה. קונבולוציה דיסקרטית בין אות הדיבור בכניסה למסננים  $s(nT')$  לבין המסנן ה-  $k$  נותנת את יציאת המסנן ה-  $k$ ,  $y_k(nT')$  [21]:

$$y_k(nT') = 2 \cdot \sum_{r=-\infty}^n s(rT')h(nT'-rT')\cos[\omega_k(nT'-rT')] =$$

$$= 2\text{Re} \{ \exp[j\omega_k nT'] S(\omega_k, nT') \} \quad (4.2)$$

כאשר:

$$S(\omega_k, nT') \triangleq \sum_{r=-\infty}^n s(rT')h(nT'-rT')\exp[-j\omega_k rT'] \quad (4.3)$$

נקרא בשם התמרת פוריה הדיסקרטית לזמן קצר

(discrete short-time Fourier transform) של אות הכניסה  $s(nT')$ . כלומר ברגע  $S(\omega_k, nT')$  הוא התמרת פוריה המחושבת בתדירות  $\omega_k$  של דגימות העבר משוקללות באמצעות  $h(nT'-rT')$ .

כעת בכדי לממש את ההזזה בתדר כלפי מטה של יציאת כל מסנן, יש לכפול את  $S(\omega_k, nT')$  ב-  $\exp[j\omega_k qnT']$ , כאשר  $q = 1/C$  ו-  $C > 1$  הוא גורם החלוקה אותה מקרבים באמצעות ההזזה. לאחר פעולה זו נקבל את  $y_k^q(nT')$  יציאת המסנן המוזזת בתדר לפי הגורם  $C$ :

$$y_k^q(nT') = 2\text{Re}\{\exp[j\omega_k qnT'] S(\omega_k, nT')\} \quad (4.4)$$

סכום של  $y_k^q(nT')$  עבור  $k=1, \dots, L$  יתן אות יציאה  $y^q(nT')$  הנמשך אותו פרק זמן כאות הכניסה  $s(nT')$  אך התחום הספקטרולי שלו צר פי  $C$  ביחס לתחום הספקטרולי של אות הכניסה. סכימה של המימוש שתואר כאן מופיעה ב- [7].

כאן מעונינים לבצע את המימוש הנייל בתכנה, ניתן לחסוך בחישובים אם מנצלים את העובדה שכל המסגנים זהים ומממשים אותם באמצעות האלגוריתם של ה-FFT. לשם כך יש לבטא את (4.3) בצורה המתאימה ל-DFT. הדבר נעשה כך. ראשית יש לבצע ב- (4.3) חילופי משתנים בסדר הבא:  $r$  מוחלף ב-  $n-r'$ , ואז  $r'$  יוחלף ב-  $r$ . לאחר זאת נהפוך את סדר הסיכום ונקבל מ- (4.3):

$$S(\omega_k, nT') = \sum_{r=0}^{\infty} s(nT'-rT') h(rT') \exp[j\omega_k rT'] \exp[-j\omega_k nT'] \quad (4.5)$$

נגדיר  $N'_p \triangleq N_p b$ .

כעת נבטא בתחום הזמן את תחום ההשתנות של  $r$  כמורכב מאינסוף קבוצות שבכל אחת מהן  $N'_p$  נקודות ובתחום בתדר נחלק את תחום ההשתנות  $-\pi/T' \leq \omega \leq \pi/T'$  ל-  $N'_p$  פסים שווים כל אחד ל-  $2\pi/N'_p T'$  כך ש-  $\omega_k = k\Delta\omega = 2\pi k/N'_p T'$ . בהתאם לכך נציג את הביטויים האקספוננציאליים בצורה הבאה:

$$\exp [j\omega_k rT'] = \exp [j2\pi kr/N'_p] = W_{N'_p}^{kr} \quad (4.6)$$

כאשר

$$W_{N'_p} \triangleq \exp [j2\pi/N'_p] \quad (4.7)$$

נכניס את כל השינויים האלה ב- (4.5), נעזר במחזוריות של (4.6) (במחזור  $N'_p$ ) ונקבל:

$$S(\omega_k, nT') = \left\{ \sum_{r=0}^{N'_p-1} \left[ \sum_{i=0}^{\infty} s(nT'-rT'-iT'_p) h(rT'+iT'_p) \right] W_{N'_p}^{kr} \right\} W_{N'_p}^{-kn} \quad (4.8)$$

כאשר  $r = 0, 1, \dots, N'_p-1$  ;  $k = 0, 1, \dots, N'_p-1$



הביטוי שבטורגריים [ ] הוא פונקציה של  $r$  ו- $n$  ויסומן כ-  $g(r,n)$ .  
 הביטוי שבטורגריים { } הוא פונקציה של  $k$  ו- $n$  ויסומן כ-  $G(k,n)$ .  
 כלומר,

$$G(k,n) \triangleq \sum_{r=0}^{N'_p-1} g(r,n) W_{N'_p}^{kr}, \quad k=0, \dots, N'_p-1 \quad (4.9)$$

פרושו של דבר ש-  $G(k,n)$  לפי (4.9), הוא ה-DFT של  $g(r,n)$ , כך שאפשר לחשבו בעזרת FFT. לצורך קבלת  $g(r,n)$  מניחים ש-  $h(nT')$  הוא בעל משך סופי כלומר, מסנן מסוג finite duration impulse response-FIR, המקיים שעבור  $n > mN'_p$  ו-  $n < 0$   $h(nT')=0$ . לפיכך  $g(r,n)$  תהיה:

$$g(r,n) = \sum_{i=0}^{m-1} s(nT' - rT' - iN'_p T') h(rT' + iN'_p T'), \quad r=0, 1, \dots, N'_p-1 \quad (4.10)$$

נציב את (4.7), (4.8) ו- (4.9) ב- (4.4) ונקבל:

$$y_k^q(nT') = 2\text{Re}\{W_{N'_p}^{nk(q-1)} G(k,n)\}, \quad k=0, 1, \dots, N'_p-1 \quad (4.11)$$

לפי (4.11) מספר הכפלים שנדרש לחישוב דגימת יציאה אינו קטן מ-  $N'_p(m+2)$  (ראה [7]). האלגוריתם החדש שפותח ב- [7] מקטין את המספר הזה ל-  $m$ . נמשיך בפיתוח המוביל לקבלת הביטוי שבו מושג הדבר. מאחר ותחום התדר  $-\pi/T' \leq \omega \leq \pi/T'$  חולק ל-  $N'_p$  פסים, נסכם לפי (4.11)  $L = (N'_p-1)/2$  מסננים ונוסיף את הפס שמרכזו ב-  $\omega=0$ . (זה בהנחה ש-  $N'_p$  אי-זוגי. עבור  $N'_p$  זוגי,  $L = N'_p/2$  והפס סביב  $\omega=0$  אינו קיים, אך התוצאות הסופיות עבור 2 המקרים זהות). תוצאת הסיכום תהיה [7]:

$$y^q(nT') = G(0,n) + 2\text{Re}\left\{ \sum_{k=1}^{(N'_p-1)/2} W_{N'_p}^{nk(q-1)} G(k,n) \right\} \quad (4.12)$$

נבצל את התכונה שסכום משתנה קומפלכסי והצמוד שלו נותן פעמיים החלק הממשי של המשתנה הקומפלכסי, ונרשום את (4.12) כך:

$$y^q(nT') = \sum_{k=0}^{N'_p-1} G(k,n) W_{N'_p}^{nk(q-1)} + A_q(nT') \quad (4.13)$$

כאשר

$$A_q(nT') = \sum_{k=1}^{(N'_p-1)/2} G^*(k,n) W_{N'_p}^{-nk(q-1)} - \sum_{k=\left(\frac{N'_p-1}{2}+1\right)}^{N'_p-1} G(k,n) W_{N'_p}^{nk(q-1)} \quad (4.14)$$

כעת לגבי המחובר השני ב- (4.14), נהפוך את סדר הסיכום, נחליף משתנים  $N'_p - k$  ב-  $k$ , ואחר את  $k$  ב-  $k$ , נבצל את תכונת המחזוריות של  $W_{N'_p}$  לפי (4.7), ולבסוף מאחר ו-  $g(r,n)$  לפי (4.10) היא ממשית, אזי  $G(N'_p - k, n)^p$  ניתן להחלפה ב-  $G^*(k,n)$ . לאחר כל זאת נקבל

$$A_q(nT') = [1 - W_{N'_p}^{nN'_p q}] \sum_{k=1}^{(N'_p-1)/2} G^*(k,n) W_{N'_p}^{-nk(q-1)} \quad (4.15)$$

עתה אנו מעוניינים לאפס את  $A_q(nT')$ . כאשר  $A_q(nT') = 0$  כאשר  $W_{N'_p}^{nN'_p q} = 1$ . את זאת נשיג ע"י דגימה  $y^q(nT')$  כל  $aT'$ , כלומר נחליף  $n$  ב-  $na$  ונקבל ש-

$$W_{N'_p}^{na(b/a)N'_p} = 1$$

לאחר הפעולה האחרונה נקבל מ- (4.13):

$$y^q(naT') = \sum_{k=0}^{N'_p-1} G(k,na) W_{N'_p}^{-nk(a-b)} \quad (4.16)$$

נגדיר משתנה שלם חדש  $p_a$ :

$$p_a \triangleq n(a-b) \pmod{N'_p}, \quad 0 \leq p_a \leq N'_p-1 \quad (4.17)$$

נציב את  $p_a$  ב- (4.16) ונקבל שביטוי זה מהווה את ה-DFT ההפוכה של  $G(k, na)$  עבור  $r=p_a$ , מוכפלת ב-  $N'_p$ , כלומר לפי (4.9):

$$y^q(naT') = N'_p g(p_a, na) \quad (4.18)$$

לפיכך, קבלנו עבור  $y^q(naT')$  ביטוי הניתן לחישוב בתחום הזמן. כעת נפתח את (4.18) לצורה נוחה יותר לחישוב. נגדיר את  $N_{sc}$ :

$$N_{sc} \triangleq \lceil [N'_p / (C-1)] \rceil \quad (4.19)$$

כאשר סוגריים כפולים מסמנים את השלם הקרוב ביותר למנה שבתוכם, וכן נגדיר את  $\alpha$  כשלם הגדול ביותר במנה  $n/N_{sc}$ , כלומר,

$$\alpha \triangleq \text{Int} [n/N_{sc}] \quad (4.20)$$

על סמך (4.20) נרשום את (4.17) כ-

$$p_a = n(a-b) - \alpha N'_p \quad (4.21)$$

נציב את (4.10) ו- (4.21) ב- (4.18) ונקבל

$$y^q(naT') = \sum_{i=0}^{m-1} s(nbT' + \alpha N'_p T' - iN'_p T') \hat{h}[(n(a-b) \bmod N'_p)T' + iN'_p T'] \quad (4.22)$$

כאשר  $\hat{h}(t) \triangleq N'_p h(t)$

לסיום נסתמך על הקשרים  $T' \triangleq T/b$ ,  $N'_p \triangleq N_p b$  ו-  $C = a/b$ .

על סמך (4.19) נקבל \* ש-  $N_p T = N_{sc} T(C-1)$  ו-  $(n(a-b) \bmod N_p) T = (n \bmod N_{sc}) T(C-1)$  נציב ביטויים אלה ב- (4.22) ונקבל את הביטוי עבור סדרת היציאה הדחוסה בזמן:

$$y^q(nCT) = \sum_{i=0}^{m-1} s(nT + \alpha N_p T - i N_p T) \hat{h}[(n \bmod N_{sc}) T(C-1) + i N_{sc} T(C-1)] \quad (4.23)$$

יש לציין שעבור חישוב (4.23) דרושים ערכי הדגימות של  $s(t)$  אך ורק במחזור דגימה  $T$ . לעומת זאת, את  $\hat{h}(t)$  הנחון כפונקציה רציפה יש לדגום במחזור  $T(C-1)$ .

לפי (4.23) מספר הכפלים הנדרש לקבלת דגימת יציאה הוא  $m$ .

לאחר שחושבה הסדרה  $y^q(nCT)$  לפי (4.23) ודגימותיה אגורות במחשב, ניתן לקבל את אות הדחוס ע"י העברת הדגימות האלה דרך ה-  $D/A$  בקצב הדגימה המקורי  $T$ . מכיוון שבסדרת היציאה  $y^q(nCT)$  יש פי  $C$  פחות דגימות מאשר בסדרת הכניסה  $s(nT)$ , האות שיתקבל ביציאת ה-  $D/A$  יהיה קצר יותר בזמן פי  $C$  מאשר האות  $s(t)$ . אולם, אות היציאה יהיה בעל תחום תדרים זהה לתחום התדרים של אות הכניסה  $s(t)$ . לפי (4.23) נראה ש-  $TSC$  מושג ע"י שקלול מתאים של מספר קטעי דבור בכדי לקבל קטע בודד. למעשה, הפעולה נעשית בקטעים - בכל צעד נלקח לעיבוד קטע דיבור באורך  $N_{sc} + N_p$ , והתוצאה קטע באורך  $N_{sc}$ , כלומר קטע הקצר פי  $C$  בערך מהקטע שנלקח לעיבוד, (בדיוק, רק אם  $N_p/(C-1)$  הוא שלם).

#### 4.2 שיקולים בבחירת פרמטרים

נדון בסעיף זה בבחירת ערכים מתאימים עבור הפרמטרים של האלגוריתם שתואר בסעיף הקודם. פרמטרים אלה הם:  $m$ ,  $N_p$  ו-  $C$ , פרט להם יש לקבוע גם את  $h(t)$ .

כאמור בתחילת הסעיף הקודם, מניחים שאות הדיבור הקולי הוא קואזי-מחזורי, ואז מעוניינים לחלק את תחום התדרים של אות הדיבור למספר פסים רב ככל האפשר ולא פחות מאשר מספר הדגימות במחזור הארוך ביותר הצפוי. כלומר, דרוש שמספר הפסים לא יהיה קטן ממספר הדגימות במחזור ה-  $pitch$  שהוא המחזור היסודי בדיבור. בדרך זו מובטח שבכל פס לא יהיה יותר מקו ספקטרלי אחד ואז, הקרוב שמתבצע כאן לחלוקה בתדר, יהיה טוב יותר.

\* אם  $N_p/(C-1)$  אינו שלם נקבל רק בקירוב.

בכדי להפחית את מידת החפיפה בין מסננים צמודים בבנק המסננים, רצוי  $m$  גדול ככל האפשר, מכיוון שככל ש-  $m$  יהיה גדול יותר נקבל עבור המסנן האב טיפוס קרוב טוב יותר למסנן מעביר נמוכים אידיאלי הנדרש. מצד שני מטעמי חסכון בחישובים רצוי  $m$  קטן ככל האפשר, כי מספר החישובים יחסי ל-  $m$ . אם  $h(t)$  סימטרי ומקיים אילוצים מסוימים שיובאו בהמשך, אזי  $m$  המינימלי הוא 2.

נקודה אחרת שיש לקחת בחשבון היא, שאות הדיבור נחשב סטציונרי רק בקטע זמן קצר, בדרך כלל באורך של 20-40 msec. מתוך (4.23) נובע, שהתחום מהעבר של האות, ממנו נלקחות דגימות כניסה לצורך השקלול הוא  $(m-1)N_p$ . נסמן ב-  $M_s$  את מספר הדגימות בקטע דיבור שניתן להחשב כסטציונרי, אזי  $m$  ו-  $N_p$  צריכים למלא את הדרישה

$$M_s \geq (m - 1) N_p \quad (4.24)$$

מילוי דרישה זו הכרחי כדי להבטיח שבחישוב אות היציאה  $y^q$ , ישוקללו רק דגימות כניסה השייכות לאותו קטע סטציונרי.

שיקול אחרון לגבי  $m$  נובע מתחום הזמן. מתוך (4.23) מתקבל שאם  $m < C$ , לא כל דגימות הכניסה משוקללות לצורך יצירת סדרת היציאה החדוטה  $y^q$ . אולם, הודות ליתרות (redundancy) שבאותות דיבור ומכיוון ש-  $C$  בדרך כלל אינו עולה על 3 (כך הדבר במערכת שפומחה), לא נראה שהכרחי להתחשב בהגבלה זו. בחירת  $N_p$  ו-  $m$  לפי השיקולים הנ"ל קובעת את  $mN_p T$  כמשכה של  $h(t)$  - התגובה להלם של המסנן האב טיפוס, אך יש דרישות נוספות שעל  $h(t)$  למלא וכעת נדון בהן.

כאשר מתכננים בנק של מסננים, דורשים שהתגובה של הבנק כולו תהיה אחידה באמפליטודה ולינארית בפאזה בתחום התדרים של אות הכניסה. קיום דרישה זו מבטיח שסכום היציאות של כל אחד מהמסננים יתן אות השונה מאות הכניסה רק בהשהיה קבועה. בכדי לקיים את הדרישה הזאת מספיק ש-  $h(t)$  תהיה סימטרית (לצורך קבלת פאזה לינארית) וכן שתהיה בעלת התכונה [7]:

$$\begin{aligned} N_p^i h(t) \Big|_{t=iN_p T} &= 1 \text{ for } i = 1 \\ &= 0 \text{ for } i \neq 1 \end{aligned} \quad (4.25)$$

כאשר 1 הוא שלם סופי כלשהו.

דרישה חשובה אחרת עבור  $h(t)$  מתקבלת ישירות מתוך (4.23). אם אות הכניסה הוא מחזורי, ובעל מחזור הנמשך  $N_p T$ , אזי דרוש שאות היציאה הדחוסה המתקבל מחישוב (4.23) יהיה מחזורי גם כן ובעל אותו מחזור. מ- (4.23) מתקבל שכך יושג אם

$$\sum_{i=0}^{m-1} \hat{h}[(n+iN_{sc}) T(C-1)] = N_p' \sum_{i=0}^{m-1} h(t) \Big|_{t=nT(C-1)+iN_p T} = 1 \quad (4.26)$$

$$n = 0, 1, \dots, N_{sc} - 1 \quad \text{כאשר}$$

ראוי לציין שקיום (4.25) אינו סותר את קיום (4.26).

עבור  $m = 2$ , פונקציה פשוטה המקיימת את כל הדרישות הנ"ל היא הפונקציה של משולש מודד:

$$\begin{aligned} \hat{h}(t) &= t/(N_p T), & 0 \leq t \leq N_p T \\ &= 1 - (t - N_p T)/(N_p T), & N_p T < t \leq 2N_p T. \end{aligned} \quad (4.27)$$

עבור ערכים גבוהים יותר של  $m$ , ניתן להשתמש בפונקציה  $\sin(\pi t/N_p T)/(\pi t/N_p T)$  משוקללת ומוזזת כמתואר ב- [7].

הפרמטר האחרון שיש לדון בו הוא גורם הדחיסה  $C$ . במערכת שפותחה הגיות בעלות משך זמן שונה נדחסות לאותו אורך, לכן  $C$  שונה בדרך כלל מהגיייה אחת להגיייה אחרת, ומחושב מחדש עבור כל הגיייה. זאת בניגוד ל-  $N_p$ ,  $m$  ו-  $h(t)$  הנקבעים עבור המערכת פעם אחת ונשארים ללא שינוי. ניתן להיווכח, שבקרוב לחלוקה בתדר באמצעות ההזזה של הקווים הספקטרלים של האות, כפי שתואר לעיל, השגיאה במיקום הקווים הספקטרלים של אות היציאה (לאחר הדחיסה), נמצאת ביחס ישר לגורם הדחיסה  $C$ , וביחס הפוך ל-  $N_p$ . באופן מעשי יש להגביל את  $C$  ל- 3. כך זה נמצא גם כגבול ההבנה עבור דיבור דחוס.

נתאר כעת את התהליך לחישוב ערכו של  $C$ . יהיה  $L_j$  מספר הדגימות בהגיית הכניסה ה- $j$  שיש לדחוס, ויהיה  $L_{c_0}$  מספר הדגימות הדרוש עבור ההגייה הדחוסה.  $L_{c_0}$  נקבע כך שגורם הדחיסה המכסימלי הצפוי במערכת יוגבל ל-3, וכן ישכולל מספר שלם של קטעי זמן, אשר מכל אחד מהם ממצים, בשלב הבא של תהליך הזיהוי, את הפרמטרים המיצגים את ההגייה. בכדי למנוע יתופעת מעברי המתקבלת אם מתחילים לחשב את (4.23) עם תנאי התחלה אפס (כלומר,  $s(n) = 0$  עבור  $n < 0$ ), מתחילים את חישוב (4.23) בנקודה  $n = N_p(m-1)$  כאשר דגימות הכניסה  $s(0), s(1), \dots, s[N_p(m-1)-1]$  נלקחות כתנאי התחלה. מכאן אפשר לקבוע את  $C_j$  - גורם הדחיסה עבור ההגייה ה- $j$  לפי הביטוי  $[L_j - (m-1)N_p] / L_{c_0}$ . אולם, מכיוון שבדרך כלל  $N_p / (C-1)$  אינו שלם, ויש לחשב את  $N_{sc}$  לפי (4.19), אזי גורם הדחיסה המעשי במקרים מסוימים, עלול להיות גבוה יותר. בכדי למנוע מצב שבו עקב גורם דחיסה אשר ערכו עלה בדרך הנייל, התקבלו פחות דגימות מאשר  $L_{c_0}$ , יש לקבוע את  $C_j$  לפי הביטוי:

$$C_j = [L_j - (m-1)N_p] / L_c \quad (4.28)$$

כאשר

$$L_c = L_{c_0} + \Delta L_c \quad (4.29)$$

ו-  $\Delta L_c$  ניתן לחישוב לפי הביטוי

$$\Delta L_c = 2L_{c_0} / (3N_p) \quad (4.30)$$

ביטוי (4.30) נובע מתוצאה שהתקבלה ב- [7], שהגיאה היחסית בחישוב  $C_j$  מוגבלת ע"י  $0.5(C_j-1)^2 / (C_j N_p)$ . מכיוון ש-  $C_j$  מוגבל ל-3 מתקבל שהגיאה עבור  $C_j$  במקרה זה היא

$$\Delta C_j = 2C_j / (3N_p) \quad (4.31)$$

כלומר, באופן מעשי עלול להתקבל גורם דחיסה  $C_j^1 = C_j [1+2/(3N_p)]$  ולכן אם נקטין את הגורם שחושב בתחילה:  $[L_j - (m-1)N_p] / L_{c_0}$  פי  $[1+2/(3N_p)]$ , נבטיח שגורם הדחיסה המעשי יהיה כנדרש. כך עושים ב- (4.28) - (4.30). חישוב  $C_j$  עפ"י ביטויים אלה מבטיח ללא רק שמספר הדגימות של ההגייה הדחוסה לא יקטן מ-  $L_{c_0}$ , אלא גם שלא יעלה על  $L_{c_0}$  במידה מופרזת. יש חשיבות לתוצאה האחרונה, מכיוון שהמערכת בנויה לקליטת  $L_{c_0}$  דגימות בשלב הבא של תהליך הזיהוי, והעודף אינו נלקח בחשבון (למעשה אף אין מחשבים אותו מטעמי חסכון בזמן חישוב ובמקום בזכרון). כעת, אם  $C_j$  מחושב כך שעלול להיווצר עודף גדול מדי, שטביר שיכיל אינפורמציה משמעותית של ההגייה, האינפורמציה הזאת תאבד עם העודף, וסיכוי הטעות בזיהוי יגדל.

מתוך (4.23) ניתן לראות שיהיה זה יעיל ונוח לעבד את אות הכניסה בקטעים. גישה כזאת תמנע את הצורך לחשב את  $\alpha$  כמוגדר ב- (4.20), וההתקדמות לפי  $n$  ב- (4.23) נעשית באופן פשוט. כל זה מודגם בתרשים הזרימה למימוש האלגוריתם המופיע בציר מס' 4.1.

לשם הבהרת אופן פעולתו של האלגוריתם לדחיסה בתחום הזמן, נמחיש בציר 4.2 את אשר מתבצע כאשר מפעילים את האלגוריתם עם  $m = 2$ , ו-  $h(t)$  פונקציה חלון משולש לפי (4.27). נעשה זאת עבור שני מקרים. במקרה הראשון (ציר 4.2(a)) גורם הדחיסה  $C = m = 2$ , ובמקרה השני (ציר 4.2(b))  $C = 3$ , כלומר  $C > m$ .

בכדי לקבל את ציר 4.2 נעזר בתרשים הזרימה למימוש האלגוריתם שבציר 4.1. תחילה נחשב את דגימות החלון לפי (4.27), כאשר  $h(t)$  נדגמת במחזור דגימה  $T(C-1)$ , וקיים  $N_p T(C-1) = N_{sc} T(C-1)$  (על סמך (4.19)). לפיכך, נקבל עבור  $\hat{h}(n)$  שבתרשים הזרימה:

$$\begin{aligned} \hat{h}(n) &= n/N_{sc}, & n &= 1, \dots, N_{sc} \\ &= 1 - (n - N_{sc})/N_{sc}, & n &= N_{sc} + 1, \dots, 2N_{sc} \end{aligned} \quad (4.32)$$

הביטוי מתרשים הזרימה, לחישוב דגימת יציאה של ההגייה הדחוסה עבור  $m = 2$  הוא:

$$Y = S(K) \hat{h}(K+N_{sc}) + S(K+N_p) \hat{h}(K) \quad (4.33)$$

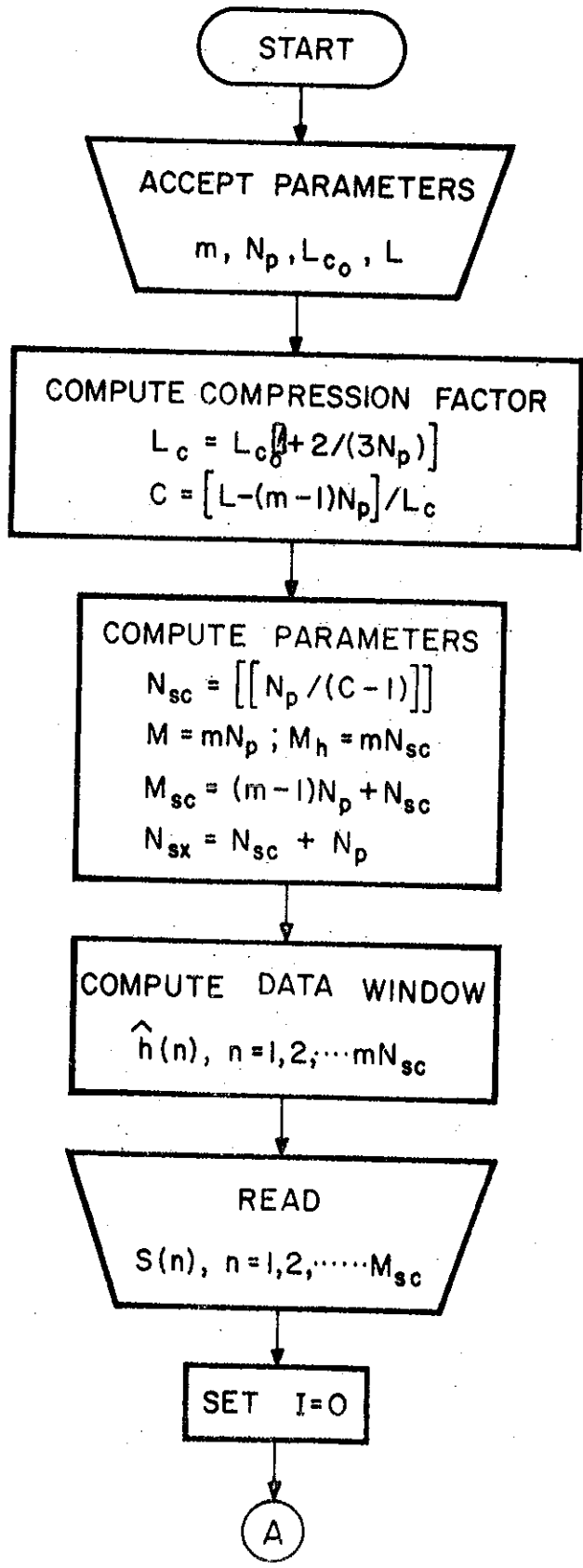


$S(K)$  היא דגימת כניסה מתוך מערך של  $N_p + N_{sc}$  דגימות כניסה. והחישוב של  $Y$  מתבצע עבור  $N_{sc}$  נקודות יציאה, כלומר, עבור (4.32):  $K = 1, \dots, N_{sc}$ . כעת נדון בנפרד בכל אחת מהמקרים הנ"ל.

$C = m = 2$ . במקרה זה  $N_{sc} = N_p / (C-1) = N_p$  (לפי (4.19)). ולכן לצורך חישוב  $N_{sc} = N_p$  דגימות יציאה, קוראים בכל צעד  $2N_p$  דגימות כניסה. החלון לפי (4.32) כולל במקרה זה  $2N_p$  דגימות.

הציור 4.2(a) מתאר כיצד יש לחשב במקרה זה את  $Y$ , עבור  $K = 1, \dots, N_p$ , לפי (4.33). הקו האופקי מסמן את דגימות ההגייה. המשולשים שמעליו הם חלקי החלון  $\hat{h}(n)$  שדגימותיו כופלות את דגימות הכניסה. דגימת יציאה  $Y(K)$  מתקבלת מהצרוף של דגימת כניסה  $S(K)$  מוכפלת בערך דגימת החלון הנמצאת מעליה, ושל דגימת הכניסה  $S(K+N_p)$  המוכפלת גם היא בדגימת החלון הנמצאת מעליה. לפיכך,  $Y(K)$  היא צרוף של 2 דגימות כניסה המשוקללות בדגימות חלון מתאים. בציור רואים בברור שעבור המקרה בו  $C = m$  כל דגימות הכניסה תורמות ליצירת דגימות היציאה. כן רואים כיצד  $2N_p$  דגימות של הגיית הכניסה נותנות  $N_p$  דגימות של הגיית יציאה על קו אופקי תחתון. הרציפות של דגימות הכניסה נשמרת מאחר והדגימה המתאימה ל-  $K = 0$  היא למעשה הדגימה ה-  $K = N_p + N_{sc}$  של הצעד הקודם.

$C = 3$ , ומכיוון ש-  $m = 2$  אזי  $C > m$ . במקרה זה  $N_{sc} = N_p / (C-1) = N_p / 2$ , ולכן לצורך חישוב  $N_{sc} = N_p / 2$  דגימות יציאה קוראים בכל צעד  $(3/2)N_p$  דגימות כניסה. החלון לפי (4.32) כולל במקרה זה  $N_p$  דגימות. מקרה זה מתואר בציור 4.2(b). הציור מראה בברור שחלק מדגימות הכניסה כלל אינו תורם ליצירת דגימות היציאה. כן רואים שדרושים 2 צעדים, כלומר יש לקרוא 3 קטעים, כל אחד מכיל  $N_p$  דגימות כניסה, כדי לקבל קטע של  $N_p$  דגימות יציאה, ואז  $N_p/2$  הראשונות של הקטע הראשון משתקללות עם  $N_p/2$  האחרונות של הקטע השני לקבלת  $N_p/2$  דגימות היציאה הראשונות. ו-  $N_p/2$  האחרונות של הקטע השני משתקללות עם  $N_p/2$  האחרונות של הקטע השלישי לקבלת  $N_p/2$  דגימות היציאה האחרונות. עבור אות דיבור מחזורי (במחזור  $N_p$ ), בלי תלות בערכו של  $C$  ביחס ל-  $m$ , נקבל ביציאה הגייה דחוסה מחזורית, בעלת המחזור של אות הכניסה. תוצאה זו מתקבלת עקב החלון המשולש שבחרנו עבור  $h(t)$  (כאשר  $m = 2$ ) המקיים את התכונה הרשומה ב- (4.26).



ציור 4.1: תרשים זרימה למימוש האלגוריתם לדחיסה בזמן במערכת לזיהוי מילים בודדות.

Fig. 4.1: Flow chart diagram for implementing the time scale compression (TSC) algorithm in isolated word recognition systems.

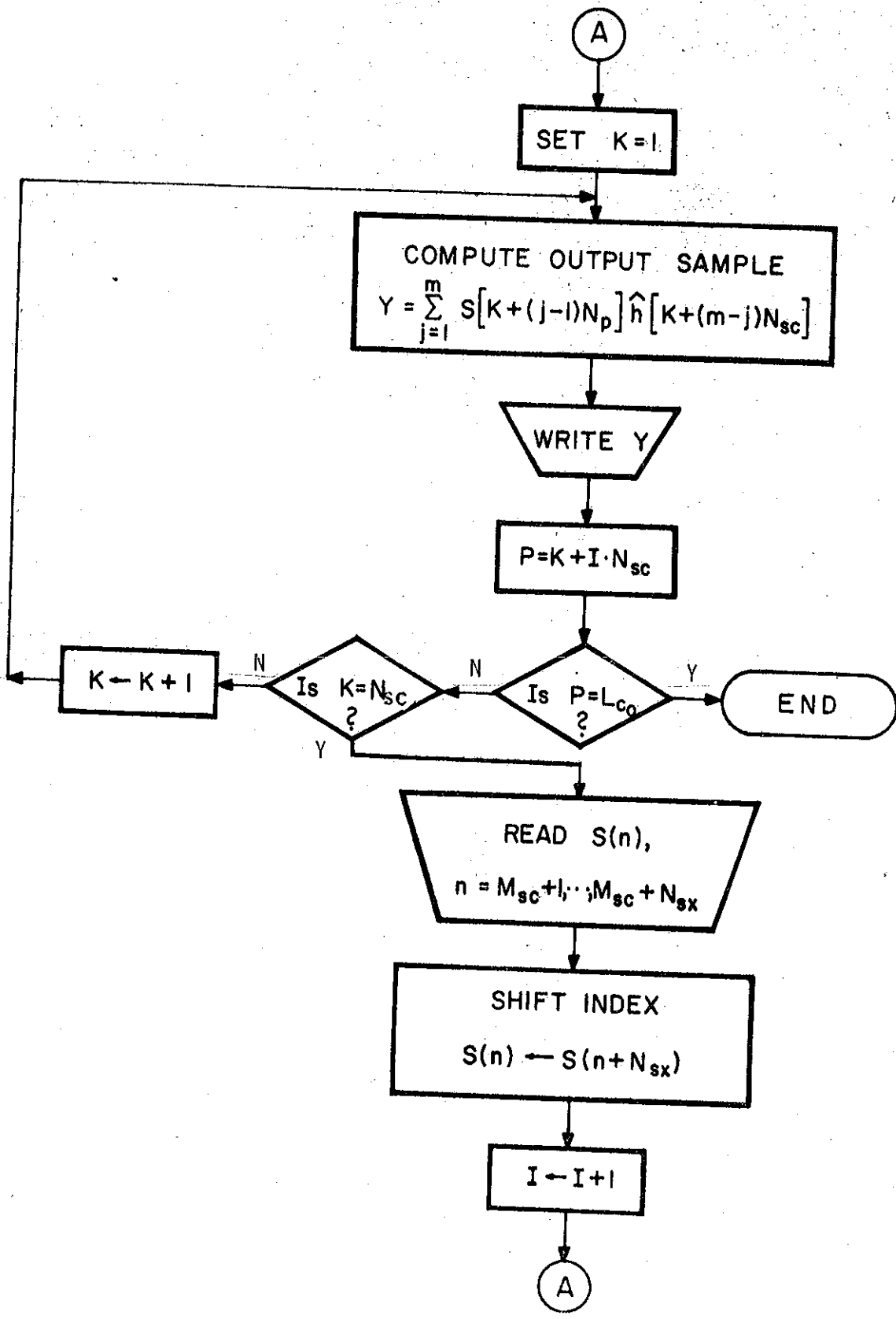
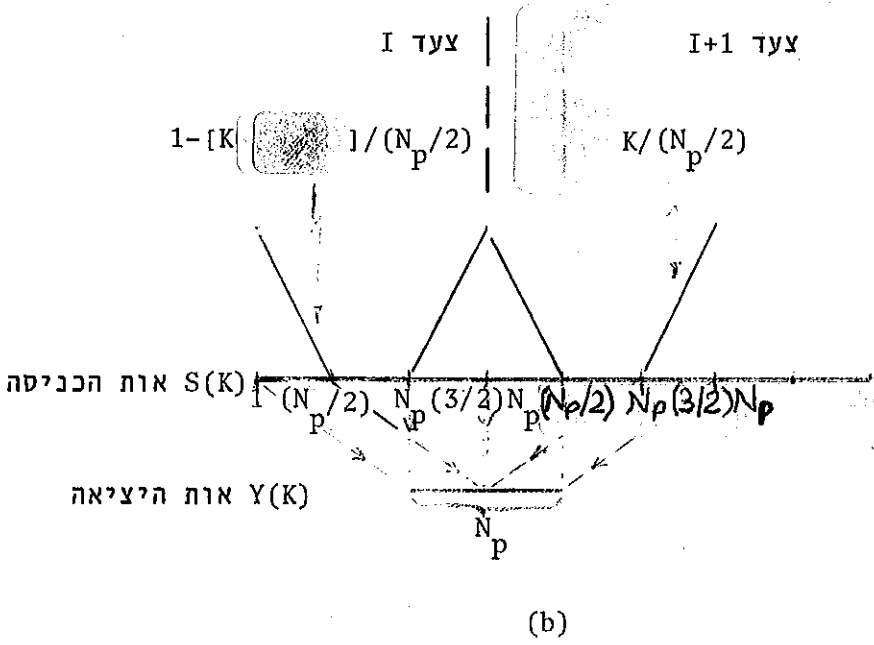
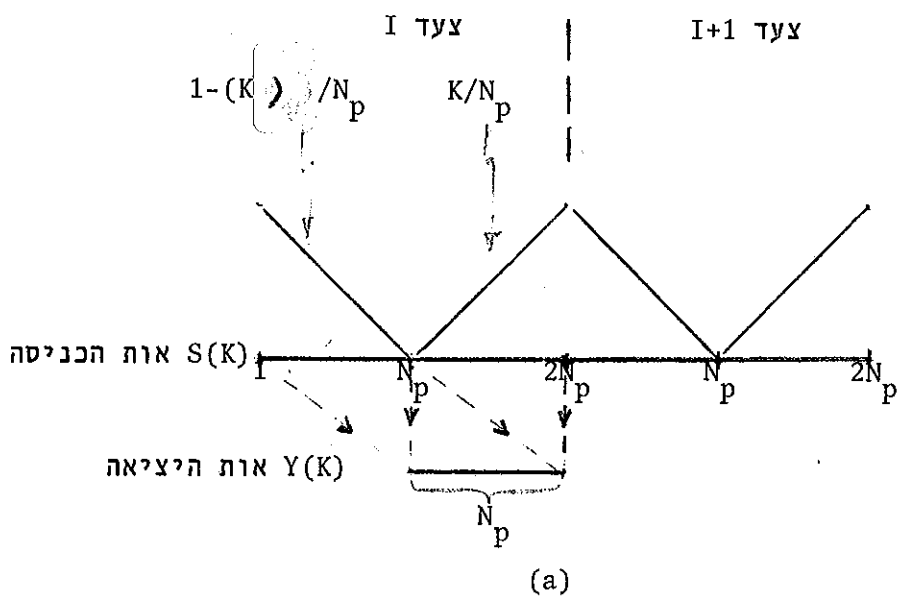


Fig. 4.1 (cont.)



ציר 4.2: הפעולה בתחום הזמן של האלגוריתם לדחיסה עבור  $m=2$  וחלון משולש.  
 .C = 2 (a)  
 .C = 3 (b)

Fig.4.2: Time domain processing by TSC algorithm for  $m=2$ , and triangular data window.  
 (a)  $C = 2$ .  
 (b)  $C = 3$ .

פרק 5: חיזוי לינארי של אותות דיבור.

פרק זה מטרתו להגיש את הבסיס התיאורטי עבור הפרמטרים - מקדמי הקורלציה החלקית (Partial correlation - PARCOR coefficients), שנבחרו לייצג את ההגיייה במערכת שפותחה. פרמטרים אלה מחושבים מתוך ההגיייה הדחוסה בשלב מיצוי התכונות. השיקולים שהביאו לבחירה במקדמי ה-PARCOR מתוארים בסעיף 3.2, ותאור תהליך חישובם מופיע בסעיף 6.3.

מודל המסנן כולו קטבים.

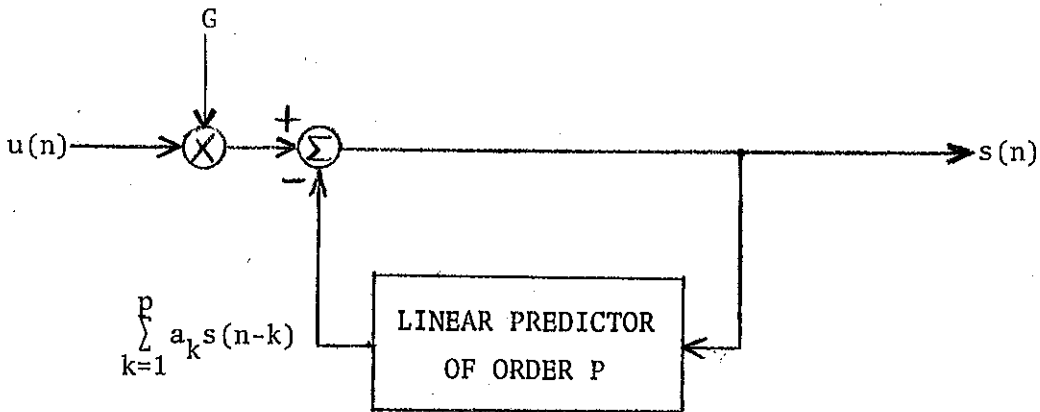
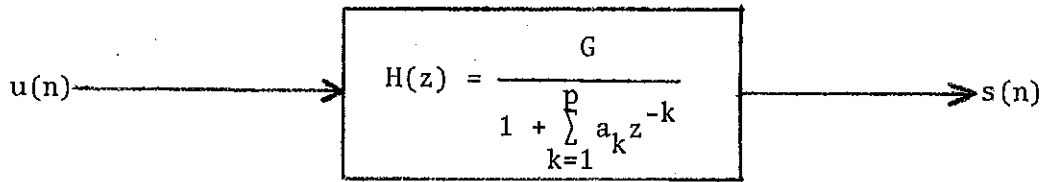
5.1

מקדמי הקורלציה החלקית מבוטסים על המודל לפיו דגימות הדיבור הן תפוקה של מסנן כולו קטבים (all-pole filter). התאור של מודל זה והקשר בינו לבין מקדמי הקורלציה החלקית הובאו כבר בסעיף 2.2.2. נחזור לתאור המודל. ציור 5.1 מתאר את מודל המסנן בתחום הזמן ואת פונקציית התמסורת שלו [15]. מתוך ציור 5.1 ניתן לראות שמודל המסנן מתואר בתחום הזמן ע"י המשוואה:

$$s(n) = - \sum_{k=1}^p a_k s(n-k) + Gu(n) \tag{5.1}$$

מודל זה יכול לשמש לשחזור הדיבור בצורה הבאה. כדי לשחזר דיבור קולי (voiced)  $u(n)$  יהיה רכבת פולסים שיופיעו בקצב התדירות היסודית (pitch frequency) של הדיבור הקולי אותו משחזרים, ועבור דיבור לא קולי (unvoiced) יהיה  $u(n)$  רעש לבן. אולם למעשה, הכניסה  $u(n)$  אינה ידועה במדויק, ולכן ניתן לחזות רק בקרוב את דגימת הדיבור  $s(n)$  בעזרת סכום של  $p$  דגימות העבר המשוקללות באופן לינארי. לפיכך  $\hat{s}_n$  החיזוי של  $s_n$  יהיה [15]:

$$\hat{s}_n(n) = - \sum_{k=1}^p a_k s(n-k) \tag{5.2}$$



(a) פונקציית התמסורת של המסנן כולו קטבים. ציור 5.1:  
(b) מודל המסנן כולו קטבים בתחום הזמן.

Fig. 5.1: (a) All - pole model - transfer function.  
(b) All - pole model in the time domain.

השגיאה בין  $s(n)$  לבין  $\hat{s}(n)$  נתונה ע"י:

$$e(n) = s(n) - \hat{s}(n) = s(n) - \sum_{k=1}^p a_k s(n-k) \quad (5.3)$$

השגיאה הריבועית הכוללת (total squared error) הנקראת גם שארית החיזוי הלינארי (Linear prediction residual-LPR) נתונה ע"י:

$$E = \sum_n e^2(n) = \sum_n \left( s(n) - \sum_{k=1}^p a_k s(n-k) \right)^2 \quad (5.4)$$

מקדמי החיזוי הלינארי (linear predictor coefficients -LPC) הם  $\{a_k\}$   $k=1, \dots, p$  המגשימים את הערך המינימלי עבור השגיאה הריבועית הכוללת הנתונה ב- (5.4). טווח הסיכום עבור  $n$  והגדרת  $s(n)$  בטווח זה הם בעלי חשיבות. נתייחס אל נושאים אלה בהמשך, כעת נראה את הביטויים הכלליים ביחס לטווח של  $n$  המתקבלים בתהליך המינימיזציה של (5.4). בכדי לקבל את ה-LPC פותרים את המשוואה הבאה:

$$\frac{\partial E}{\partial a_i} = 0, \quad 1 \leq i \leq p. \quad (5.5)$$

מ- (5.4) ו- (5.5) מתקבלת מערכת המשוואות [15]:

$$\sum_{k=1}^p a_k \sum_n s(n-k) s(n-i) = - \sum_n s(n) s(n-i), \quad 1 \leq i \leq p. \quad (5.6)$$

מערכת משוואות זו היא בעלת  $p$  משוואות עבור  $p$  נעלמים - ה-LPC:  $\{a_k\}, k=1, \dots, p$ . כאשר מפתחים את הביטוי (5.4) ומציבים בו את (5.6) מקבלים את  $E_p$  - השגיאה הריבועית הכוללת המינימלית [15]:

$$E_p = \sum_n s^2(n) + \sum_{k=1}^p a_k \sum_n s(n) s(n-k) \quad (5.7)$$

נגש כעת לקביעת טווח הסיכום עבור  $n$ . יש שתי שיטות בקשר לכך, וכל אחת מובילה לקבוצת פתרונות אחרת עבור (5.6).

## 5.2 שיטת האוטוקורלציה (השיטה הסטציונרית).

לפי שיטה זו מניחים שהשגיאה הריבועית הכוללת שב- (5.4) עוברת מינימיזציה על פני  $-\infty < n < +\infty$ , ואז המשוואות (5.6) ו- (5.7) מצמצמות ל-:

$$\sum_{k=1}^p a_k R(i-k) = -R(i), \quad 1 \leq i \leq p \quad (5.8)$$

$$E_p = R(0) + \sum_{k=1}^p a_k R(k) \quad (5.9)$$

כאשר:

$$R(i) = \sum_{n=-\infty}^{\infty} s(n) s(n+i), \quad 0 \leq i \leq p \quad (5.10)$$

היא פונקציית האוטוקורלציה של האות  $s(n)$ . מקדמי  $R(i-k)$  יוצרים את מטריצת האוטוקורלציה ולכן נקראת השיטה - שיטת האוטוקורלציה. המקור לשם השני של השיטה הוא שעבור תהליכים סטציונריים פונקציית האוטוקורלציה תלויה אך ורק בהפרש שבין כל שתי נקודות זמן בתהליך, ולא בשעוריהן, ותכונה זו מתקיימת ב- (5.10).

באופן מעשי אות הדיבור  $s(n)$  קיים במשך זמן סופי, ולכן על מנת לשמור לגביו על תנאי הסטציונריות, מחשבים את (5.10) עבור קטעי דיבור קצרים (אורך כל קטע כ- 20÷40 msec), כאשר כל אחד חופף בחלקו לבא אחריו. לשם כך מקובל לכפול את  $s(n)$  בפונקציית חלון  $w(n)$  באורך  $N$  דגימות בכדי לקבל  $s'(n)$  המקיים:



$$s'(n) = \begin{cases} s(n) w(n), & 0 \leq n \leq N-1 \\ 0, & \text{otherwise.} \end{cases} \quad (5.11)$$

ואז פונקציית האוטוקורלציה נתונה ע"י

$$R(i) = \sum_{n=0}^{N-1-i} s'(n) s'(n+i), \quad i \geq 0 \quad (5.12)$$

לצורת פונקציית החלון  $w(n)$  יש חשיבות, ועל כך מדובר בסעיף 6.3.

### שיטת הקוואריאנס (השיטה הלא סטציונרית)

5.3

בניגוד לשיטת האוטוקורלציה, כאן מבצעים מינימיזציה של השגיאה הריבועית הכוללת שב- (5.4) על פני קטע זמן סופי  $0 \leq n \leq N-1$  ואין מניחים כל הנחה על הנתונים שמחוץ לקטע זה. כעת המשוואות (5.6) ו- (5.7) מצטמצמות ל-

$$\sum_{k=1}^p a_k \phi_{ki} = -\phi_{oi}, \quad 1 \leq i \leq p \quad (5.13)$$

$$E_p = \phi_{oo} + \sum_{k=1}^p a_k \phi_{ok} \quad (5.14)$$

כאשר

$$\phi_{ik} = \sum_{n=p}^{N-1} s(n-i) s(n-k), \quad 0 \leq i, k \leq p \quad (5.15)$$

הוא הקוואריאנס של האות  $s(n)$  בקטע הנ"ל. המקדמים  $\phi_{ki}$  שב- (5.13) יוצרים את מטריצת הקוואריאנס ולכן נקראת השיטה בשם שיטת הקוואריאנס. כפי שרואים ב- (5.15) תלויים מקדמי  $\phi_{ik}$  בזמן, כך מתקיים בתהליכים לא סטציונריים ולכן מקובל לכנות שיטה זו בשם השיטה הלא סטציונרית. שיטת הקוואריאנס מתלכדת עם שיטת האוטוקורלציה כאשר הטווח של  $n$  הולך לאינסוף.

5.4 השוואה בין שיטת האוטוקורלציה לבין שיטת הקוואריאנס.

לאחר סקירת שתי השיטות המקובלות במסגרת הטיפול בחיזוי לינארי של אותות דיבור, נשווה ביניהן ממספר בחינות.

1. מידת הדיוק בקרוב של אות הדיבור באמצעות חיזוי לינארי. כמידה ראשונה להערכת הדיוק תשמש השגיאה הריבועית הכוללת המינימלית לאחר שעברה נירמול ביחס לאנרגיה של קטע הדיבור אותו מקרבים. עבור שיטת האוטוקורלציה תתקבל שגיאה זו  $\eta_0$ , מתוך (5.9):

$$\eta_0 = 1 + \sum_{k=1}^p a_k (R(k)/R(0)) \quad (5.16)$$

עבור שיטת הקוואריאנס תתקבל השגיאה הנ"ל  $\eta_c$  מתוך (5.14):

$$\eta_c = 1 + \sum_{k=1}^p a_k (\phi_{ok}/\phi_{oo}) \quad (5.17)$$

Chandra and Lin [17] ערכו השוואה בין  $\eta_0$  לבין  $\eta_c$  המתקבלות עבור דיבור קולי (voiced). קטעי הדיבור עבורם חושבו שגיאות אלה נבחרו בשתי צורות. לפי הראשונה הנקראת pitch synchronous, מחשבים את ה-LPC עבור קטעי דיבור שכל אחד מהם מתחיל בדגימה המתאימה להלם ה-pitch ונמשך פחות או יותר כמשך מחזור ה-pitch. לפי הצורה השנייה הנקראת pitch asynchronous אין קשר בין התחלת כל קטע לבין הלם ה-pitch, ומשכו אינו מוגבל למחזור ה-pitch. נסמן את הצורות הראשונה והשנייה ב-PS, PA בהתאמה. מההשוואה התקבל, שכאשר ארכו של קטע הדיבור עבורו מחושבות  $\eta_0$  ו- $\eta_c$  קטן ממחזור ה-pitch (PS) אזי  $\eta_0 \ll \eta_c$ . לעומת זאת, כאשר מגדילים את ארכו של קטע הדיבור מעבר למחזור ה-pitch ובוחרים אותו לפי PA,  $\eta_0$  הולכת וקטנה, ועבור קטע דיבור הנמשך 2 עד 3 מחזורי pitch,  $\eta_0$  ו- $\eta_c$  בקרוב שוות.

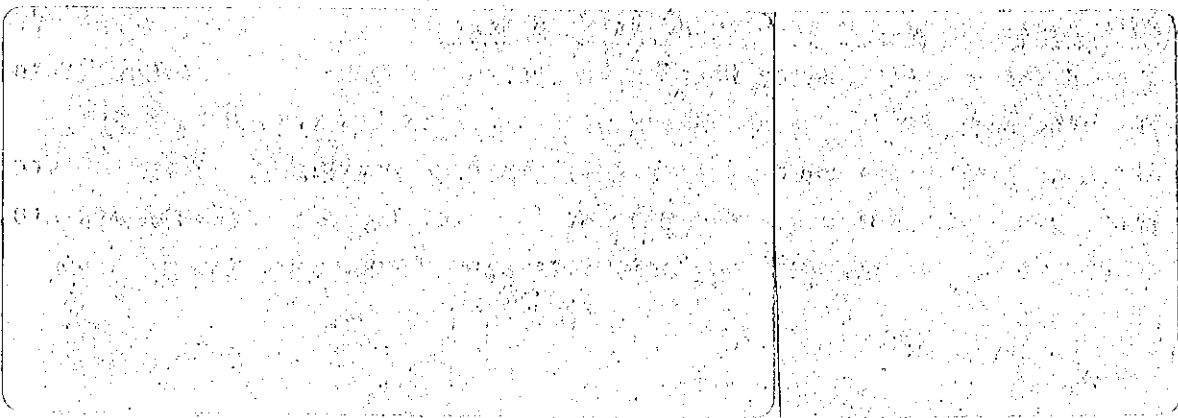
כמידה שנייה להערכת הדיוק בקרוב באמצעות חיזוי לינארי תשמש מידת הדיוק בשערוך הספקטרום. כאמור בסעיף 2.2.2 ניתן לקרב את המעטפת הספקטרלית של אות דיבור בקטע מסוים בעזרת ה-LPC שחושבו מתוך אותו קטע. ב-[17] משווים את המעטפות המתקבלות על סמך שיטת האוטוקורלציה ושיטת הקוואריאנס

עם המעטפת הספקטרלית המתקבלת מתוך התמרת פוריה הדיסקרטית הממומשת באמצעות fast Fourier transform-FFT, של קטע הדיבור הנדון עבורו חושבו ה-LPC לפי שתי השיטות הנ"ל.

מהשוואה התקבל שכאשר קטעי הדיבור נבחרו לפי PS, מקרבת המעטפת הספקטרלית שחושבה לפי שיטת האוטוקורלציה את המעטפת שהתקבלה ב-FFT. אולם שתי המעטפות הללו אינן מראות בברור את הפורמנטים [17]. לעומת זאת, המעטפת שהתקבלה לפי שיטת הקווריאנס אינה מקרבת את המעטפת שהתקבלה ב-FFT, אך היא מצליחה למסור בברור את הפורמנטים, כלומר היא מקרבת טוב יותר משאר השיטות את פונקציית התמסורת של המעבר הקולי (vocal tract). כאשר קטעי הדיבור נבחרו לפי PA, ומשך כל אחד מהם עלה על 3 מחזורי pitch, שתי השיטות קרבו בצורה מדויקת את המעטפת שהתקבלה ב-FFT.

לסיכום, מתוך תוצאות הניסויים שב- [17], מתקבל שעבור קטעי דיבור הנבחרים ב-PS, שיטת הקווריאנס מדייקת יותר בקרוב את הדיבור הקולי והפורמנטים שבו מאשר שיטת האוטוקורלציה לעומת זאת, כאשר קטעי הדיבור נבחרים ב-PA, ומשך כל אחד מהם כ-2, 3 מחזורי pitch, הביצועים של שתי השיטות פחות או יותר שווים.

2. יציבותו של המסנן כולו קטבים. כאשר מחשבים את ה-LPC לפי שיטת האוטוקורלציה, מובטחת באופן תיאורטי יציבותו של המסנן כולו קטבים הממומש באמצעותם [18]. לעומת זאת, בשיטת הקווריאנס כלל לא מובטחת באופן תיאורטי יציבותו של המסנן למעשה כאשר החישוב נעשה באורך מילה סופי ולא בנקודה צפה, גם בשיטת האוטוקורלציה היציבות אינה מובטחת.



ההשוואה שהובאה כאן בין שיטת האוטוקורלציה לבין שיטת הקוואריאנס, מהווה בסיס עליו, מסתמכת הבחירה בשיטת האוטוקורלציה עבור המערכת שפותחה. השיקולים עבור בחירה זו מתוארים בסעיף 6.3.

5.5 הקשר בין ה-LPC לבין מקדמי ה-PARCOR.

נפתח את הקשר בין ה-LPC (המחושבים בשיטת האוטוקורלציה) לבין מקדמי הקורלציה החלקית - מקדמי ה-PARCOR, שהם הפרמטרים המייצגים את ההגיייה במערכת שפותחה.

לפי שיטת האוטוקורלציה יש לפתור את מערכת המשוואות (5.8) בכדי לקבל את ה-LPC:  $\{a_k\}$ ,  $1 \leq k \leq p$ . לשם כך קיימות שיטות פתרון כלליות, אולם פותחו גם שיטות המתאימות לתכונות מטריצת האוטוקורלציה הנוצרת ע"י מקדמי  $R(i-k)$ . שיטת הפתרון המקובלת כיום כמהירה ביותר [15], מנצלת את העובדות שמטריצת האוטוקורלציה היא סימטרית, שהאלמנטים לאורך כל אחד מהאלכסונים זהים (מטריצה בעלת תכונות אלה מכונה בשם Toeplitz matrix), ושארבי וקטור העמודה שב- (5.8) מופיעים במטריצת האוטוקורלציה. שיטה זו מיוחסת ב- [15] ל-Durbin (מקור מס' 25 ב- [15]), והיא כוללת את הקשרים הרקורסיביים הבאים:

$$E_0 = R(0) \quad (5.18)$$

$$K_{i+1} = - \frac{\sum_{k=0}^i a_k^{(i)} R(i+1-k)}{E_i}; \quad a_0^{(i)} = 1 \quad (5.19)$$

$$a_k^{(i+1)} = \begin{cases} K_{i+1}, & k = i+1 \\ a_k^{(i)} + K_{i+1} a_{i+1-k}^{(i)}, & k = 1, 2, \dots, i \\ a_0^{(i)}, & k = 0 \end{cases} \quad (5.20)$$

$$E_{i+1} = (1 - K_{i+1}^2)E_i \quad (5.21)$$

המשוואות (5.19) - (5.21) נפתרות באופן רקורסיבי עבור  $i = 1, 2, \dots, p$  כאשר  $p$  הוא מספר המקדמים הרצוי (סדר המסנן). עבור כל ערך חדש של  $i$  נוצרת קבוצת מקדמים חדשה  $\{a_k^{(i)}\}$ ,  $1 \leq k \leq i$ , ורק כאשר מגיעים ל-  $i = p$  מתקבלת הקבוצה הדרושה. לעומת זאת  $\{K_j\}$ ,  $1 \leq j \leq i$ , שהם מקדמי הקורלציה החלקית - PARCOR, לא ישתנו כאשר יחושב המקדם הבא  $K_{i+1}$ .

הביטויים (5.19), (5.20) מתארים את מקדם ה- PARCOR ( $K_{i+1}$ ) כמוצר לוואי ל- LPC בפתרון המשוואות (5.8), אולם בסעיף 2.2.2 ניתנה לו משמעות המרומזת בשמו - מקדם קורלציה חלקית (שלילית) בין השגיאה בחיזוי קדימה  $f_p(n)$  (2.5) לבין השגיאה בחיזוי לאחור  $b_p(n)$  (2.7), ובהתאם למשמעות זו נרשם עבורו הביטוי

$$K_{p+1} = - \frac{E[f_p(n) b_p(n)]}{\sqrt{E[f_p^2(n)] E[b_p^2(n)]}} \quad (5.22)$$

כאשר  $E[y]$  מבטא את התוחלת של  $y$ .

נקשור כעת בין הביטוי (5.19) שמקורו בפתרון המשוואות (5.8) לבין הביטוי (5.22).

לפי (2.5) ו- (2.7) קיים:

$$E[f_p(n) b_p(n)] = \sum_{k=0}^p \sum_{i=0}^p a_k^{(p)} a_i^{(p)} E[s(n-k) s(n-(p+1-i))] \quad (5.23)$$

עבור אות סטציונרי, או כאשר נוקטים בשיטת האוטוקורלציה מתקיים

$$E[s(n-k) s(n-(p+1-i))] = R(p+1-i-k) \quad (5.24)$$

כאשר מציינים את (5.24) ב- (5.23) מתקבל:

$$E[f_p(n) b_p(n)] = \sum_{k=0}^p \sum_{i=0}^p a_k^{(p)} a_i^{(p)} R(p+1-i-k) \quad (5.25)$$

נחשב את  $E[f_p^2(n)]$  מתוך (2.5) ועל סמך השיקולים שהביאו ל- (5.24):

$$E[f_p^2(n)] = \sum_{k=0}^p \sum_{i=0}^p a_k^{(p)} a_i^{(p)} R(i-k) \quad (5.26)$$

חישוב  $E[b_p^2(n)]$  מתוך (2.7) ועל סמך השיקולים שהביאו ל- (5.24) נותן

$$E[b_p^2(n)] = E[f_p^2(n)] \quad (5.27)$$

לפיכך (5.26) ניתן לרישום גם כך

$$(E[f_p^2(n)] E[b_p^2(n)])^{1/2} = \sum_{k=0}^p \sum_{i=0}^p a_k^{(p)} a_i^{(p)} R(i-k) \quad (5.28)$$

נשתמש במשוואות (5.8):

$$\sum_{k=0}^p a_k^{(p)} R(i-k) = 0, \quad a_0^{(p)} = 1, \quad 1 \leq i \leq p \quad (5.29)$$

לשם הצבה ב- (5.25), (5.28) ואת הביטויים המתקבלים לאחר ההצבה הזאת נציב ב- (5.22) ונקבל

$$K_{p+1} = - \left[ \sum_{k=0}^p a_k^{(p)} R(p+1-k) \right] / \left[ \sum_{k=0}^p a_k^{(p)} R(k) \right] \quad (5.30)$$

מכיוון שהביטוי שבמכנה של (5.30) שווה ל-  $E_p$  לפי (5.9), אזי הביטוי עבור  $K_{p+1}$  ב- (5.30) זהה לביטוי (5.19) (כאשר  $i = p$ ). לפיכך הביטויים עבור  $K_{p+1}$  ב- (5.22) וב- (5.19) הם אקויוולנטיים.

לסיכום, נזכיר כאן תכונה חשובה של מקדמי ה- PARCOR הקשורה בנושא יציבות של המסנן כולו קטבים. נאמר קודם בסעיף 5.4 שעבור חישובים באורך מילה סופי, אין למעשה בטחון שהמסנן כולו קטבים יהיה יציב, לכן, הכרחי לעיתים, לבדוק את יציבותו של המסנן. הבדיקה האם השרשים של פולינום המכנה בפונקציה התמסורת שלו נמצאים במעגל היחידה, דורשת הרבה חישובים. אולם לפי [15] אפשר לנצל את העובדה שהתנאי  $E_i > 0$  עבור  $1 \leq i \leq p$  הוא הכרחי ומספיק ליציבות המסנן. ולכן על סמך (5.18) ו- (5.21) מתקבל שתנאי אקויוולנטי ליציבות המסנן הוא

$$|K_i| < 1 \quad 1 \leq i \leq p \quad (5.31)$$

ותנאי זה קל לבדיקה מתוך (5.20).

פרק 6: פרטי מימוש המערכת הכוללת דחיסה בזמן.

בפרק זה נדון במימוש המערכת לזיהוי מילים בודדות הכוללת דחיסה, באמצעות מיני מחשב NOVA 2. הדיון יתפתח בהתאם לשלבי הפעולה הקיימים במערכת. תכנית המחשב (תכנית ראשית ושגרות), המבצעת את תהליך הזיהוי תקרא בפרק זה בשם תכנית הזיהוי, והיא מודפסת בנספח א'.

דגימת ההגייה וקביעת גבולותיה.

6.1

הפעולות המתוארות בסעיף זה מהוות את הצעד הראשון בשלב עיבוד הקדם. מכיוון שהמערכת פותחה עבור מימוש בתכנה בלבד, כל תהליך הזיהוי מתבצע על הייצוג הספרתי של אות הדיבור, כלומר על דגימותיו. לכן יש לדגום את האות האנלוגי המכיל הגייה לפני תחילת תהליך הזיהוי שלה. האות האנלוגי הזה מתקבל באחת משתי צורות: הדובר הקליט את הגייתו במכשיר הקלטה אנלוגי, במערכת הנדונה המכשיר היה cassette recorder מתוצרת AKAI, או אמר אותה אל מיקרופון (מתוצרת SONY) שהפיק אות חשמלי אנלוגי להגייה.

לפני ביצוע הדגימה מועבר האות האנלוגי דרך מסנן המעביר תחום של 200:3200 Hz. תחום תדרים זה מכיל כמעט את כל האינפורמציה הספקטרלית של אות הדיבור והוא מקובל בקוי טלפון. לאחר הסינון עובר האות האנלוגי לממיר אות אנלוגי לספרתי A/D converter של NOVA 2, בו הוא נדגם במדירות של 10 KHz כאשר כל דגימה היא בת 12 סיביות.

מכיוון שמיותר להעביר דגימות לזכרון המחשב לפני שמופיע הדיבור באות הנדגם, יש צורך לגלות את התחלת ההגייה - את גבולה הראשון. במערכת שפותחה מעשה הדבר כך. כאשר מעונינים להעביר אל המערכת הגייה לזיהוי, כלומר להעביר אות אנלוגי המכיל את ההגייה אל ערוץ ה-A/D, מעלים רמת מתח כ-5 V בערוץ אחר של ה-A/D. שגרה מתאימה (A2DM1) בתכנית הזיהוי דוגמת ערוץ זה, כלומר מפעילה את ה-A/D עבורו, וכאשר היא מגלה את רמת המתח היא פונה לדגימת הערוץ שבו עתיד להופיע אות הדיבור לאחר הסינון כנ"ל. כל דגימה נבדקת ע"י שגרה וכאשר נספרות 5 דגימות שמעל לערך מתאים, מחליטה השגרה שההגייה החלה. ערך זה תלוי בסביבה שבה הוקלטה ההגייה או נאמרה אל המיקרופון. כאשר הסביבה היא שקטה, ללא רעש של מערכת מחשב או מזגן, ערך זה הוא 20, והוא



מתאר  $50 \text{ mV}$  בערך, מכיוון שערך דגימה של  $2047$  מתאים ל- $5 \text{ V}$ . ההנחה שביסוד ההחלטה הנייל היא ש- $5$  דגימות, כל אחת בערך שלמעלה מ- $50 \text{ mV}$  מתארות התחלתה של הגייה, כאשר הסביבה היא כנייל, וכל עוד לא נמנו  $5$  דגימות כאלה, כל אשר נדגם מהווה רעש רקע שאינו מכיל דיבור. כל הערכים הללו ( $5$  דגימות בערך של  $20$ ) ועצם השיטה לגילוי התחלת ההגייה התקבלו מתוך ניסויים עבור הגיות רבות בסביבה כנייל. כאשר הסביבה היא רועשת כבחדר המחשב שבו מופעלת מערכת המחשב כולל יחידת הסרט המגנטי, הערך המינימלי עבור  $5$  הדגימות נקבע ל- $80$  (כ- $200 \text{ mV}$  בערך). גם ערך זה התקבל מתוך ניסויים רבים בסביבה רועשת. הוא גבוה יותר מהערך שנמצא עבור סביבה שקטה, בגלל הגברת רעש הרקע שבו טובלה ההגייה. שימוש בערך קטן יותר הביא לכך ש- $5$  הדגימות שנספרו תארו למעשה רעש רקע ולא את תחילת ההגייה, ולכן התקבלה החלטה לא נכונה על התחלת ההגייה.

כאשר הוחלט שההגייה החלה מועברות  $9945$  הדגימות הבאות מה- $A/D$  אל זכרון המחשב. מספר זה נבחר בהתאם לאורכי המילים שיש לזהות, וכן בהתאם לקיבול הזכרון הנתון במיני מחשב NOVA 2. זכרון זה מכיל  $32 \text{ K}$  מילים בנות  $16$  סיביות.  $8 \text{ K}$  מילים דרושות עבור מערכת ההפעלה (RDOS), והקובץ המכיל את תכנית הזיהוי בשפת מכונה דורש כ- $9 \text{ K}$ . מ- $15 \text{ K}$  המילים הנוותרות יש להקצות כ- $5 \text{ K}$  עבור מערכי נתונים (שאינם דגימות הגייה), ולכן עבור דגימות ההגייה נותרו כ- $10 \text{ K}$  מילים בזכרון המחשב.

הפעולה הבאה שמתבצעת בשלב עיבוד הקדם היא בדיקה וגילוי האם מתח האות האנלוגי של ההגייה לא יצא במשך ההגייה מהתחום של  $\pm 5 \text{ V}$  שהוא תחום העבודה של ה- $A/D$ . גלישת האמפליטודה של האות האנלוגי מתחום זה בזמן כלשהו תגרום לכך שהאות הדגום שיתקבל בזמן זה יהיה חתוך ולכן הספקטרום שלו יכיל עיוותים. תופעה זו תשבש את הזיהוי ולכן נדחית כל הגייה שהבדיקה הנייל גילתה גלישה של אמפליטודת האות האנלוגי שלה. הבדיקה נעשית כך, תכנית הזיהוי סוקרת את הערכים המוחלטים של הדגימות וכאשר היא מגלה דגימה שערכה המוחלט אינו קטן מ- $2047$  (כ- $5 \text{ V}$ ), היא מפסיקה את תהליך הזיהוי ומודיעה על ביטולו עקב הגלישה. אם לא התגלתה גלישה של אמפליטודת האות האנלוגי, תהליך הזיהוי מתקדם ועובר לפעולה האחרונה של הצעד הראשון בשלב עיבוד הקדם - קביעת סוף ההגייה, כלומר גבולה השני.

כאמור לעיל, בזכרון המחשב אגורות  $9945$  דגימות המכילות הגייה, כאשר לפי החלטה קודמת הדגימה הראשונה מהווה את התחלת ההגייה. אולם לגבי הדגימות

האחרונות, יש צורך להחליט על הדגימה האחרונה שעדיין שייכת להגייה, ואז אלה הבאות אחריה מתארות רעש רקע בלבד. החלטה זו מתחייבת מהעובדה שמילות המילון של המערכת - עשר הספרות העבריות (אפס עד תשע), נמשכות באופן טבעי פחות מאשר 994.5 msec (פרק הזמן המתאים ל- 9945 דגימות ב- 10 KHz תדר דגימה), ולכן לאחר שההגייה מסתיימת נדגם גם רעש רקע שיש להתעלם ממנו בשלבים הבאים של תהליך הזיהוי, כי הוא עלול לגרום לשגיאה בזיהוי.

במערכת שפותחה מבצעים את ההחלטה הקובעת את הגבול השני של ההגייה בדרך הבאה. תכנית הזיהוי סוקרת את מערך הדגימות החל מסופו. כאשר נספרות 5 דגימות מעל לערך מתאים, מחליטה התכנית שהיא הגיעה לסיום ההגייה - לגבולה השני וכל הדגימות שמגבול זה עד לדגימה ה- 9945 מהוות רעש רקע שמתעלמים ממנו בהמשך תהליך הזיהוי.

הערך המתאים שנקבע עבור 5 הדגימות בהחלטה זו תלוי בסביבה שבה הוקלטה ההגייה או נאמרה אל המיקרופון, וכן הוא תלוי בערך השיא של אמפליטודת ההגייה הדגומה. ערך זה מחושב לפי הביטוי הבא:

$$\text{Max} \{ [(Fe \cdot LEVT \cdot \text{Max} |x(i)| / 2048) + 0.5], LEVT \} \quad (6.1)$$

[y] מסמן השלם הקטן מ- y.

עבור סביבה שקטה ערכו של LEVT, המופיע ב- (6.1), הוא 20 (כ- 50 mV) ועבור סביבה רועשת ערכו 72 (כ- 180 mV). Fe עבור שתי הסביבות נקבע באופן נסויי ל- 1.25. {x(i)} היא סדרת הדגימות בזכרון המחשב ו-  $\text{Max} |x(i)|$  הוא ערכה המוחלט של דגימת השיא מתוך {x(i)}. הגדרת הסביבות היא כהגדרה הנ"ל.

תלות (6.1) בשיא ההגייה מבוססת על ההנחה שאמפליטודת זנב ההגייה נמצאת בייחס ישר לשיא ההגייה. ערכיו של LEVT עבור הסביבות השונות והכפל בגורם  $Fe = 1.25$  התקבלו על סמך ניסויים. בניסויים אלה התברר גם שאין זה רצוי שערכן המינימלי של 5 הדגימות ירד מערכו של LEVT.

יש לציין שקביעה לא מהימנה של גבולות ההגייה עלולה לחבל בדיוק הזיהוי.

אמנם, קביעה מדויקת שלהם, במיוחד בסביבה רועשת, מהווה ענין מסובך ואולי אף בלתי אפשרי עבור חלקי דיבור מסוימים, לדוגמא, "ת" בסוף הספרה אחת, "מ" בסוף הספרה שתיים, אולם היא אינה הכרחית. כל אשר נדרש הוא לקבוע את גבולות ההגיייה כך שתכיל את כל התכונות האקוסטיות המשמעותיות הדרושות לצורך מיונה הנכון, ויחד עם זאת לא תכיל רעש רקע בכמות העלולה לפגוע בדיוק המיון. הדרישות האלה מרמזות על כך שכל אלגוריתם לגילוי גבולות ההגיייה יתבסס על ניסויי שמיעת ההגיות לאחר שהוגבלו ובדיקת הדיוק בזיהוין. כך גם נבנה האלגוריתם שמימשו תואר בסעיף זה. אך בצד השיקול של השגת מהימנות גבוהה, היה צורך להתחשב בזמן בנית האלגוריתם, בדרישה המרכזית - חסכון בזמן חישוב, ועל סמך שני שיקולים אלה נוצר האלגוריתם שבמערכת זו.

לאחר ביצוע ההחלטה על סיוס ההגיייה ידוע לתכנית הזיהוי ארכו של וקטור הדגימות המתאר את ההגיייה. ומכאן עובר וקטור זה אל הצעד הבא בשלב עיבוד הקדם, צעד הדחיסה בזמן. מימוש צעד זה יתואר בסעיף הבא.

#### מימוש האלגוריתם לדחיסה בזמן של ההגיייה.

6.2

בצעד השני של שלב עיבוד הקדם מתבצעת דחיסה בזמן של ההגיייה לאורך האחיד עבור כל ההגיות. מספר דגימות ההגיייה העוברות דחיסה ( $L_j$ ) נקבע בצעד הראשון כמתואר בסעיף 6.1. בסעיף זה נדון בפרטי המימוש של האלגוריתם לדחיסה בזמן, שתואר בפרק 4, במערכת שפותחה.

לצורך המימוש יש לבחור בערכים מתאימים עבור הפרמטרים של האלגוריתם. בחירה זא נעשית כמובן לפי השיקולים שתוארו בסעיף 4.2.

ערכי הפרמטרים שנבחרו היו:  $m=2$ ,  $N_p=200$ ,  $L_{c_0}=3000$ , ופונקציה של חלון משולש כנתון ב- (4.27) נבחרה עבור  $h(t)$ . מכיוון שאף אחת מההגיות שזיהוין נבדק במערכת, לא נמשכה יותר מ- 0.95 שניה (כלומר 9500 דגימות עבור קצב הדגימה שבמערכת - 10 KHz), אזי לפי (4.28) - (4.30) והערכים הנייל מתקבל ש-  $C_j < 3.09$  כרצוי. הערך  $L_{c_0} = 3000$  נקבע מכיוון שהוא מבטיח פרט לשמירה על ההגבלה עבור  $C_j$ , גם את הקבלה של מספר שלם של קטעי זמן מהם ימוצו הפרמטרים (ראה סעיף 6.3).

מטעמי הדרישה לחסכון בזמן חישוב שהיא הדרישה המרכזית עבור המערכת שפותחה, נבחר  $m=2$  כי הוא הערך המינימלי האפשרי עבור  $m$ . מכיוון שמשתמשים בפונקציה

חלון משולש סימטרית, זמן החישוב הנדרש לביצוע הדחיסה הוקטן עוד יותר. כאשר משתמשים בביטוי (ראה 4.27):

$$\hat{h}(t) = 1 - \hat{h}(t - N_p T), \quad N_p T \leq t \leq 2N_p T \quad (6.2)$$

אזי הביטוי הכללי לחישוב דגימת יציאה  $Y$ , כפי שנתון בתרשים הזרימה למימוש האלגוריתם שבציור 4.1, מוחלף ב-

$$Y = h_o(K) [S(k + N_p) - S(k)] + S(k) \quad (6.3)$$

כאשר

$$h_o(k) = k/N_{sc}, \quad k = 1, 2, \dots, N_{sc} \quad (6.4)$$

מחליף את  $\hat{h}(k)$  שבתרשים הזרימה של ציור 4.1. עקב שינוי זה, דרושים רק פעולת כפל אחת ושתי פעולות חיבור עבור חישוב דגימת יציאה אחת, במקום שני כפלים וחיבור אחד במקרה הרגיל עבור  $m=2$ . לכן, לחישוב  $L_{c0} = 3000$  דגימות יציאה של הגייה דחוסה, נדרשו 3000 כפלים ו- 6000 חיבורים.

עבור  $N_p$  נבחר הערך 200 על סמך מספר סיבות. מתוך נסויי מובנות שנערכו עבור הגיות דחוסות עם ערכים שונים של  $N_p$ , התברר שהערך 200 מאפשר מובנות טובה יותר של ההגייה הדחוסה. פרט לכך, ערך זה מבטיח ש-  $N_p$  יהיה בדרך כלל גדול ממחזור pitch כלשהו העשוי להופיע בהגייה נבדקת. ולבסוף,  $N_p = 200$  פרושו עבור  $m=2$ :  $(m-1)N_p = 200$ . כלומר, עבור קצב דגימה של 10 KHz, בזמן חישוב דגימת יציאה דחוסה, ילקחו עבור השקלול דגימות מקטע זמן של 20 msec, כך שמובטח שהן שייכות לאותו קטע סטציונרי.

לסיכום, נדגים בציורים 6.1, 6.2 ו- 6.3 את תוצאות הדחיסה עבור המילה "שלוש". העיבוד לפני הדחיסה והדחיסה עצמה נעשו במערכת שמומשה, כמתואר בפרק הנוכחי. מציור 6.1 נובע שלאות הדחוס יש כמעט אותו מחזור pitch (כ- 7 msec בערך), שיש לאות המקורי. מציור 6.2 נובע שהאות הדחוס תופש

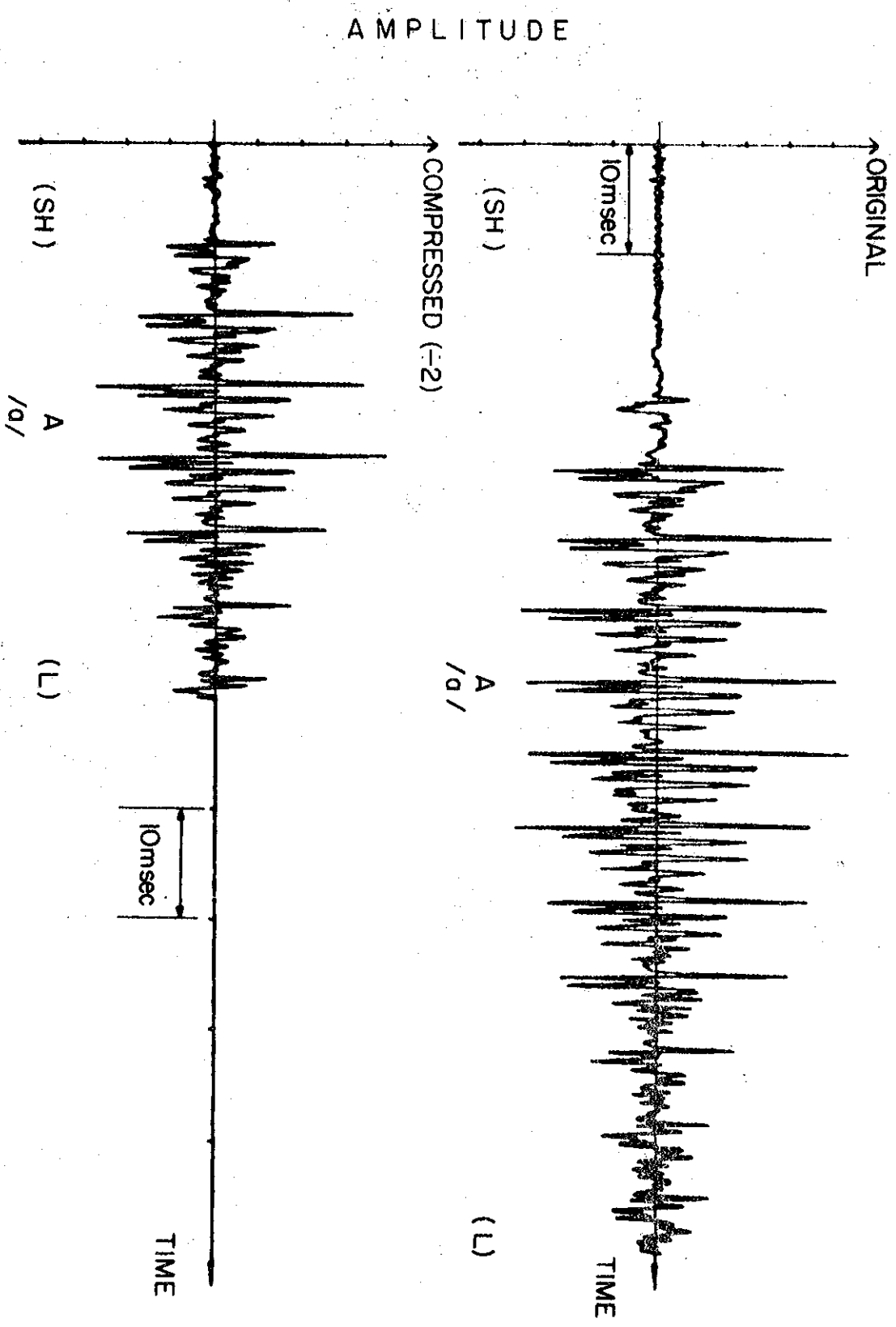
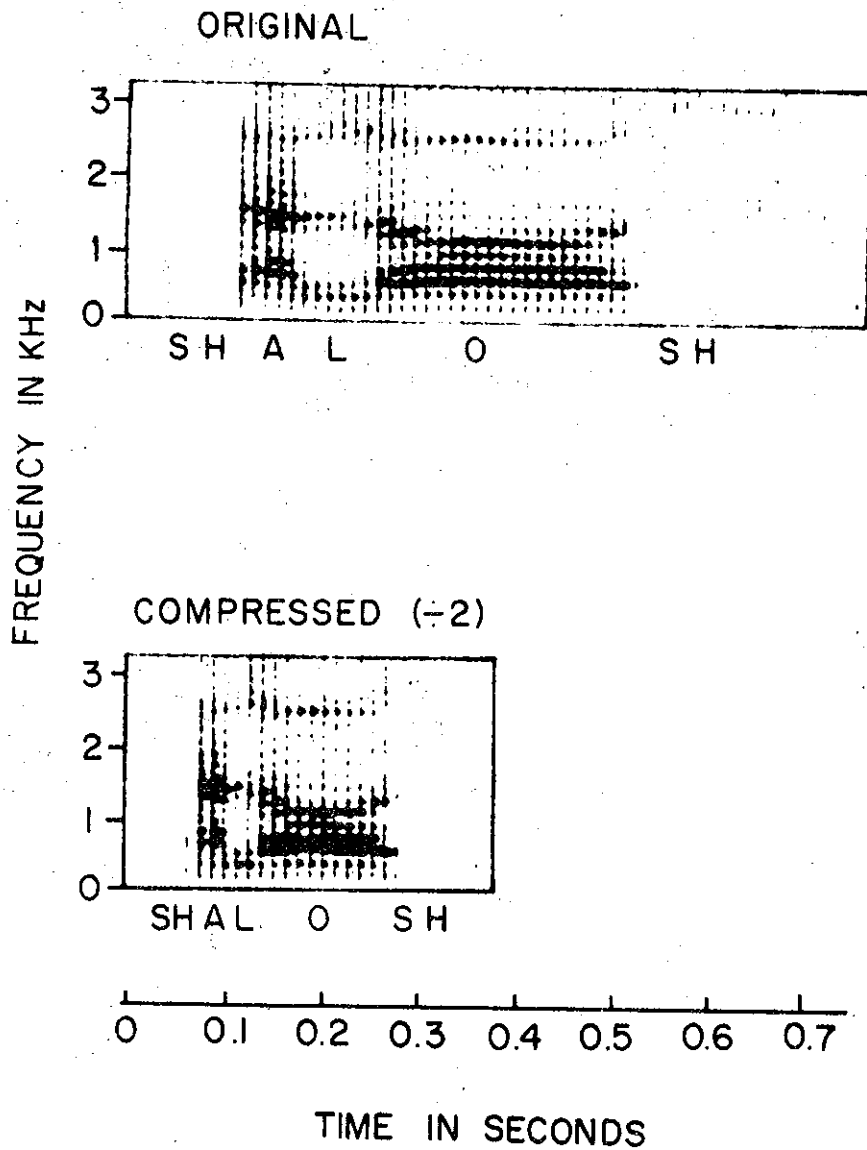
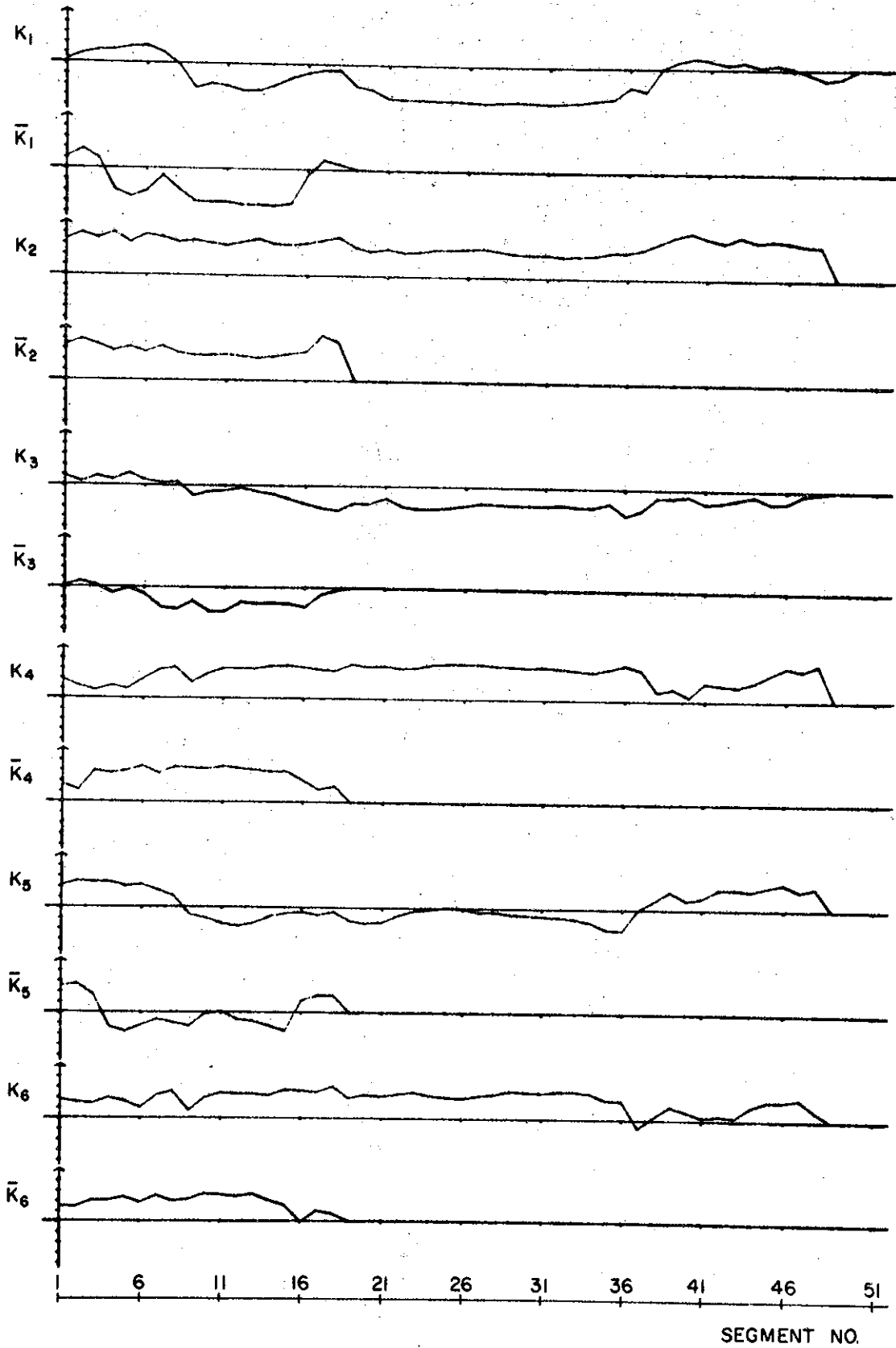


Fig. 6.1: Original (upper wave form) and time compressed, by a factor of 2, (lower wave form) of the voiced part of the first syllable in the Hebrew word "shalosh" /shaloʃ/.  
 ציור 6.1: החלק הקולי המקורי (צורת הגל העליונה), והחלקים בזמן פי 2 (צורת הגל התחתונה), של התבונה הראשונה במילה "שלוש" /שלוש/.



צִיּוֹר 6.2: ספקטרוגרמות צרות סרט שנוצרו באמצעות מחשב, של המילה "שלוש" המקורית (הספקטרוגרמה העליונה), ושל הדחוסה פי 2 (הספקטרוגרמה התחתונה).

Fig 6.2: Computer generated narrow band spectrograms of the original (upper spectrogram) and compressed by two (lower spectrogram) of the Hebrew word "shalosh".



ציור 6.3: מקדמי הקורלציה החלקית של המילה "שלוש" המקורית  $(K_i, i=1,2,\dots,6)$  ושל הדחוסה בזמן  $(\bar{K}_i, i=1,2,\dots,6)$ .

Fig 6.3: Partial Correlation Coefficients of the original  $(K_i, i=1,2,\dots,6)$  and time compressed  $(\bar{K}_i, i=1,2,\dots,6)$  signal of the Hebrew word "shalosh".

אותו תחום תדרים כאות המקורי ושלטניהם מבנה ספקטרלי דומה. הספקטרוגרמות של ציור 6.2 הן צרות סרט (40 Hz) בכדי להראות את ההרמוניות של ה-pitch. ציור 6.3 מראה את השפעת הדחיסה על הפרמטרים המייצגים את ההגייה - מקדמי ה-PARCOR. המקדמים שמוצו מההגייה הדחוסה נראים לפי ציור 6.3, כאילו התקבלו מדחיסה ושקלול של המקדמים שמוצו מההגייה הלא דחוסה. תוצאה זו צפויה, מכיוון שאלגוריתם הדחיסה משקלל מספר קטעי זמן של הגיית הכניסה כדי ליצור קטע זמן אחד של הגייה דחוסה.

תוצאות דומות התקבלו גם עבור הגיות אחרות שזיהוין נבדק.

### 6.3 חישוב מקדמי הקורלציה החלקית

חישוב מקדמי הקורלציה החלקית (מקדמי ה-PARCOR) מבוצע בשלב מיצוי תכונות לפי שיטת האוטוקורלציה (ראה סעיף 5.2). השיטה הזאת נבחרה על סמך הנימוקים הבאים:



1. מבחינת טיב הקרוב של המודל המבוסס על מסנן כולו קטבים, שתי השיטות מתלכדות כאשר קטע הזמן ממנו מחשבים את המקדמים, נמשך 2 עד 3 מחזורי pitch (ראה סעיף 5.4). שיטת הקוואריאנס משיגה אמנם קרוב טוב יותר כאשר קטעי הזמן נבחרים בשיטת pitch synchronous (PS) (ראה סעיף 5.4), אולם לשם כך יש לגלות את רגע הופעת הלם ה-pitch בדיבור ופעולה זו דורשת כמות חישובים רבה נוספת.

2. שיטת האוטוקורלציה בניגוד לשיטה השניה, מבטיחה את יציבות המסנן באופן תיאורטי, ולמעשה מושגת יציבות זו בחישובים הנעשים בנקודה צפה (floating point) [18].

כעת נעבור לדיון בפרטי החישוב של מקדמי ה-PARCOR. תחילה נדון בבחירת אורך קטע הזמן ממנו ימוצו מקדמי ה-PARCOR. נימוך מסי 4 הני"ל, מחייב בחירה בקטע זמן הנמשך 2 עד 3 מחזורי pitch. על סמך נסויים שנעשו במטרה לקבוע את אורך הקטע (מספר הדגימות שבו), המתאים ביותר לקבלת מידע ברור על הפורמנטים שבקטע, הסיק Markel [19] שאורך הקטע הנדרש תלוי רק בצורה חלשה בהגייה שאותה מעבדים, ובצורה חזקה בתדר הדגימה. הוא גם המליץ על שמירת הקשר  $N = \delta F_s$ , כאשר  $F_s$  תדר הדגימה (ב-KHz) ו- $\delta$  גורם בעל תחום



השתנות בין 20 ל- 35, על מנת לקבל גבולות סבירים עבור N - מספר הדגימות הנדרש בקטע הזמן. לכן, מכיוון שמחזור ה-pitch הצפוי לא יעלה על 15 msec ו-  $F_s = 10 \text{ KHz}$ ,  $N=300$  (עבור תדר דגימה של 10KHz), הוא ערך שיתאים לשתי המסקנות הנ"ל, והוא הערך שנבחר עבור אורך קטע הזמן ממנו מוצו מקדמי ה-PARCOR.

כאשר מחשבים את מקדמי ה-PARCOR לפי שיטת האוטוקורלציה, כופלים את דגימות ההגייה בפונקציית חלון ראה סעיף (5.2). אורך החלון הוא N הנ"ל, הקובע כמובן, עבור תדר דגימה נתון, את אורך קטע הזמן עבורו מחשבים את המקדמים. פונקציית החלון הפשוטה ביותר היא חלון מלבני, כלומר דגימות ההגייה מקטע הזמן הנדון נלקחות באופן ישיר לחישוב המקדמים. אולם, לפונקציה זו יש חסרון והוא שעקב התכונה שיש להתמרת פוריה של החלון המלבני - אונות הצד (side lobes) שלה אינן מונחתות בצורה מספיק חזקה יחסית לאונה הראשית, מתקבלת תופעת זליגה בתחום התדר הגורמת להחלשת הפורמנטים השני והשלישי [19]. פונקציית חלון מתאימה לפי Markel [19] היא חלון Hamming:

$$w(n) = 0.54 - 0.46 \cos [2\pi n / (N-1)] \quad n=0,1,\dots,N-1$$
$$= 0, \quad \text{elsewhere}$$

(6.5)

והיא מקובלת גם אצל Itakura [2], Ichikawa [9], ואצל White and Neely [6]. לפיכך, עבור המערכת הנדונה נבחר חלון Hamming לפי (6.5) לשמש כפונקציית חלון.

לפני הכפל בחלון עובר אות הדיבור פעולת גזירה ע"י החסרת הדגימה הנוכחית מהדגימה הבאה אחריה. לפיכך, הדגימות הנלקחות לחישוב המקדמים ע"י תכנית הזיהוי הן:

$$\chi(n) = w(n) [s(n+1) - s(n)] , \quad n=0,1,\dots,N-1$$

(6.6)

כאשר  $\{s(n)\}$  הן דגימות ההגייה החדוטה, ~~מנומנות, כך ש-~~  $-1 < s(n) < 1$  וכאשר מניחים  $s(N) = 0$ , ו-  $w(n)$  נתונה ב-(6.5).

גזירה זו היא אמצעי חסכוני מבחינת זמן חישוב לביצוע preemphasis של 6 dB/octave. כלומר, הגזירה יחד עם פונקציית החלון גורמים להגברת העצמה של התדרים הגבוהים בספקטרום האות, כך שמודל המסנן כולו קטבים יקרב אותם ביתר דיוק. נוסף לכך, הגזירה גורמת לסילוק ה-d.c מאות הדיבור, שיתכן שהצטרף לאות הדיבור בזמן הדגימה, ובכך מונעת הדגשה מיותרת של התדרים הנמוכים סביב האפס.

בכדי לקרב נאמנה שינויים מהירים בדיבור מחשבים את מקדמי ה-PARCOR מתוך קטעי זמן החופפים, כל אחד באופן חלקי את קודמו. כלומר החלון הנייל אינו מתקדם כל פעם בצעד של  $N$  דגימות כאשר  $N_K < N$ . ככל ש- $N_K$  יהיה קטן יותר, נקבל יותר קטעי של  $N_K$  דגימות. כאשר  $N_K < N$ . ככל ש- $N_K$  יהיה קטן יותר, נקבל יותר קטעי זמן לחישוב המקדמים וזמן החישוב יגדל. לכן בכדי לחסוך בזמן חישוב אך גם לשמר על קצב עדכון מתאים של המקדמים, נבחר  $N_K = 150$  כלומר 15 msec עבור תדר דגימה של 10 KHz. בחירת  $N_K = 100$  לדוגמא, מוסיפה 50% חישובים עבור המיצוי ביחס ל- $N_K = 150$ . Itakura במערכת המתוארת ב-[2] בחר גם כן בחלון הנמשך 30 msec והמתקדם בצעד של 15 msec.

לפיכך, עבור הגייה המכילה לאחר הדחיסה  $L_{c_0} = 3000$  דגימות. מספר קטעי הזמן לחישוב המקדמים -  $L_s$  יהיה:  $L_s = (L_{c_0} - N_w) / N_K + 1 = 19$  כאשר  $N_w \equiv N = 300$  ו- $N_K = 150$  כמתואר לעיל.

מכיוון שמספר החישובים בשלב המיצוי הוא יחסי ל- $p$ , אזי השאיפה היא לבחור את  $p$  המינימלי שעדיין אינו פוגע בביצועי המערכת. בעקבות עבודותיהם של Itakura [2] ו-Ichikawa [9] נבחר  $p = 6$  כערך המקובל עבור המערכת. סביר ש- $p = 6$  יתאר נאמנה את 3 הפורמנטים הראשונים ומכיוון שאות הדיבור לפני דגימתו מועבר דרך מסנן מעביר תחום של 200-3200 Hz והפורמנטים הגבוהים יותר יונחתו או יסוננו, לא יפגע הדיוק של יצוג האות באמצעות 6 מקדמי PARCOR.

חישוב מקדמי ה- PARCOR מבוצע על סמך המשוואות הרקורסיביות (5.19) - (5.21), המהוות כיום את הדרך המהירה ביותר לחישובם. השגרה המחשבת את המקדמים בתכנית הזיהוי נקראת LPF (ראה נספח א'), והיא מבוססת על השגרה שהוצגה ע"י Markel and Gray [20].

מספר הפעולות בחישוב  $pL_s = 114$  מקדמים עבור הגייה נבדקת נתון בסעיף 3.2: 45,600 כפלים ו- 39,900 חיבורים.

#### 6.4 תהליך המיון

תהליך המיון מתבצע בשלב האחרון של תהליך הזיהוי. במערכת שפותחה, נמצאים בזכרון המחשב 10 וקטורי ייחוס המיצגים כל אחד מילה אחרת של המילון. כאשר מתבצע תהליך זיהוי של הגייה נבדקת, מתחיל תהליך המיון עם סיום חישוב וקטור הפרמטרים (מקדמי ה- PARCOR) המייצג את ההגייה הנבדקת. בתחילת תהליך המיון מחשבת תכנית הזיהוי את מידת הדמיון  $S_{r,j}$  בין וקטור ההגייה הנבדקת בתצוין ע"י  $j$ , לבין כל אחד מוקטורי היחוס שיצויינו ע"י  $r$ ,  $1 \leq r \leq 10$ , לפי הביטוי

$$S_{r,j} = \max_q \{1 - d_{r,j}^q / [2p(L_s - 2)]\}; \quad q = 1, 2, 3 \quad (6.7)$$

כאשר

$$d_{r,j}^q = \sum_{\ell=q}^{L_s-3+q} \left( \sum_{i=1}^p |K_{i1}^j - K_{im}^r| \right); \quad q = 1, 2, 3 \quad (6.8)$$

כאשר  $m = 1 - q + 1$  ו-  $K_{in}$  הוא מקדם ה- PARCOR ה-  $i$  מתוך  $p$  מקדמים המיצגים את קטע הזמן ה-  $n$ . במערכת שפותחה  $p = 6$ .

הביטוי (6.8) מבטא את שיטת ההזדהות שתוארה בסעיף 3.3, לפיה  $L_s - 2 = 17$  קטעי זמן של הגיית היחוס מושוים בעזרת מרחק צ'ביצ'ף, 3 פעמים עם ההגייה הנבדקת המכילה לאחר הדחיסה  $L_s = 19$  קטעי זמן, כאשר בכל פעם

מתבצעת הזזה מתאימה. לפי (6.7) מידת הדמיון בין שני הוקטורים המשווים, היא המידה המקסימלית שהתקבלה משלושת ההשוואות. המכנה שבביטוי זה הוא גורם שבא לנרמל את  $d_{r,j}^q$ . כאמור בסעיף (5.5)  $|K| < 1$ , לכן סכום של  $p$  ערכים מוחלטים של הפרשי  $K$ -ים יתן גודל הקטן מ-  $2p$ . ומכיוון שמסכמים ב- (6.8)  $(L_S - 2)$  סכומים כאלה, יש לחלק את  $d_{r,j}^q$  ב-  $2p(L_S - 2)$  כדי להבטיח שיתקיים תמיד  $0 < S_{r,j} \leq 1$ . (במיוחד  $0 < S_{r,j}$ ).

לאחר שהתקבלו  $\{S_{r,j}\}$ ,  $1 \leq r \leq 10$ , מחליטה תכנית הזיהוי שהגגייה הנבדקת  $j$  שייכת למילה ה- $r$  שעבורה התקבל  $\text{Max}_r (S_{r,j})$ . תכנית הזיהוי מודיעה את החלטתה - שהמילה היא:  $r$  ומידת הדמיון שלה היא:  $\text{Max}_r (S_{r,j})$ . ובכך מסתיים הגגייה נבדקת. תהליך הזיהוי עבור/מספר פעולות החישוב, בתהליך המיון (פרט למספר הזניח של חיבורים לצורך מציאת  $\text{Max}_r (S_{r,j})$ ), הוא 9180 חיבורים הדרושים לחשוב  $\{S_{r,j}\}$ ,  $1 \leq r \leq 10$ , כפי שצויין כבר בסעיפים 3.2 ו- 3.3.

פרק 7: פיתוח מערכת השוואתית ללא דחיסה.

7.1 מבוא

מערכת לזיהוי מילים בודדות הפועלת ללא דחיסה, פותחה על מנת שניתן יהיה להשוות כלפיה את הביצועים של המערכת הכוללת דחיסה. כלומר, מתוך השוואת תוצאות הזיהוי שתתקבלנה בשתי המערכות, ניתן יהיה לדעת עד כמה אם בכלל, פוגעת הדחיסה בזמן של ההגיות, בהצלחת הזיהוי.

מכיוון שמעוניינים לבדוק את התלות של הישגי הזיהוי רק בגורם אחד - דחיסה, שתי המערכות חייבות להיות זהות פרט לכך שבאחת יופעל תהליך הדחיסה ובשניה לא. כמובן שהבדיקות חייבות להעשות עבור אותן הגיות. לפיכך, במערכת ההשוואתית שפותחה דגימת ההגיה וקביעת גבולותיה מתבצעים כמו במערכת הכוללת דחיסה, כמתואר בסעיף 6.1. הדחיסה המתוארת בסעיף 6.2 איננה מתבצעת, ולכן לאחר קביעת גבולות ההגיה, מתבצע מיד תהליך חישוב מקדמי ה- PARCOR המקובל במערכת האחרת כמתואר בסעיף 6.3. אך בהבדל אחד - במערכת ההשוואתית אורך ההגיות אינו אחיד ולכן מספר קטעי הזמן מהם מחשבים את המקדמים משתנה מהגיה להגיה כך שמתקבלים וקטורי פרמטרים באורך שונה, כל וקטור בעל אורך היחסי למשך ההגיה אותה הוא מייצג. במצב זה, יש צורך לבצע בשלב המיון התאמה של צירי הזמן של שני וקטורי הפרמטרים, אשר את מידת הדמיון ביניהן מעוניינים לחשב. שיטה קיימת לבצוע פעולה זו שהשיגה עד כה את התוצאות הטובות ביותר, היא השיטה המבוססת על תכנות דינמי dynamic programming (ראה סעיף 2.3.2), ולכן, היא נבחרה למימוש תהליך המיון במערכת ההשוואתית.

בסעיף הבא של הפרק נדון בשיטה שנבחרה ופותחה עבור המערכת ההשוואתית.

7.2 התאמת צירי הזמן ומיון באמצעות תכנות דינמי.

קיימות מספר שיטות לביצוע התאמת צירי הזמן ומיון באמצעות תכנות דינמי, ביניהן השיטות של Itakura [2], Ichikawa [9] ושל Velichko [10]. השיטה שנבחרה ופותחה עבור המערכת ההשוואתית היא השיטה של Velichko.

נדון תחילה בשיטה הזאת ובפיתוחה, ואחר נסביר במה יתרונה על פני השתיים האחרות.

לפנינו שתי הגיות, הגייה נבדקת והגיית יחוס שבדרך כלל הן בעלות אורך שונה. יש לחשב את מידת הדמיון בין וקטורי הפרמטרים המיצגים אותן. לשם כך נחשב את מטריצת הדמיון עבור שתי ההגיות. אלמנט  $s_{i,j}$  במטריצת הדמיון הוא מידת הדמיון בין קטע הזמן ה-  $j$  של ההגייה הנבדקת לבין קטע הזמן ה-  $i$  של הגיית הייחוס (למעשה בין הרכיבים מוקטורי הפרמטרים, המיצגים קטעים אלה). כאשר  $j = 1, \dots, N$  מציינ כל אחד מ-  $N$  קטעי הזמן של ההגייה הנבדקת ו-  $i = 1, \dots, M$  מציינ כל אחד מ-  $M$  קטעי הזמן של הגיית הייחוס. נחלק מלבן שרחבו  $N$  וארכו  $M$  ל-  $M \times N$  ריבועים שווים, כאשר הצלע העליונה מתארת את ציר הזמן של ההגייה הנבדקת שעליו מסומנים  $N$  קטעי זמן שווים, והצלע השמאלית מתארת את ציר הזמן של הגיית הייחוס שעליו מסומנים  $M$  קטעי זמן שווים. הקדקד השמאלי העליון של המלבן יהיה ראשית הצירים  $B$  ובה מתלכדות ההתחלות של שתי ההגיות המשוות, ובקודקוד הימני התחתון  $E$  מתלכדים סיומי ההגיות, ראה ציור 7.1. כעת נרשום בכל רבוע את האלמנט המתאים ממטריצת הדמיון הנ"ל  $s_{i,j}$ , כלומר  $s_{i,j}$  ירשם בריבוע שהקודקוד הימני התחתון שלו הוא  $(i,j)$  כמתואר בציור 7.1. מסלול בין  $B$  ל-  $E$ , מוגדר כרצף קטעים מכוונים עוקבים כאשר הראשון מתחיל ב-  $B$  והאחרון מסתיים ב-  $E$ . קטע מכוון הוא אחת משלוש אפשרויות, צלע של רבוע בכוון ציר הזמן של ההגייה הנבדקת או ציר הזמן של הגיית הייחוס, או אלכסון של רבוע שכוונו מ-  $(i-1, j-1)$  אל  $(i,j)$ ;  $1 \leq i \leq M$ ,  $1 \leq j \leq N$ . מסלול אחד מרבים אפשריים מסומן בציור 7.1.

כל מסלול כזה מתאר תאום אפשרי בין צירי הזמן של ההגיות המשוות. תאום המתואר ע"י מסלול מסוים הוא בעל המשמעות הבאה. קטע זמן של ההגייה הנבדקת המתאים לקטע מכוון אפקית, אינו נלקח בחשבון לצורך השוואה בין ההגיות, וכך נוהגים גם בקטע זמן של הגיית הייחוס המתאים לקטע מכוון אנכית. לעומת זאת, קטע מכוון אלכסוני, מציינ שקטעי הזמן של ההגיות, המתאימים לצלעות הריבוע שבו עבור האלכסון, נלקחים לצורך חישוב מידת הדמיון, לדוגמא הקטעים  $i$  ו-  $j$  במסלול שבציור 7.1. לפיכך, מידת דמיון המתאימה למסלול מסוים, היא סכום מידות הדמיון שבין קטעי הזמן שנלקחו להשוואה, כלומר סכום מידות הדמיון הרשומות בריבועים שבכל אחד מהם עובר אלכסון השייך למסלול. מידת הדמיון  $S$  בין ההגיות המשוות, מוגדרת כמידת הדמיון המתאימה למסלול המאפשר עבורה את קיום:

$$S = \text{Max} \left( \sum_{\{k,1\}} s_{k,1} \right) \quad (7.1)$$

כאשר  $\{k,1\}$  מציין את קטעי הזמן שנלקחו לצורך חישוב מידת הדמיון, כלומר  $\{s_{k,1}\}$  רשומים בריבועים שבכל אחד מהם עובר אלכסון השייך למסלול. הבעיה היא למצוא את המסלול הנותן את  $S$ , המוגדרת ב-(7.1). נכנה מסלול זה - המסלול האופטימלי. המסלול האופטימלי מתאר באמצעות אלכסוני את קטעי הזמן משתי ההגיות, הדומים ביותר ואם נסדר על ציר הזמן את האלכסונים לפי סדר הופעתם, נקבל משתי ההגיות המקוריות שתי הגיות חדשות, באותו אורך שהן הדומות ביותר שניתן לקבל מהשתיים המקוריות.

המסלול האופטימלי מתקבל באמצעות תכנות דינמי. למעשה מתקבלת  $S$ , אך ממנה ובעזרת תוצאות הביניים שמהתהליך לקבלתה, ניתן לקבל את המסלול האופטימלי. נתאר כעת את התהליך לקבלת  $S$  באמצעות תכנות דינמי.

נגדיר  $S_{i,j}$  - מידת דמיון חלקית בין חלק ההגיה הנבדקת הנמשך מ- $B$  ל- $j$ , לבין חלק הגיית הייחוס הנמשך מ- $B$  ל- $i$ , כמתואר בצירור 7.1.  $S_{i,j}$  מקיימת את הקשר הרקורסיבי הבא:

$$S_{i,j} = \text{Max} [(S_{i-1,j-1} + s_{i,j}), S_{i-1,j}, S_{i,j-1}] \quad (7.2)$$

תנאי ההתחלה עבור (7.2) הם  $S_{i,0} = 0$ ,  $0 \leq i \leq M$ , ו-  $S_{0,j} = 0$ ,  $0 \leq j \leq N$ . לפיכך, מחשבים את  $S_{i,j}$  לפי (7.2), עם תנאי ההתחלה הנ"ל, עבור  $1 \leq j \leq N$  כאשר  $1 \leq i \leq M$ . מידת הדמיון החלקית המחושבת אחרונה היא  $S_{M,N}$  והיא מהווה את מידת הדמיון  $S$  בין שתי ההגיות המשוויות. (למעשה יש לנרמל את  $S_{M,N}$  בגורם מתאים אך בכך נדון בהמשך).

כמות החישובים לקבלת  $S$  לפי התהליך הנ"ל יחסית ל- $MN$ . זו כמות חישובים רצינית ומעונינים להקטינה. עושים זאת ע"י הגבלת תחום הקיום של המסלולים האפשריים לאזור ברוחב מתאים סביב האלכסון שבין  $B$  ל- $E$ . כעת כמות החישובים יחסית ל- $N_c$ , כאשר  $N_c$  הוא מספר הריבועים שבתחום הקיום הנ"ל. פעולה זו מתבססת על ההנחה, שעבור שתי הגיות של אותה מילה, המרחקים של קטעי הזמן הדומים, ביחס להתחלת ההגיות, אינם נבדלים זה מזה בצורה קיצונית. כלומר, יש סבירות קטנה מאד שקטע הזמן הראשון של ההגיה הנבדקת יהיה דומה לקטע הזמן האחרון של הגיית הייחוס. קביעת רחבו של תחום הקיום הנ"ל, נעשית בצורה ניסויית. כלומר, על סמך תוצאות זיהוי עבור רוחבים שונים נבחר המתאים מבחינת הצלחה בזיהוי וכמות חישובים.

יש לזכור, שעל התהליך הנ"ל לחישוב  $S$ , אנו חוזרים עבור כל אחד מוקטורי הייחוס ובהתאם לשיטת שתוארה, מחפשים את המסלול האופטימלי עבור כל אחד מהם. כלומר, מעונינים למצוא בכל פעם את התאמת צירי הזמן הטובה ביותר בין וקטור הפרמטרים המיזג את ההגיייה הנבדקת לבין וקטור הייחוס הנדון. תהליך התכנות הדינמי מבצע זאת בלי הבחנה כמובן, בין וקטור הייחוס המיזג את המילה לה שיזכת ההגיייה הנבדקת לבין וקטורי הייחוס האחרים. לכן קיימת אפשרות לקבלת טעות בזיהוי, אם לצורך ההחלטה נסתמך רק על  $S_{M,N}$  שנקבל עבור כל אחד מוקטורי הייחוס, בכדי להקטין אפשרות זו עבור המערכת השוואתית, נעשתה הבדיקה הבאה.

כאשר נבדק זהוי של הגיייה, חושב עבור כל אחד מוקטורי הייחוס אורך המסלול האופטימלי, המוגדר כמספר האלכסונים שבמסלול, כלומר, מספר מידות הדמיון הנלקחות בחשבון לצורך הסיכום ב- (7.1). חישוב המסלול האופטימלי וארכו נעשה לאחר שהתקבלה  $S_{M,N}$ , בעזרת מידות הדמיון החלקיות ו-  $S_{M,N}$  בדרך הבאה. תחילה בודקים, מי מבין שלושת האיברים הרשומים בסוגריים [ ] שבביטוי (7.2), שווה ל-  $S_{i,j}$  כאשר מציבים  $i = M, j = N$ . רק אם זה האיבר  $(S_{M-1,N-1} + S_{M,N})$  רושמים ו במונה האלכסונים, מכיוון שהגענו ל-  $S_{M,N}$  מ-  $S_{M-1,N-1}$  באמצעות אלכסון. בכל מקרה, האיבר שבסוגריים השווה ל-  $S_{M,N}$  מגדיר את הצומת ממנה הגענו לצומת  $(M,N)$ . כעת, מציבים ב-  $i, j$  את מציני הצומת שמצאנו (זו שממנה הגענו ל-  $(M,N)$ ) וחוזרים על הפעולה הנ"ל. כך ממשיכים לנוע לאחור על פני המסלול האופטימלי, כאשר בכל צעד מגלים את הצומת שממנה הגענו אל הצומת שמצאנו בצעד הקודם, עד אשר מגיעים לראשית B. כעת לאחר שספרנו את האלכסונים שגילינו במסלולי האופטימלי, ידוע גם ארכו, נסמן אותו ב-  $L_p$ . יש לציין שתמיד  $L_p \leq L_{SH}$ , כאשר  $L_{SH}$  הוא מספר קטעי הזמן של ההגיייה הקצרה יותר בין השתיים המשוות.

בניסויי זיהוי רבים שנערכו עבור אותה ספרייה של וקטורי ייחוס, התברר שבדרך כלל, כאשר נעשתה השוואה עם וקטור הייחוס שיזג את המילה לה שייכת ההגיייה שנבדקה, הייחס  $L_p/L_{SH}$  שהתקבל עבור השוואה זו, היה גדול יותר מאשר התקבל עבור השוואות עם וקטורי הייחוס האחרים.

תוצאה זו הביאה לרעיון להשתמש בייחס  $L_p/L_{SH}$  לנרמול  $S_{M,N}$ , כך שתוקטן האפשרות לקבלת טעות בזיהוי, אותה תארנו קודם. לפיכך, נבדק  $\hat{S}$  המתקבל עקב נרמול  $S_{M,N}$  לפי הביטוי:



$$\hat{S} = (S_{M,N}/L_{SH})(L_p/L_{SH}) = S_{M,N}L_p/L_{SH}^2 \quad (7.3)$$

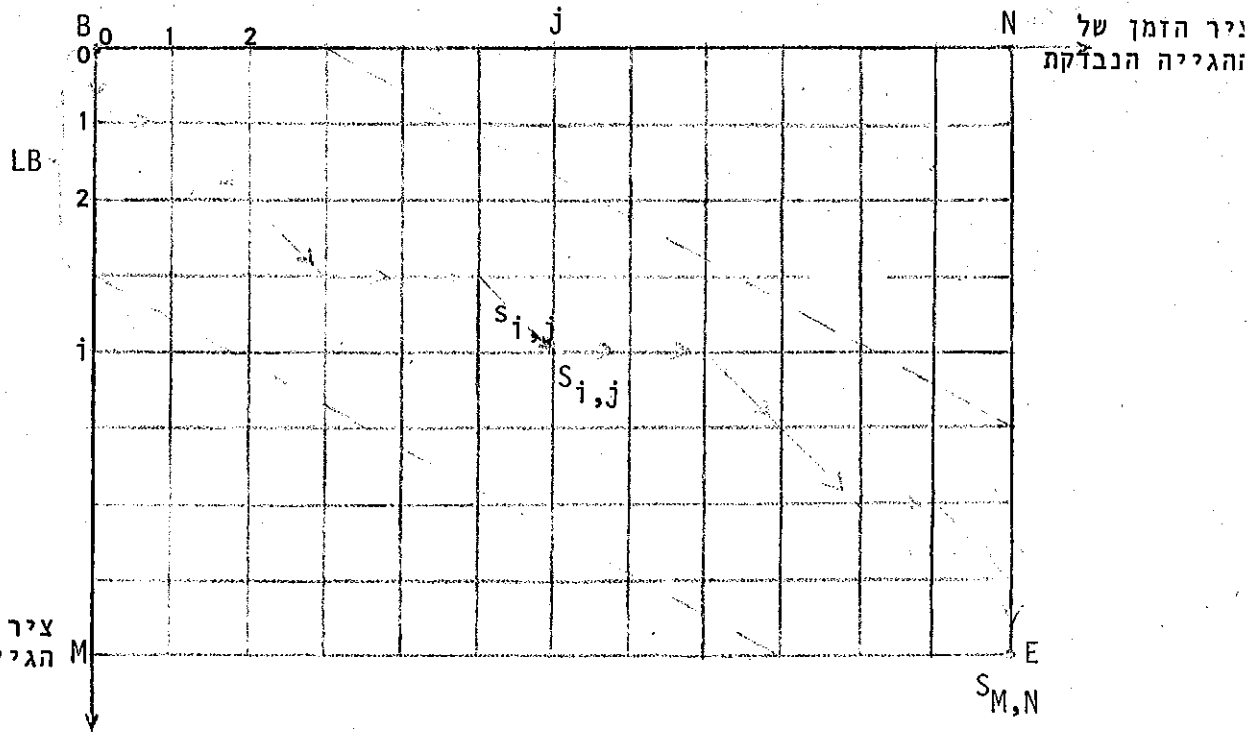
כאשר  $(S_{M,N}/L_{SH})$  בא להבטיח  $\hat{S} \leq 1$  (הדבר מובטח מאחר ו-  $S_{M,N} \leq L_p$  כי לפי בחירתנו  $s_{i,j} \leq 1$ ).

מבדיקות זיהוי הגיות, התברר שחלק מהטעויות שהתקבלו כאשר מידת הדמיון חושבה ללא הכפל ב-  $(L_p/L_{SH})$ , המופיע ב- (7.3), תוקן כאשר מידת הדמיון חושבה לפי (7.3).

לסיכום, נציין שלדעתנו עבור ביצוע התאמת צירי הזמן ומיון באמצעות תכנות דינמי, יש לשיטה של Velichko [10] יתרון על השיטות של Itakura [2] ושל Ichikawa [9]. יתרון זה הוא בכך, שבעוד שאצל Ichikawa קיים תנאי קצה הקובע שקטע הזמן הראשון בהגייה הנבדקת דומה לראשון בהגיית הייחוס ומידת הדמיון שביניהן חייבת להלקח בחשבון ואצל Itakura נוסף אותו תנאי גם עבור קטע הזמן האחרון של כל הגייה, אזי בשיטה של Velichko אין כל תנאים כאלה על קטעי הזמן שבקצוות והדבר מאפשר לתכנית הדינמית עצמה לקבוע מהם קטעי הזמן המתאימים עבור הקצוות. לכך יש חשיבות יתר, כאשר לוקחים בחשבון שבקביעת קצוות ההגייה בשלב עיבוד הקדם, יתכנו אי דיוקים העלולים להגביר את סיכוי הטעות בזיהוי.

פרט לכך, ישנו הבדל נוסף בין השיטה של Itakura [2] לבין השיטה של Velichko [10]. האחרונה אינה מאפשרת יותר מפעם אחת בחירה של קטע זמן כלשהו מההגייה הקצרה יותר, מבין השתיים המשוויות, לצורך חישוב מידת הדמיון. לכן עבורה  $L_p \leq L_{SH}$ , לעומת זאת, בשיטה הראשונה אפשר לעשות זאת פעמיים, ולכן אורך המסלול האופטימלי לפי שיטה זו גדול מ-  $L_{SH}$ , והוא מוגבל בהתאם להגדרתו רק עיני מספר קטעי הזמן של ההגייה הארוכה יותר.

מכיוון ששיטת Velichko פשוטה יותר משיטות אחרות, והיא גם בעלת יתרון עליהן כמתואר לעיל, בחרנו בה לפתוח עבור המערכת השוואתית.



ציור 7.1: מסלול לתאום צירי הזמן לצורך חישוב מידת הדמיון בין שתי הגינות.

Fig. 7.1: Time alignment path for computing the similarity measure between two utterances.

7.3 הפעלת תכנות דינמי במערכת ההשוואתית.

השיטה להתאמת צירי הזמן ומיון באמצעות תכנות דינמי, שנדונה בסעיף הקודם, הופעלה במערכת ההשוואתית בשלב המיון. היא מומשה בתכנית הזיהוי עבור המערכת ההשוואתית, המופיעה בנספח ב'. נביא כעת את פרטי המימוש.

תחילה נציין שמידת הדמיון  $s_{i,j}$ , בין שני קטעי זמן של שתי ההגיות המשוות, חושבה לפי פונקצית המרחק של צ'ביצ'ף, כפי שבוצע במערכת הכוללת דחיסה. לפיכך,

$$s_{i,j} = 1 - \frac{F}{2p} \sum_{l=1}^p |K_{li}^r - K_{lj}^q|; \quad 1 \leq i \leq M, \quad 1 \leq j \leq N. \quad (7.4)$$

כאשר  $K_{li}^r$  הוא מקדם ה- PARCOR ה-1 מתוך  $p = 6$  מקדמים המייצגים את קטע הזמן ה- $i$  של הגיית הייחוס ה- $r$ ,  $K_{lj}^q$  הוא מקדם ה- PARCOR ה-1 מתוך  $p = 6$  מקדמים המייצגים את קטע הזמן ה- $j$  של ההגייה הנבדקת  $q$ -ו,  $F$  הוא גורם נרמול מתאים.

מידת הדמיון בין שתי ההגיות ה- $x$  וה- $q$  חושבה לפי ביטוי (7.3) לאחר התהליך הרקורסיבי המוגדר ב- (7.2) עם תנאי ההתחלה שהוגדרו בסעיף 7.2. התהליך הרקורסיבי הזה מבוצע באמצעות השגרה SBWL והפונקציה SKL, ראה נספח ב'. תחום הקיום של המסלולים הוגבל, כפי שהוצע בסעיף 7.2, לאזור סביב האלכסון BE, התחום בצירור 7.1 בין שני קווים מרוסקים. מכיוון שבדרך כלל שתי ההגיות המשוות אינן שוות בארכן, אין אפשרות להגדיר את גבולות התחום, כלומר אלו קטעי זמן לא יבדקו בתהליך, באמצעות קוים ישירים כבצירור 7.1. באופן מעשי, שומרים בקרוב על השיפוע של הקוים הישרים  $(M-LB)/(N-LB)$  אך כל קו נשבר לאלכסונים העוברים דרך ריבועים, שכל אחד מהם מגדיר קטעי זמן גבוליים עבור תחום הקיום. השגרה שקובעת את תחום הקיום בדרך זו נקראת BAL והיא מופיעה בנספח ב'.

כעת נותר לקבוע את ערכם של שני משתנים שהוזכרו לעיל.  $F$  גורם הנרמול המופיע ב- (7.4), ו-LB הגורם הקובע את רחבו של תחום הקיום, ראה צירור 7.1. קביעת ערכם קשורה בהגיות הנבדקות, מכיוון שהיא נעשית על סמך תוצאות נסויי זיהוי הגיות. ערכיהם של  $F$  ו-LB משפיעים על תוצאות

הזיהוי. ערכו של  $F$  במיוחד, מכיוון שהוא משפיע ישירות על ערך  $s_{i,j}$  לפי (7.4). הדבר מתבטא בכך, שעבור  $F$  מתחת לערך מסוים (בניסויים התקבל ערך בסביבות 5) אורכי המסלולים האופטימלים שהתקבלו היו כמספר קטעי הזמן של ההגייה הקצרה יותר. כלומר נבחרו המסלולים האופטימלים הארוכים ביותר האפשריים. לעומת זאת, עבור  $F$  מעל לערך גבוה יותר (בניסויים התקבל ערך בסביבות 15) המסלולים האופטימלים שנבחרו היו קצרים ביותר. עבור שני הקצוות האלה תוצאות הזיהוי לא היו טובות לעומת ערכי הביניים. בניסויים שנערכו נמצא ש-  $F = 10$  נותן את התוצאות הטובות ביותר.

לערכו של  $LB$  השפעה קטנה יותר, אך גם הוא מתחת לערך מסוים, גורם לטעויות בזיהוי, מכיוון שהמסלולים, במקרים אלה, מוגבלים לתחום צר מדי. מתוך הניסויים שנערכו נמצא ש-  $LB = 12$  הוא ערך מינימלי הנותן את התוצאות הטובות ביותר.

בניגוד למערכת הכוללת דחיסה, כמות החישובים הנדרשת לצורך זיהוי הגייה במערכת ההשוואתית, תלויה בארכן של הגיות הייחוס ומשתנה כמובן עבור כל הגייה נבדקת, ולכן ניתן להעריכה רק במוצע. פרטים על כך ינתנו בפרק 9.

פרק 8: ניסויים בזיהוי הגיות ותוצאותיהם.

בפרקים הקודמים נדון פתוחן של שתי מערכות. האחת, הכוללת דחיסה של ההגיייה בזמן, והשניה שאינה כוללת דחיסה ופוחחה לשם השוואה עם הראשונה. ההשוואה בין שתי מערכות אלה תעשה משתי בחינות, האחת - הצלחה בזיהוי אותן ההגיות, והשניה כמות החישובים הנדרשת לצורך הזיהוי. בפרק זה נתאר את הניסויים בזיהוי ההגיות שנערכו בשתי המערכות ונביא את תוצאותיהם. בפרק הבא תערך ההשוואה לפי שתי הבחינות הנ"ל.

8.1 הכנת שני בסיסי נתונים

לצורך ביצועי הניסויים בזיהוי הגיות הוכנו שני בסיסי נתונים שונים, אך שניהם עבור אותו מילון שנבחר עבור המערכת - עשר הספרות העבריות "אפט" עד "תשע" (במין נקבה).

בסיס הנתונים הראשון כלל הגיות של שלושה דוברים שונים אשר הוקלטו בחדר שקט ללא רעש מחשב או מזגן, באמצעות Akai cassette recorder. כל דובר הקליט 4 קבוצות של עשר הספרות. ההקלטות נערכו בשני מועדים, הנפרדים זה מזה במספר ימים. בסיס הנתונים השני כלל את הגיותיהם של שני דוברים (מתוך השלושה הקודמים), אלה הוקלטו בחדר מחשב רועש (כלומר, בזמן ההקלטה הופעלה מערכת המחשב כולל יחידת הסרט המגנטי), גם כן באמצעות Akai cassette recorder\* הוחס אות לרעש (signal to noise ratio-S/N) בחדר זה, בזמן ההקלטות, היה בין 18 ל- 21 dB בתחום התדרים 200-3200 Hz שהוא התחום בו הוגבל אות הדיבור לפני הדגימה. גם בסביבה זו הקליט כל דובר 4 קבוצות של עשר הספרות בשני מועדים הנפרדים זה מזה במספר ימים.

בסיסי הנתונים הנ"ל הוכנו עבור יותר מדובר אחד, בכדי לבדוק האם מידת ההשפעה של הדחיסה על הצלחה בזיהוי, תלויה בדובר.

בסיס הנתונים השני הוכן לאחר שעבור הראשון התקבלו תוצאות זיהוי מבטיחות לגבי המערכת הכוללת דחיסה (השווה טבלאות 8.1 ו- 8.2 עבור חדר שקט), ומן הראוי היה לבדוק האם כך יהיה גם כאשר ההקלטות תעשינה בסביבה בעלת S/N נמוך.

---

\*כל ההקלטות בוצעו עם מיקרופון רב כיווני רגיל מתוצרת SONY.

עבור שני בסיסי הנתונים נערכו ההקלטות של 4 קבוצות הגיות, בשני מועדים, בכדי לבדוק מה השפעת הריחוק בזמן בין מועדי אמירתן של קבוצת ההגיות הנבדקות וקבוצת הגיות הייחוס, על תוצאות הזיהוי.

8.2 ארגון הניסויים והתוצאות.

הניסויים עבור כל בסיס נתונים בנפרד, אורגנו בדרך הבאה: מכל קבוצת הגיות שהוקלטו הוכנה בכל אחת מהמערכות ספריית וקטורי פרמטרים (מקדמי ה-PARCOR). עבור כל דובר בנפרד, שמשה בכל פעם ספריה אחרת שלו בספריית וקטורי ייחוס, כאשר השלוש האחרות משמשות כספריות של הגיות נבדקות. בדרך זו נערכו 120 ניסויים עבור כל דובר, כלומר סה"כ 360 ניסויים עבור בסיס הנתונים הראשון ו- 240 ניסויים עבור בסיס הנתונים השני. צורה זו של ארגון הניסויים הונהגה גם ע"י White and Neely [6].

בכל מערכת הונהג סדר הניסויים הנ"ל עבור ספריות וקטורי הפרמטרים שהוכנו בה. התוצאות שהתקבלו עבור המערכת הכוללת דחיסה מרוכזות בטבלה 8.1. והתוצאות עבור המערכת השוואתית ללא דחיסה מרוכזות בטבלה 8.2.

טבלה 8.1: תוצאות של ניסויי זיהוי במערכת הכוללת דחיסה.

Table 8.1: Recognition Test Results for Isolated Word Recognition System with TSC.

הסביבה Type of Environment	מספר הניסויים Number of Tests	מספר שגיאות לדובר Number of Errors per Speaker			מספר שגיאות כולל Total Number of Errors	הצלחה בזיהוי Recognition Score (%)
		IE	DM	DA		
שקטה	360	3	0	0	3	99.1
רועשת (S/N = 20 dB)	240	3	2	-	5	97.9

הערות לגבי התוצאות שבטבלה 8.1:

1. 3 השגיאות עבור הסביבה השקטה היו: "ארבע" זוהתה כ- "אחת", "אחת" זוהתה כ- "ארבע" ו- "שתיים" זוהתה כ- "שמונה". כל השגיאות חלו כאשר ספרית היחוס והספריה הנבדקת יצגו קבוצות של הגיות שהוקלטו במועדים נפרדים.

2. עבור הסביבה הרועשת, 3 השגיאות אצל הדובר IE היו: "חמש" זוהתה כ- "שתיים", "שש" זוהתה כ- "שבע" ו- "שש" זוהתה כ- "חמש".  
2 השגיאות אצל הדובר DM היו מאותו סוג: "שש" זוהתה כ- "חמש".  
גם כאן כמו עבור הסביבה השקטה, כל השגיאות חלו כאשר ספרית היחוס והספריה הנבדקת יצגו קבוצות של הגיות שהוקלטו במועדים נפרדים.

טבלה 8.2: תוצאות של ניסויי זיהוי במערכת ההשוואתית ללא דחיסה.

Table 8.2: Recognition Test Results for Isolated Word Recognition System with Dynamic Programming.

הסביבה Type of Environment	מספר הניסויים Number of Tests	מספר השגיאות לדובר Number of Errors per Speaker			מספר השגיאות הכולל Total Number of Errors	הצלחה בזיהוי Recognition Score(%)
		IE	DM	DA		
שקטה	360	6	0	0	6	98.3
רועשת (S/N = 20 dB)	240	5	0	-	5	97.9

הערות לגבי התוצאות שבטבלה 8.2:

1. 6 השגיאות עבור הסביבה השקטה היו: "אחת" זוהתה כ- "ארבע" (שגיאה זו זהה לשגיאה השנייה במערכת האחרת, בסביבה שקטה, והתקבלה עבור אותן הגיות), "שתיים" זוהתה כ- "אחת" (התקבלה עבור אותן הגיות כמו השגיאה השלישית במערכת האחרת, בסביבה השקטה), "שלוש", "ארבע", "שבע" ובפעם נוספת "שבע" זהו כ- "אחת". כל השגיאות פרט לאחרונה חלו כאשר ספריית היחוס והספריה הנבדקת יצגו קבוצות של הגיות שהוקלטו במועדים נפרדים.

2. עבור הסביבה הרועשת, 5 השגיאות היו: פעמיים "אחת" זוהתה כ- "תשע", פעמיים "אבע" זוהתה כ- "אחת", ו- "חמש" זוהתה כ- "אפס". כל השגיאות, פרט לאחת מהשגיאות בהן "ארבע" זוהתה כ- "אחת", חלו כאשר ספרית הייחוס והספריה הנבדקת יצגו קבוצות של הגיות שהוקלטו במועדים נפרדים.

פרט לניסויים הנ"ל, נערכו במערכת הכוללת דחיסה שתי סדרות ניסויים נוספות עבור בסיס הנתונים הראשון ("השקטי"), על מנת לבחון את השפעת ערכו של  $p$  (מספר המקדמים המייצגים קטע זמן) על הצלחת הזיהוי. בסדרה הראשונה השתתפו וקטורי פרמטרים שהכילו  $p = 12$  מקדמים לכל קטע זמן בניגוד ל-  $p = 6$  מקדמים בניסויים הנ"ל. בסדרה זו התקבלה שגיאה אחת בלבד מתוך 360 ניסויים, כלומר הצלחה בזיהוי של 99.7%. בסדרה השנייה השתתפו וקטורי פרמטרים שהכילו מקדם אחד בלבד לכל קטע זמן ( $p = 1$ ), והתקבלו 21 שגיאות מתוך 360 ניסויים, כלומר הצלחה בזיהוי של 94.1%. התוצאות של שתי הסדרות הנוספות האלה, בצרוף התוצאות שהתקבלו באותם תנאים עבור  $p = 6$ , מרוכזות בטבלה 8.3.

טבלה 8.3: תוצאות של ניסויי זיהוי עבור ערכים שונים של  $p$ . (הניסויים נערכו במערכת הכוללת דחיסה בזמן בסביבה השקטה).

Table 8.3: Recognition Test Results for Several Values of  $p$ . (The tests were performed for the IWRS with TSC at a Quiet Environment).

p	מספר הניסויים Number of Tests	מספר השגיאות לדובר Number of Errors per Speaker			מספר השגיאות הכולל Total Number of Errors	הצלחה בזיהוי Recognition Score (%)
		IE	DM	DA		
1	360	17	0	4	21	94.1
6	360	3	0	0	3	99.1
12	360	0	1	0	1	99.7



פרק 9: סכום ומסקנות

9.1 השוואה בין שתי המערכות מבחינת כמות החישובים

נסכם כעת את כמות החישובים במערכת הכוללת דחיסה בהתאם לשלבי הפעולה השונים.

חיפוש שיא האמפליטודה וקביעת סוף ההגיייה:	42,000	חיבורים	-
דחיסה:	6,000	"	3,000 כפלים
גזירה (19 קטעי זמן × 300 דגימות):	5,700	"	-
מיצוי מקדמי ה-PARCOR:	39,900	"	45,600 "
מיון ההגיייה:	9,180	"	-
סה"כ:	102,780	חיבורים	48,600 כפלים

עבור המערכת ההשוואתית אין אפשרות לקבל כמות חישובים קבועה, מכיוון שהיא תלויה במשכיהן של הגיות הייחוס וההגיייה הנבדקת. לפיכך, נוכל רק להעריך כמות חישובים ממוצעת, שתקבע בהתאם למשך ההגיות שהשתתפו בניסויים שבוצעו.

מספרי קטעי הזמן בהגיות שהשתתפו בניסויים נע בין 33 ל-62, כאשר הממוצע על פני כל ההגיות הוא 47.3 קטעי זמן. מספר זה יש להשוות עם 19 קטעי זמן שהתקבלו עבור כל אחת מההגיות הדחוסות. פרושו של דבר שכמות החישובים לצורך גזירה ומיצוי מקדמי ה-PARCOR גדולה בממוצע במערכת ללא דחיסה, פי  $2.49 = 47.3/19$  מאשר כמות החישובים הדרושה לביצוע כל הפעולות האלה במערכת הכוללת דחיסה.

לצורך כמות החישובים בשלב המיון, נחשב קודם את  $\bar{N}_c$  - מספר מידות הדמיון החלקיות המתאים לתחום הקיום של המסלולים, כאשר משווים בית שתי הגיות "ממוצעות", כלומר בעלות 47 קטעי זמן כל אחת. כאשר משווים הגיות השוות בארכן  $N_c = (N-LB)(N-LB-1)$ , כאשר N הוא מספר קטעי הזמן של כל אחת מהן ו-LB כמוגדר בציור 7.1. נציב עבור LB את הערך שהשתמשנו בו בכל הניסויים,  $LB = 12$  ועבור N את מספר קטעי הזמן של ההגיייה "הממוצעת" ונקבל  $\bar{N}_c = 1,190$ . כמות החישובים לקבלת מידת הדמיון בין שני קטעי זמן היא כפל אחד ו-18 חיבורים. לכן כמות החישובים הנדרשת לחישוב מידת הדמיון בין שתי הגיות היא:  $\bar{N}_c$  כפלים ועוד  $21\bar{N}_c$  חיבורים. ( $3\bar{N}_c$  דרושים לצורך חישוב הביטוי (7.2)). לקבלת כמות החישובים הנדרשת בשלב המיון יש לכפול ביטויים אלה ב-10.

נסכם כעת, בהתאם לתוצאות הנ"ל את כמות החישובים הממוצעת במערכת ההשוואתית ללא דחיסה.

חיפוש שיא האמפליטודה וקביעת סוף ההגיה:	42,000	חיבורים	-
גזירה:	14,200	"	-
מיצוי מקדמי ה-PARCOR:	99,300	"	113,500 כפלים
מיון ההגיה:	249,900	"	11,900
ס ה " כ:	405,400	חיבורים	125,400 כפלים

לפיכך, לצורך זיהוי הגיה אחת דרושים בממוצע במערכת ההשוואתית ללא דחיסה פי 3.9 חיבורים ופי 2.5 כפלים מאשר דרושים במערכת הכוללת דחיסה. אם נמיר לצורך ההשוואה כל כפל ב- 3 חיבורים (כאשר החישובים הם בנקודה צפה), נקבל שבמערכת הכוללת דחיסה כמות החישובים קטנה לפחות פי 3.

## 9.2 השוואה בין שתי המערכות מבחינת תוצאות הזיהוי

כאשר משווים את תוצאות הניסויים שנערכו בזיהוי הגיות, הרשומות בטבלאות 8.1 ו- 8.2, מתקבל שביצועי המערכת הכוללת דחיסה אינם נופלים מביצועי המערכת ההשוואתית. עבור סביבה שקטה, התוצאות שהתקבלו במערכת הכוללת דחיסה אף היו טובות יותר.

בהקשר לפילוג השגיאות בין הדוברים, מתברר שבשתי המערכות, עבור הסביבה השקטה, הצטברו השגיאות אצל דובר אחד - IE. עבור הסביבה הרועשת, נשתמרה תופעה זו רק במערכת ההשוואתית, בעוד שבמערכת הכוללת דחיסה מתקבלות שגיאות גם אצל דובר אחר.

יש לשים לב לעובדה, שלא ניתן על סמך ניסויים אלה להשוות בין המערכות מבחינת חסינותן לרעש, מכיוון שעבור שתי הסביבות השונות נבדקו שני בסיסי נתונים שונים ולא בוצע ניסוי מבוקר שבו מוסיפים רעש סינתטי להגיות שהוקלטו בסביבה שקטה. אולם, ניתן בכל זאת להסיק מהתוצאות שגם בסביבה רועשת אין המערכת הכוללת דחיסה גופלת מהמערכת ההשוואתית.

נתבונן כעת בהערות שנרשמו לגבי התוצאות. עבור הסביבה השקטה, מסתבר שיש קשר בעיתוי הופעת השגיאות בשתי המערכות. השגיאות בהן "ארבע" ו"אחת" מוחלפות בניהן התגלו כטיפוסיות למילון, עקב הופעתן השכיחה גם בניסויים

אחרים שלא דווח עליהם כאן. עבור הסביבה הרועשת, ניתן לראות שבמערכת הכוללת דחיסה, כמעט כל השגיאות התרכזו סביב הגיות הכוללות קטעים ניכרים של דיבור לא קולי - "חמשי" ו- "ששי". תופעה זו לא התגלתה בשגיאות של המערכת ההשוואתית (פרט למקרה "חמשי" - "אפסי"). יתכן ושתופעה זו מצביעה על רגישות מסוימת של הראשונה לרמת רעש גבוהה.

מסקנה אחרונה, אך חשובה, היא שהמערכת הכוללת דחיסה מגלה יתר יציבות מבחינת אימונה לדובר מאשר המערכת ההשוואתית. הדבר מתבטא בכך שבמערכת הכוללת דחיסה לא חלה כל שגיאה כאשר הגיות היחוס וההגיות הנבדקות הוקלטו באותו מועד. לעומת זאת, במערכת ההשוואתית, בכל סביבה, שגיאה אחת נוצרה גם כאשר הגיות היחוס וההגיות הנבדקות הוקלטו באותו מועד.

יש לציין, כפי שכבר נאמר בסעיף 7.3, שגישת התכנות הדינמי גילתה רגישות ניכרת לשינויים בערכו של קבוע הנרמול F (ראה (7.4)), ורגישות חלשה יותר לשינויים בערכו של LB (ראה ציור 7.1). לכן, קביעת ערכם המתאים לקבלת הצלחה מכסימלית בזיהוי, דרשה מספר ניכר של בדיקות. לעומת זאת, המערכת הכוללת דחיסה בזמן היתה חופשית מבעיות אלה ותהליך המיון שלה היה פשוט יותר, ודבר זה מקנה לה יתרון מה על פני המערכת ללא דחיסה המשתמשת בתכנות דינמי.

### 9.3 מסקנות

מערכת לזיהוי מילים בודדות פותחה ומומשה בתוכנה באמצעות מיני-מחשב NOVA 2. מהמערכת נדרש זיהוי נכון בזמן קצר ככל האפשר. על הדרישה האחרונה עונה המערכת בעיקר ע"י ביצוע דחיסה בזמן של ההגיה בשלב עיבוד הקדם באמצעות אלגוריתם יעיל. עקב דחיסה זו קטנה כמות החישובים לצורך הזיהוי במערכת הכוללת דחיסה לפחות פי 3 מאשר כמות החישובים במערכת ללא דחיסה שפותחה ומומשה לצורך השוואה (ראה סעיף 9.1)\*. פרט לכך, היצוג הפרמטרי ותהליך המיון נבחרו על סמך כושרם לבצע את הנדרש בזמן מינימלי. תוצאות הניסויים בזיהוי הגיות שנערכו במערכת הכוללת דחיסה בזמן ובמערכת שאין בה דחיסה, מעידים על קיום הדרישה הראשונה - זיהוי נכון. הצלחה בזיהוי בשעור שלמעלה מ- 97% גם בסביבה רועשת ( $S/N = 20$  dB), מראה שביצועי המערכת הכוללת דחיסה נמצאים ברמה משביעת רצון יחסית לביצועים המקובלים כיום עבור מערכות מסוגה. לכך נוסיף שמתוצאות ניסויי הזיהוי המרוכזות בטבלאות

\* במערכת הכוללת דחיסה שמומשה, זמן זיהוי הגייה הוא כ- 10 שניות.

8.1 ו- 8.2 נובע שהדחיסה לא פגעה בהצלחת הזיהוי, ועבור הגיות שנאמרו בסביבה שקטה היא אף שיפה אותה.

אמנם, מספר הניסויים שבוצעו, גודל המילון ומספר הדוברים שהשתתפו בניסויים הני"ל, אינם מספיקים לצורך הסקת מסקנות סופיות ביחס לישומים של האלגוריתם לדחיסה בזמן המוצע, עבור כל מערכת לזיהוי מילים בודדות הממומשת בתכנה בלבד. אך התוצאות שהתקבלו תומכות בגישה המוצעת ומצביעות על כוחה.

הסיבה שביצועי המערכת הפועלת עם דחיסה אינם נופלים מהמערכת שלא דחיסה, היא כנראה הדרך שבה מבצע האלגוריתם המוצע את הדחיסה בזמן - שקלול מתאים של כל דגימות הכניסה, או רובן, ללא השמטת קטעים בצורה שרירותית. יתכן והדחיסה מסלקת, במידה מסוימת, יתרות הקיימת בהגייה הנבדקת, כך שהפרמטרים המחושבים מההגייה הדחוסה, מאפיינים אותה ביתר מהימנות מאשר הפרמטרים המחושבים מההגייה הטבעית. זו אולי הסיבה ליתרון הקל שהתגלה עבור המערכת הכוללת דחיסה בסביבה השקטה.

לצד התוצאות החיוביות שהתקבלו עבור הדחיסה בזמן, יש לשים לב גם למשמעותן עבור היצוג הפרמטרי - מקדמי ה- PARCOR. ההצלחה הגבוהה שהתקבלה בזיהוי - מעל 97%, עבור יצוג של קטע זמן באמצעות  $p = 6$  מקדמי ה- PARCOR, מצביעה כל כושרם של מקדמים אלה ליצג בצורה מהימנה את ההגיות כך שבתהליך המיון ניתן להפריד ביניהן באחוז הצלחה גבוה כני"ל. עדות נוספת על כך ניתנת בטבלה 8.3, המראה שעבור הגיות שנאמרו בסביבה שקטה,  $p = 12$  מעלה את % ההצלחה ל- 99.7%, ומצד שני עבור  $p = 1$  עדיין מתקבלת תוצאת זיהוי טובה - 94.1%.

בסעיף הקודם צויין שעבור המערכת הכוללת דחיסה, כאשר הגיות היחוס וההגיות הנבדקות הוקלטו באותו מועד לא התקבלו שגיאות זיהוי. תוצאה זו תומכת בצורת הפעלה של המערכת, לפיה מאמנים את המערכת מחדש בכל יום בו מעונינים להשתמש בה, כי סביר שבדרך זו לא תתקבלנה שגיאות בזיהוי. זמן אימון המערכת הוא קצר מכיוון שהוא אינו דורש יותר מאשר אמירה פעם אחת של עשר מילות המילון, והשימוש בדחיסה מבטיח מיצוי פרמטרים מהיר. לכן אימון המערכת מחדש בכל יום, לא יהווה מעמסה על המשתמש.

לסיום, נביא הצעה להקטנת זמן הזיהוי של המערכת הכוללת דחיסה שפותחה.

כאמור בטעיף 6.1 נדגמה ההגייה ב- 10 KHz לאחר שהוגבלה לתחום 200-3200 Hz. הגבלה זו מאפשרת להוריד את קצב הדגימה ב-  $1/3$ , כלומר לדגום בקצב של 6.67 KHz מבלי לפגוע בתנאי משפט הדגימה. הפחתה זו של קצב הדגימה, תגרום להורדת מספר הדגימות ליצוג הגייה בעלת אורך נתון ב-  $1/3$ , וזה יגרור הקטנת זמן הזיהוי. אולם יש לבדוק, אם אין פעולה זו פוגעת בהצלחת הזיהוי.

נספח א': תכנית הזיהוי עבור המערכת הכוללת דחיסה.

1.א. התכנית הראשית ICWRS

הקלט:

1. LEV - סף התחלת ההגייה - לאחר שנספרו 5 דגימות שעוברות ערך זה, נקבע שההגייה החלה ודגימותיה מועברות למערך ISA.
2. LEVT - סף סיום ההגייה - ערך מינימלי עבור 5 הדגימות שהתכנית סופרת לצורך גילוי סוף ההגייה.
3. REF INP FILE - קובץ המכיל באופן עוקב עשרה וקטורי יחוס.
4. IO - כאשר מקנים לו את הערך 1, ההגייה הנבדקת (דגומה) תקרא מה-DISK. כאשר 2, ההגייה הנבדקת תמסר בצורה אנלוגית ותדגם.
5. K OUT FILE - שם הקובץ שבו יכתבו הפרמטרים של ההגייה הנבדקת.
6. TEST INPUT - הקובץ שבו הדגימות של ההגייה הנבדקת. נדרש רק כאשר מקנים ל-IO את הערך 1.
7. האות האנלוגי של ההגייה הנבדקת, הנמסר מסרט הקלטה אנלוגי או ממיקרופון, כאשר IO = 2.

הפלט:

1. הודעה על ערכה של דגימת השיא בהגייה (=PEAK), וערך הסף שחושב לצורך גילוי סוף ההגייה (=LEV).
2. הודעה על המילה שוקטור הייחוס המתאים לה נמצא כדומה ביותר לוקטור הנבדקת, ועל המילה הנמצאת במקום השני מבחינת הדמיון. כן נמסר על מספר ההזזה (SHIFT) שבה נמצא הדמיון המקסימלי.
3. מידות הדמיון בין כל אחד מוקטורי הייחוס לבין הוקטור של הנבדקת.
4. וקטור הפרמטרים של הנבדקת.

המשתנים העיקריים:

- ISA(10000) - מכיל את דגימות ההגייה.
- REF (1020) - מכיל את וקטורי הייחוס בסדר עוקב מ-1 עד 0.
- CK (12) - מכיל את הפרמטרים המחושבים מקטע הזמן הנדון של הנבדקת.

- TEST(120) - מכיל את וקטור הנבדקת.
- SWK(3,10) - מכיל את מידות הדמיון עבור עשרת וקטורי היחוס שהתקבלו עקב שלוש ההזזות.
- H(320) - מכיל את דגימות החלון המשולש המחושבות לצורך הדחיסה.
- MPS - ערכו הסופי הוא שיא ההגייה.
- IWS - ערכו הסופי הוא מספר הדגימות שמכילה ההגייה לאחר גילוי סיומה.
- Q - גורם הדחיסה C (ראה פרק 4).
- LC -  $L_{c_0} + \Delta L_c$  (ראה סעיף 4.2).
- LF - m (ראה פרק 4).
- NP, NSC -  $N_p, N_{sc}$  בהתאמה (ראה פרק 4).

השגרות הנקראות:

- A2DM1 - שגרה לדגימת ההגייה (ראה א.2).
- LPF - שגרה לחישוב מקדמי ה- PARCOR עבור קטע זמן (ראה א.3).
- INOUT - שגרת קלט/פלט, הפותחת קובץ ומקצה לו מספר ערוץ. בנויה סביב השגרה של FORTRAN :FOPEN.

הערות:

1. הגבלת IWS הסופי לתחום:  $5110 \leq IWS \leq 9500$ , נעשתה על סמך כמות הזכרון המוקצבת עבור תכנית הזיהוי. אורכי ההגיות שנבדקו בניסויי הזיהוי לא יצאו מתחום זה.
2. וקטור הפרמטרים של הנבדקת (TEST) מכיל 114 מקדמים, אולם רק 102 מקדמים המתאימים לקטעי הזמן 2÷18 (מתוך 19), נכתבים בקובץ K.OUT FILE.
3. התכנית בודקת את ערכה המוחלט של כל דגימת הגייה (NT) ואם  $NT \geq 2047$  התכנית מבטלת את בדיקת הזיהוי עקב גלישת אמפליטודה, ומוסרת על כך הודעה מתאימה. לאחר הדפסת הודעה זו, התכנית חוזרת למקום בו היא ממתינה להופעת הגייה חדשה, כאשר היא שומרת על המידע שנמסר לה לפני הביטול, ואין היא דורשת את עידכוננו.

4. לאחר הדפסת תוצאות הבדיקה, שואלת התכנית אם ממשיכים לבדיקה הבאה (CONTINUE?...?). אם המשתמש מדפיס 1(YES), היא מחזירה אותו למקום בו היא מדפיסה את הכותרת RECOGNITION TEST ומצפה לשם קובל עבור K.OUT FILE. כל המידע שנמסר לה בבדיקה הקודמת לפני המקום הזה, נשמר ואין היא דורשת את עדכונו.

בעמודים הבאים מובאת הדפסה של התכנית הראשית - ICWRS. דוגמת הרצה שלה תנתן ב- 4.א.



תכנית הראשית עבור המערכת הכוללת דחיקה - ICWRS

```
DIMENSION ISA(10000),IX(1040)
DIMENSION WINT(299),H(320),TEST(120),REF(1020)
DOUBLE PRECISION Y(600),X(300),CK(12),SWK(3,10),TPI,DPDF,MAX
DOUBLE PRECISION MIN,SUMRT2
EQUIVALENCE (ISA(9501),H(1))
EQUIVALENCE (ISA(3101),Y(1)),(ISA(5501),X(1)),(ISA(7301),CK(1))
EQUIVALENCE (ISA(9401),TEST(1)),(ISA(9701),SWK(1,1))
```

```
*** ACCEPT LEVELS FOR UTTERANCE END POINTS DETECTION ***
```

```
ACCEPT "HEAD LEVEL :",LEV
ACCEPT "TAIL LEVEL :",LEVT
LC=3010
LF=2
```

```
*** READ REFERENCE FILE ***
```

```
CALL INOUT(2,"REF INP FILE :",0)
ACCEPT "IN. TEST INPUT (1) , EX. TEST INPUT (2) :",10
READ BINARY (2) (REF(1),I=1,1020)
```

```
*** COMPUTE HAMMING WINDOW ***
```

```
TPI=8.0D0*DATAN(1.0D0)
DO 6 J=1,299
WINT(J)=.54-.46*DCOS(TPI*(J-1)/299.)
750 TYPE "***** R E C O G N I T I O N T E S T *****"
CALL INOUT(1,"K OUT FILE :",1)
751 CONTINUE
IF(10.EQ.2) GOTO 7
```

```
*** READ INTERNAL TEST INPUT UTTERANCE ***
```

```
CALL INOUT(0,"TEST INPUT :",0)
READ BINARY (0) (ISA(I),I=1,9945)
```

(המשך) ICWRS

```
CALL FCLOS(0)
GOTO 8
CONTINUE
?
C
C *** READY FOR TEST INPUT UTTERANCE FROM MICROPHONE ***
C
C PAUSE
C TYPE "<7>"
C
C **** THE STARTING OF THE RECOGNITION PROCESS ****
C
C *** CALL FOR SAMPLING ***
C
C CALL A2DMI(ISA,LEV)
C
C *** DETECT THE PEAK AND THE FINAL END POINT ***
C
8
MPS=ISA(1)
IF(MPS.LT.0) MPS=-MPS
DO 600 I=2,9945
NT=ISA(I)
IF(NT.LT.0) NT=-NT
IF(NT.GE.2047) GOTO 700
IF(NT.LE.MPS) GOTO 600
MPS=NT
600 CONTINUE
RATIO=FLOAT(MPS)/FLOAT(2048)
VP=1.25*FLOAT(LEV)
LEV=INT(VP*RATIO+0.5)
IF(LEV.LT.LEVT) LEV=LEVT
TYPE "PEAK= ",MPS," LEV= ",LEV
NAL=5
NP=0
IUS=9945
```

(המשך) ICWRS

```
4 NT=ISA(IWS)
  IF(NT.LT.0) NT=-NT
  IWS=IWS-1
  IF(NT.LE.LEV) GOTO 4
  NP=NP+1
  IF(NP.LT.NAL) GOTO 4
  IF(IWS.LT.5110) IWS=5110
  IF(IWS.GT.9500) IWS=9500
```

```
C
C ***** TIME SCALE COMPRESSION PHASE *****
C
```

```
LOGICAL LAST
LAST=.FALSE.
```

```
C
C *** ACCEPT PARAMETERS, COMPUTE COMPRESSION FACTOR AND PARAMETERS ***
C
```

```
NP=200
Q=FLOAT(IWS-(LF-1)*NP)/FLOAT(LC)
NS=200
IWC=0
TPI=8.*ATAN(1.)
PI=TPI/2.
SCF=FLOAT(NP)/(Q-1.)
NSC=INT(SCF+0.5)
MH=LF*NSC
MHH=MH/2
M=LF*NP
MP=M-NP
MA=MP+NSC
MT=M/2
NSX=NSC+NP
NSCP=NSC+MP-MT
```

```
C
C *** READ INITIAL CONDITIONS ((LF-1)*NP) AND NSC SAMPLES ***
C
```

(המשך) ICWRS

```
DO 12 I=1,MA
12 IX(I)=ISA(I)
C
C *** COMPUTE TRIANGULAR DATA WINDOW ***
C
CSN=1./FLOAT(NSC)
H(1)=CSN
DO 24 I=2,NSC
24 H(I)=H(I-1)+CSN
KT=MA
DO 100 J=1,NS

C
C *** COMPUTE NSC OUTPUT SAMPLES ***
C
36 DO 50 K=1,NSC
A=H(K)*(IX(K+NP)-IX(K))+IX(K)
IWC=IWC+1
ISA(IWC)=INT(A)
IF(IWC.EQ.3000) GOTO 300
50 CONTINUE
IF(LAST) GOTO 300

C
C *** READ NEW NSC+NP INPUT SAMPLES ***
C
DO 55 I=1,NSX
IM=I+MA
KT=KT+1
IF(KT.GT.IWS) GOTO 200
55 IX(IM)=ISA(KT)

C
C *** SHIFT INDEX ***
C
DO 60 I=1,MA
60 IX(I)=IX(I+NSX)
100 CONTINUE
```

(המשך) ICWRS

```
200      DO 250 J=1,NSX
          JM=J+MA
250      IX(JM)=0
          DO 260 I=1,MA
260      IX(I)=IX(I+NSX)
          LAST=.TRUE.
          GOTO 36
300      CONTINUE
C
C ****      PARAMETERS EXTRACTION PHASE      ****
C
          IP=0
          IPK=0
C
C ***      READ THE FIRST 300 SAMPLES      ***
C
          DO 62 I=1,300
          IP=IP+1
62      Y(I)=DFLOAT(ISA(IP))
          DO 99 K=1,10
          IF(K.EQ.10) GOTO 53
C
C ***      READ NEW 300 SAMPLES      ***
C
          DO 52 I=301,600
          IP=IP+1
52      Y(I)=DFLOAT(ISA(IP))
53      DO 110 I=1,2
          L1=(I-1)*150+1
          L2=(I-1)*150+300
          DO 310 J=L1,L2
          M=J-150*(I-1)
310      X(M)=Y(J)
C
C ***      DIFFERENCE AND WINDOWING      ***
```

(המשך) ICWRS

```
C
210 DO 210 J=1,299
X(J)=(X(J+1)-X(J))*WINT(J)
X(300)=-X(300)*.08
SUMRT2=0.0D0

C
C *** CALL FOR THE EXTRACTION OF THE PARCOR COEFFICIENTS ***
C
CALL LPF(X,CK,300,6)
DO 75 J=1,6
IPK=IPK+1
75 TEST(IPK)=SINGL(CK(J))
IF(K.EQ.10) GOTO 99
110 CONTINUE
DO 98 I=301,600
J=I-300
98 Y(J)=Y(I)
99 CONTINUE

C
C **** CLASSIFICATION PHASE ****
C
DO 520 KW=1,10
IR=(KW-1)*102
DO 490 JT=1,3
MIN=0.0D0
LJ=(JT-1)*6
II=LJ
IJ=IR
DO 495 J=1,17

C
C *** COMPUTE CHEBISHEV NORM ***
C
SUMRT2=0.0D0
DO 500 I=1,6
IJ=IJ+1
```

(7777) ICWRS

```
      II=II+1
      DPDF=DBLE(TEST(II))-DBLE(REF(IJ))
      IF(DPDF.LT.0) DPDF=-DPDF
      SUMRT2=SUMRT2+DPDF
500   CONTINUE
      MIN=MIN+SUMRT2
495   CONTINUE
C
C *** STORE THE SIMILARITY MEASURE ***
C
      SWK(JT,KW)=1.0D0-MIN/DFLOAT(204)
490   CONTINUE
520   CONTINUE
C
C *** DECIDE ABOUT THE WORD AND THE SECOND NEAREST ***
C
      MAX=SWK(1,1)
      NW=1
      ND=1
      DO 400 I=1,10
      DO 450 J=1,3
      IF(SWK(J,I).LE.MAX) GOTO 450
      MAX=SWK(J,I)
      NW=I
      ND=J
450   CONTINUE
400   CONTINUE
      TPI=0.0D0
      DO 301 I=1,10
      IF(I.EQ.NW) GOTO 301
      IF(SWK(ND,I).LE.TPI) GOTO 301
      TPI=SWK(ND,I)
      NWS=I
301   CONTINUE
      TYPE "<7>"
```

(המשך) ICWRS

\*\*\*\* THE END OF THE RECOGNITION PROCESS \*\*\*\*

\*\*\* TYPE RECOGNITION RESULTS \*\*\*

IF(NW.EQ.10) NW=NW-10

IF(NWS.EQ.10) NWS=NWS-10

WRITE (10,95)

95 FORMAT(2X,"THE SIMILARITY BETWEEN TEST WORD AND REF.")

DO 96 I=1,10

WRITE (10,94) I,SWK(1,I),SWK(2,I),SWK(3,I)

96 CONTINUE

94 FORMAT(1X,"REF. WORD #",I3,2X,3D16.8)

WRITE (10,97) NW,MAX,ND

97 FORMAT(2X,"THE WORD :",2X,I3,2X,"SIMIL. :",D16.8," SHIFT#",I4)

WRITE (10,101) NWS,TPI

101 FORMAT (1X,"THE SECOND:",2X,I3,2X,"SIMIL. :",D16.8)

WRITE BINARY (1) (TEST(I),I=7,108)

CALL FCLOS(1)

ACCEPT "CONTINUE ?,YES(1),NO(0) ",JCO

IF(JCO.EQ.1) GOTO 750

GOTO 1

700 TYPE "<7>"

TYPE "<7>"

TYPE "AMPLITUDE OVERFLOW -- TEST CANCELLED"

GOTO 751

CONTINUE

STOP

END



2.א. השגרה לדגימת ההגיייה A2DM1  
(כתובה בשפת אסמבלי של NOVA)

הפרמטרים:

ISA - הכתובה הראשונה של המערך בזכרון בו יאוחסנו דגימות הגיייה.

LEV - ערך הסף לצורך גילוי תחילת ההגיייה (ראה א.1).

הסבר:

A2DM1 בודקת הופעת רמת מתח (מעל 1.25 V) בערוץ מס' 2. (קטע התכנית L1 עד L2). עם גילוי הרמה הנייל סופרת השגרה 1512 (3000 g) דגימות, בכדי לדלג על רעש הנקלט עקב הפעלת מתג המיקרופון, (הקטע L3-L2). לאחר זאת מעבירה השגרה את ערכו של הפרמטר LEV אל אקומולטור 1 וסופרת את הדגימות העוברות ערך זה (הקטע LOOP-L3). לאחר שנספרו 5 דגימות מעבירה השגרה את כתובת הזכרון שממנה והלאה ימצא המערך ISA, אל אקומולטור 1 ועוברת לדגימה רצופה של שלושה בלוקים כאשר כל בלוק מכיל 3315 דגימות. הדגימות האלה מועברות אל הזכרון למערך ISA. (הקטע OUT-LOOP). לאחר זאת השגרה מסתיימת וחוזרים לתכנית הראשית.

הדפסה של השגרה מובאת בהמשך.

A2DM1 השגרה לדגימת ההגיייה

```
.TITL A2DM1
.ENT  A2DM1
.EXTU
.ZREL
A2DM1: START
.NREL
START: JMP @.SAVE
      24
      JSR @.MKST
      64
L1:   LDA @,CHM
      LDA 1,LEVEL
      DOAS @,ADCV
      SKPBZ ADCV
      JMP  .-1
      DIC @,ADCV
      MOVZL# @,@,SZC
      NEG @,@
      AND# @,1,SNR
      JMP L1
      LDA @,CH0
L2:   DOAS @,ADCV
      SKPBZ ADCV
      JMP  .-1
      DSZ J
      JMP L2
L3:   LDA @,CH0
      LDA 1,@-4,3 ;TRANSFER THE VALUE OF LEV
      DOAS @,ADCV
      SKPBZ ADCV
      JMP  .-1
      DIC @,ADCV
      MOVZL# @,@,SZC
      NEG @,@
      SUBZ# @,1,SZC
```

(המשך) A2DM1

```
JMP L3
DSZ N
JMP L3
LOOP: LDA I,-3,3           ;TRANSFER THE LOCATION OF ISA
      DOB 1,ADCV
      LDA 0,CHAN
      DOA 0,ADCV
      LDA 0,WRD
      LDA 1,MASK
      AND 0,I
      DOCP 1,ADCV
      SKPBZ ADCV
      JMP .-1
      DSZ I
      JMP NEXT
      JMP OUT
NEXT:  DIB 1,ADCV
      JMP LOOP
OUT:   LDA 0,I0
      STA 0,I
      LDA 0,J0
      STA 0,J
      LDA 0,N0
      STA 0,N
      JMP @.RETN
      JMP @.RETN
```

WRD:171415

I:3

MASK:7777

CHAN:400

LEVEL:17000

CHM:2

CH0:0

J:3000

N:5

LEV:60

I0:3

J0:3000

N0:5

.END A2DM1

3.א השגרה לחישוב מקדמי ה- LPF PARCOR

(נכתבה על סמך השגרה של [20] Makel and Gray)

הפרמטרים:

X - מערך כניסה המכיל את הדגימות של קטע הזמן 'עבורו' יחושבו מקדמי ה- PARCOR.

CK - מערך יציאה שיכיל את המקדמים שיחושבו.

NP - מספר הדגימות של קטע הזמן הנייל.

M - מספר המקדמים שיחושבו.

הסבר:

לאחר חישוב  $R(i)$ ,  $i = 1, \dots, M+1$ , מקדמי אוטוקורלציה  $R(1)$  מתאים למקדם אוטוקורלציה עבור השהייה אפט), מחושבים בתהליך רקורסיבי  $CK(j)$ ,  $j=1, \dots, M$ , מקדמי ה- PARCOR. יש לציין שהמערך A מכיל בכל צעד של התהליך את ה- LPC המתאימים לצעד. אמנם, חוזרים מהשגרה לפני שמחושבים ה- LPC של הצעד ה- M, אך השמטת המשפט: GOTO 10 IF(IP1.EQ.M) תאפשר את חישובם של ה- LPC עבור מסנן מסדר M.

הדפסת השגרה בעמוד הבא.

השגרה לחישוב מקדמי ה- PARCOR לרפ

```

SUBROUTINE LRF(X,CK,MP,M)
DOUBLE PRECISION X(303),A(12),CK(12),R(13),SS,AL,BE,TA
C C C *** COMPUTE M+1 AUTOCORRELATION COEFFICIENTS ***
MP1=MP+1
DO 11 JJ=1,MP1
J=JJ-1
NMJ=MP-J
SS=0.0D0
DO 9 I=1,NMJ
IPJ=I+J
SS=SS+X(I)*X(IPJ)
R(JJ)=SS
11 CONTINUE
C C C *** COMPUTE M PARCOR COEFFICIENTS ***
A(1)=-R(2)/R(1)
CK(1)=A(1)
AL=R(1)+A(1)*R(2)
BE=R(3)+A(1)*R(2)
MM1=M-1
DO 10 I=1,MM1
IP1=I+1
CK(IP1)=-BE/AL
IF(IP1.EQ.M) GOTO 10
I2=IP1/2
DO 20 J=1,I2
IJ=IP1-J
TA=A(J)+CK(IP1)*A(IJ)
A(IJ)=A(IJ)+CK(IP1)*A(J)
20 A(J)=TA
A(IP1)=CK(IP1)
AL=AL+CK(IP1)*BE
BE=R(I+3)
DO 30 J=1,IP1
NJ=I-J+3
30 BE=BE+A(J)*R(NJ)
10 CONTINUE
RETURN
END
```

4.א דוגמת הרצה של ICWRS

בדוגמא זו מובאת בדיקה זיהוי ההגייה שדגימותיה בקובץ DM71. זו ההגייה "אחת" שנאמרה ע"י הדובר DM בסביבה הרועשת, ובהתאם לכך ניתנו ערכי LEV ו- LEVT. הקובץ LX6DM5 מכיל באופן עוקב את וקטורי היחוס של עשר הגיות (אחת, שתיים, ..... , תשע, אפס) שנאמרו ע"י DM באותה סביבה אך מספר ימים קודם לכן.

ההודעה הראשונה המודפסת היא PEAK = LEV = שהם בהתאמה ערך הסף שחושב לצורך גילוי סוף ההגייה וערך דגימת השיא של ההגייה. לאחר שתהליך הזיהוי מסתיים מדפיסה התכנית את מידות הדמיון בין וקטור הנבדקת לבין כל אחד מוקטורי היחוס, ואת החלטה.

ICWRS

HEAD LEVEL : 80  
TAIL LEVEL : 72

REF INP FILE :  
LX6DM5

IN. TEST INPUT (1) , EX. TEST INPUT (2) : 1

\*\*\*\* RECOGNITION TEST \*\*\*\*

K OUT FILE :  
KTMP

TEST INPUT :  
DM71

PEAK= 1836 LEV= 81

THE SIMILARITY BETWEEN TEST WORD AND REF.

REF. WORD # 1	.92354123D 00	.93288038D 00	.93735456D 00
REF. WORD # 2	.88984073D 00	.89262311D 00	.89698325D 00
REF. WORD # 3	.87353658D 00	.87619407D 00	.88158498D 00
REF. WORD # 4	.91940072D 00	.90758395D 00	.89606977D 00
REF. WORD # 5	.87129309D 00	.86777135D 00	.86717850D 00
REF. WORD # 6	.84287242D 00	.84449488D 00	.84803561D 00
REF. WORD # 7	.87931544D 00	.87353970D 00	.87495288D 00
REF. WORD # 8	.87099046D 00	.87176530D 00	.87990593D 00
REF. WORD # 9	.86937945D 00	.86005171D 00	.85536060D 00
REF. WORD # 10	.86981436D 00	.86393786D 00	.86828429D 00

THE WORD : 1 SIMIL. : .93785456D 00 SHIFT# 3

THE SECOND: 2 SIMIL. : .89698325D 00

CONTINUE ?, YES(1), NO(0) 0

STOP

נספח ב': התכנית עבור המערכת ההשוואתית ללא דחיסה.

המערכת ללא דחיסה פותחה ומומשה לצרכי השוואה, ולכן לא נבנתה עבורה תכנית זיהוי שלמה במתכונת של תכנית הזיהוי עבור המערכת הכוללת דחיסה (ראה א.1). המערכת ההשוואתית ממומשת באמצעות שתי תכניות ראשיות: הראשונה WFEL, המשמשת לחישוב וקטורי הפרמטרים עבור סדרה הכוללת עד 10 הגיות (ספריה), והשניה WRDPK, הקולטת וקטורי פרמטרים של הגיות נבדקות וממיינת אותן.

1. ב. התכנית הראשית לחישוב וקטורי פרמטרים(עבור הגיות לא דחוסות) WFEL

הקלט:

1. LEV - כמו עבור ICWRS (ראה א.1).
2. LEVT - כנ"ל.
3. IO - כנ"ל.
4. INPUT LIB - הקובץ שבו הדגימות של עד 10 הגיות, קובץ זה מכונה ספריה (LIB). נדרש רק כאשר מקנים ל- IO את הערך 1.
5. הקלט מסי 7 עבור ICWRS (ראה א.1).
6. K OUT FILE - שם הקובץ שבו יכתבו הפרמטרים המחושבים.

הפלט:

1. הודעה על ערכה של דגימת השיא בהגייה (=PEAK), וערך הסף שחושב לצורך גילוי סוף ההגייה (=LEV).
2. הודעה על מספר הדגימות של ההגייה ("נטוי") לאחר גילוי סיומה (IWS), ומספר קטעי הזמן לצורך חישוב הפרמטרים (NOS).
3. וקטורי הפרמטרים שחושבו. יכתבו בקובץ אחד, באופן עוקב, לפי הסדר בו קלטה התכנית את ההגיות. לפני כל וקטור יכתב מספר קטעי הזמן שהוא מייצג (NOS).

המשתנים העיקריים:

- ISA(10000) - מכיל את דגימות ההגייה.
- TEST (750) - מכיל את וקטור הפרמטרים שחושב עבור ההגייה.

CK (12) - מכיל את הפרמטרים המחושבים מקטע הזמן הנדון.

MPS - ערכו הסופי הוא שיא ההגייה.

IWS - ערכו הסופי הוא מספר הדגימות של ההגייה לאחר גילוי סיומה.

NOS - מספר קטעי הזמן מהם מחושבים הפרמטרים.

#### השגרות הנקראות:

A2DM1 - שגרה לדגימת ההגייה (ראה א.2).

LPQ - שגרה לחישוב מקדמי ה-PARCOR עבור קטע זמן. זהה לשגרה LPF (ראה א.3), פרט לכך שב-LPQ X הוא מערך REAL בדיוק רגיל (32 סיביות).

INOUT - שגרת קלט/פלט כמתואר ב-א.1.

#### הערות:

1. ראה הערה מס' 1 ב-א.1.

2. ראה הערה מס' 3 ב-א.1, אך עם השינוי הבא - התכנית הנוכחית מבטלת את חישוב וקטור הפרמטרים עקב גלישת אמפליטודה.

בעמודים הבאים מובאת הדפסה של התכנית הראשית - WFEL. תוכנית זו זהה לחלק הראשון של ICWRS (עד לשלב המיון), פרט לכך שאין בה שלב הדחיטה ולכן ארכו של וקטור הפרמטרים המחושב משתנה מהגייה להגייה. עקב זהות זו, לא הוכנסו הערות (COMMENTS) בגוף התכנית WFEL, וניתן להעזר בהערות הרלבנטיות המופיעות בגוף התכנית ICWRS.



תכנית הראשית לחישוב וקטורי פרמטרים (עבור הגיות לא דחוסות) .WFEL

```
DIMENSION ISA(10000)
DIMENSION TEST(750)
DIMENSION Y(600),X(300),WINT(299)
DOUBLE PRECISION CK(12)
EQUIVALENCE (ISA(1),TEST(1))
ACCEPT "HEAD LEVEL :",LEV
ACCEPT "TAIL LEVEL :",LEVT
750 ACCEPT "IN. TEST INPUT (1) , EX. TEST INPUT (2) :",IO
CALL INOUT(0,"INPUT LIB. :",0)
CALL INOUT(1,"K OUT FILE :",1)
TPI=8.0*ATAN(1.0)
DO 6 J=1,299
6 WINT(J)=.54-.46*COS(TPI*(J-1)/299.)
DO 1 IT=1,10
TYPE "WORD # :",IT
IF(IO.EQ.2) GOTO 7
READ BINARY (0) (ISA(I),I=1,9945)
GOTO 8
7 CONTINUE
PAUSE
TYPE "<7>"
CALL A2DMI(ISA,LEV)
8 MPS=ISA(1)
IF(MPS.LT.0) MPS=-MPS
DO 600 I=2,9945
NT=ISA(I)
IF(NT.LT.0) NT=-NT
IF(NT.GE.2047) GOTO 700
IF(NT.LE.MPS) GOTO 600
MPS=NT
600 CONTINUE
RATIO=FLOAT(MPS)/FLOAT(2048)
VP=1.25*FLOAT(LEVT)
LEV=INT(VP*RATIO+0.5)
IF(LEV.LT.LEVT) LEV=LEVT
```

(המשך) WFEL

```

TYPE "PEAK= ",MPS," LEV= ",LEV
NAL=5
NP=0
IWS=9945
4 NT=ISA(IWS)
IF(NT.LT.0) NT=-NT
IWS=IWS-1
IF(NT.LE.LEV) GOTO 4
NP=NP+1
IF(NP.LT.NAL) GOTO 4
80 IF(IWS.LT.5110) IWS=5110
81 IF(IWS.GT.9500) IWS=9500
TYPE "<7>"
NOS=(IWS-150)/150
NOB=INT(FLOAT(NOS)*0.5+0.5)
IND=MOD(NOS,2)
TYPE "REAL UTTERANCE LENGTH : ",IWS," TOTAL # OF SEGMENTS : ",N

```

C  
C \*\*\*\* PARAMETERS EXTRACTION PHASE \*\*\*\*  
C

```

IP=0
IPK=0
DO 62 I=1,300
IP=IP+1
62 Y(I)=FLOAT(ISA(IP))
ISG=0
DO 99 K=1,NOB
IF((K.EQ.NOB).AND.(IND.EQ.1)) GOTO 53
DO 52 I=301,600
IF((K.EQ.NOB).AND.(I.GT.450)) GOTO 53
IP=IP+1
52 Y(I)=FLOAT(ISA(IP))
53 DO 110 I=1,2
L1=(I-1)*150+1
L2=(I-1)*150+300

```

(המשך) WFEL

```
DO 310 J=L1,L2
M=J-150*(I-1)
310 X(M)=Y(J)
DO 210 J=1,299
210 X(J)=X(J+1)-X(J)
X(300)=-X(300)
DO 540 J=1,299
X(J)=X(J)*INT(J)
540 CONTINUE
X(300)=X(300)*.08
CALL LPQ(X,CK,300,12)
DO 75 J=1,12
IPK=IPK+1
75 TEST(IPK)=SNGL(CK(J))
IF((K.EQ.NOB).AND.(IND.EQ.1)) GOTO 99
110 CONTINUE
IF(K.EQ.NOB) GOTO 99
DO 98 I=301,600
J=I-300
98 Y(J)=Y(I)
99 CONTINUE
TYPE "<7>"
RNOS=FLOAT(NOS)
WRITE BINARY (1) RNOS
WRITE BINARY (1) (TEST(I),I=1,IPK)
GOTO 1
700 TYPE "<7>"
TYPE "<7>"
TYPE "AMPLITUDE OVERFLOW -- TEST CANCELLED"
1 CONTINUE
STOP
END
```

2.ב התכנית הראשית למיון הגיות לא דחוסות WRDPK

הקלט:

1. INPUT REF LIB - קובץ המכיל באופן עוקב עשרה וקטורי יחוס.
2. NK - מספר מקדמי ה- PARCOR (מתוך 12 הנקלטים), שיש לקחת בחשבון לצורך יצוג קטע זמן. (בניסויים שנערכו לצרכי השוואה בין המערכות הוצב  $(NK = 6)$ ).
3. LB - פרמטר הקובע את מיקום גבולותיו של תחום הקיום עבור המסלולים האופטימלים (לצורך שיטת התכנות הדינמי). ראה ציור 7.1.
4. MD - הגורם F בביטוי (7.4) עבור מידת הדמיון בין שני קטעי זמן.
5. K TEST LIB - קובץ המכיל באופן עוקב עד 10 וקטורי פרמטרים, הפלט של התכנית הראשית WFEL, ראה ב.1, פלט סמי 3, וקטורים אלה מיצגים הגיות אותן מעונינים למיין.
6. LL - מספר הוקטורים בקובץ K TEST LIB.

הפלט:

1. אורך המסלול האופטימלי (ICNT).
2. מידות הדמיון בין כל אחד מוקטורי הייחוס לבין וקטור הנבדקת.
3. המילה שוקטור הייחוס שלה הוא הדומה ביותר לוקטור הנבדקת ומידת הדמיון ביניהם.

המשתנים העיקריים:

- SIM(63, 63) מטריצה שערכיה - מידות הדמיון החלקיות, מחושבים בתהליך התכנות הדינמי, כמוגדר בביטוי (7.2).
- TEST(12, 62) מכיל את וקטור הפרמטרים של ההגיייה הנבדקת.
- REF(12, 62) מכיל את וקטור הפרמטרים של אחת מעשר הגיות הייחוס.
- IY1(63) מכיל את המספרים הסידוריים של קטעי הזמן בהגיייה הייחוס, הנמצאים בגבול העליון (לפי ציור 7.1) של תחום הקיום עבור המסלולים האופטימליים. תחום ההשתנות של מציין המערך IY1 הוא ציר קטעי הזמן של ההגיייה הנבדקת. ערכי IY1 מתאריך את ערכי הפונקציה המקרבת את הקו המרוסק העליון שבציור 7.1.

- IY2(63) - כמו IY1, אולם הוא מתאר את הגבול התחתון, כלומר את הקו המרוסק התחתון בציר 7.1.
- DWK(10) - מכיל את מידות הדמיון בין כל אחד מוקטורי היחוס לבין וקטור הנבדקת.
- MT - מספר קטעי הזמן המיוצגים בוקטור ההגייה הנבדקת.
- MR - מספר קטעי הזמן המיוצגים בוקטור ייחוס.
- ICNT - ערכו הסופי הוא אורך המסלול האופטימלי,  $L_p$  בביטוי (7.3).
- I - מציין קטעי הזמן של הגיית יחוס.
- J - מציין קטעי הזמן של ההגייה הנבדקת.
- NF -  $Min(MR, MT)$ , כלומר  $L_{SH}$  בביטוי (7.3).
- SIM(M1,N1) - מידת הדמיון המתאימה למסלול האופטימלי, כלומר S או  $S_{M,N}$  בסעיף 7.2.

השגרות הנקראות:

- BAL - שגרה לחישוב ערכי IY1, IY2.
- SBWL - שגרה לחישוב ערכי SIM לשם קבלת SIM(M1,N1).
- SKL - שגרת פונקציה לחישוב מידת הדמיון בין שני קטעי זמן לפי פונקצית המרחק של ציביצי'ף.
- INOUT - שגרת קלט/פלט כמתואר ב- א.1.

הערות:

1. התכנית WRDPK קולטת וקטורי פרמטרים המיצגים כל קטע זמן באמצעות 12 מקדמים. ערכו של NK קובע את מספר המקדמים שילקחו (החל מהמקדם הראשון) לצורך חישוב מידת הדמיון בין שני קטעי זמן.

בעמודים הבאים מובאת הדפסה של התכנית הראשית - WRDPK.

התכנית הראשית למיון הגיות לא דחוסות WRDPK

```
DIMENSION SIM(63,63),TEST(12,62),REF(12,62)
DIMENSION IY1(63),IY2(63)
DOUBLE PRECISION DNR
DOUBLE PRECISION DWK(10),MIN,SUMRT2,DPDF,TPI
DOUBLE PRECISION SKL
EXTERNAL SKL
CALL INOUT(0,"INPUT REF LIB:",0)
ACCEPT "NUMBER OF K'S :",MK
1 TYPE " *** RECOGNITION TEST *** "
C
C *** ACCEPT PARAMETERS (LB,MD=F) ***
C
ACCEPT "BAND LIMIT : ",LB
ACCEPT "DISTANCE FACTOR : ",MD
DNR=DFLOAT(MD)/DFLOAT(2*MK)
2 CALL INOUT(1,"K TEST LIB :",0)
ACCEPT "NUMBER OF WORDS IN TEST LIB : ",LL
DO 50 KT=1,LL
C
C *** READ THE TEST PARAMETERS VECTOR ***
C
READ BINARY (1) RMT
MT=IFIX(RMT)
READ BINARY (1) ((TEST(I,J),I=1,12),J=1,MT)
DO 520 KW=1,10
C
C *** READ REFERENCE PARAMETERS VECTOR ***
C
READ BINARY (0) RMR
MR=IFIX(RMR)
READ BINARY (0) ((REF(I,J),I=1,12),J=1,MR)
C
C **** CLASSIFICATION PHASE ****
C
480 M1=MR+1
```

(המשך) WRDPK

```
NI=NT+1
C
C *** CALL FOR BAND LIMITTING ***
CALL BAL(NI,M1,IY1,IY2,LB)
C
C *** CALL FOR THE COMPUTATION OF THE SIMILARITY MEASURE (DR) ***
CALL SBVL(SKL,SIM,IY1,IY2,NI,M1,REF,TEST,NK,DNR)
C
C *** COMPUTE OPTIMAL PATH LENGTH ***
ICNT=0
I=M1
J=N1
555 IF(I.NE.1) GOTO 550
J=J-1
GOTO 502
550 IF(J.NE.1) GOTO 560
I=I-1
GOTO 502
560 IF(SIM(I,J).EQ.SIM(I-1,J)) GOTO 500
IF(SIM(I,J).EQ.SIM(I,J-1)) GOTO 501
I=I-1
J=J-1
ICNT=ICNT+1
GOTO 502
500 I=I-1
GOTO 502
501 J=J-1
502 IF((I.EQ.1).AND.(J.EQ.1)) GOTO 556
GOTO 555
556 TYPE "PATH LENGTH :",ICNT
NF=MIN0(MR,MT)
```

(המשך) WRDPK

```
C *** STORE THE NORMALIZED SIMILARITY MEASURE ***
C
520 DWK(KW)=SIM(M1,M1)*ICNT/DFLOAT(NF*NF)
CONTINUE
C
C *** DECIDE ABOUT THE WORD ***
C
MIN=DWK(1)
NW=1
DO 400 I=2,13
IF(DWK(I).LE.MIN) GOTO 400
MIN=DWK(I)
NW=I
400 CONTINUE
TYPE "<7>"
C
C *** TYPE RECOGNITION RESULTS ***
C
95 WRITE (10,95)
FORMAT (2X,"THE SIMILARITY BETWEEN TEST WORD AND REF.")
DO 96 I=1,13
WRITE (10,94) I,DWK(I)
96 CONTINUE
94 FORMAT (1X,"REF. WORD #",I3,2X,D16.8)
IF(NW.EQ.10) NW=NW-10
WRITE (10,97) NW,MIN
97 FORMAT (2X,"THE WORD :",2X,I3,2X,"SIMILARITY :",D16.8)
REWIND 0
702 ACCEPT "CONTINUE ?,YES(1),NO(2) ",JCO
CALL FCLOS(1)
IF(JCO.EQ.1) GOTO 1
STOP
END
```



3. ב. שגרה לחישוב ערכי  $IY1$ ,  $IY2$  (גבולות תחום הקיום עבור המסלולים האופטימליים) .BAL

#### הפרמטרים:

- LX - 1 + מספר קטעי הזמן המיוצגים בוקטור ההגיייה הנבדקת.
- LY - 1 + מספר קטעי הזמן המיוצגים בוקטור יחוס.
- IY1 - הגבול העליון של תחום הקיום, ראה ב.2. - מערך יציאה.
- IY2 - הגבול התחתון של תחום הקיום, ראה ב.2. - מערך יציאה.
- LB - קובע את מיקום הגבולות של תחום הקיום, ראה ב.2.

#### הסבר:

על סמך ערכי 'LX, LY, LB מחושב SLP - השיפוע של קוי הגבול. לאחר זאת מחושבים ערכי  $IY1$ ,  $IY2$  בדרך הבאה: מציבים במשוואות הישרים (קווי הגבול), את המספרים הסידוריים של קטעי הזמן מההגיייה הנבדקת (המשתנה המציין את המספרים האלה הוא IX). את תוצאות ההצבה מעגלים למספרים שלמים ואלה נרשמים ב-  $IY1(IX)$  וב-  $IY2(IX)$ . לבסוף (ב- DO 40) מחשבים את מספר מידות הדמיון החלקיות שיש לחשב בתהליך התכנות הדינמי - הערך הסופי של ISUM.

הדפסת השגרה בעמוד הבא.

BAL (חשבה לחישוב ערכי  $IY2$ ,  $IY1$ ) (גבולות תחום הקיום עבור המסלולים האופטימליים)

```

SUBROUTINE BAL(LX,LY,IY1,IY2,LB)
DIMENSION IY1(63),IY2(63)
SLP=FLOAT(LY-LB-1)/FLOAT(LX-LB-1)
LB1=LB+1
DO 5 IX=2,LB1
5  IY1(IX)=2
LB2=LB+2
DO 10 IX=LB2,LX
10  IY1(IX)=INT(FLOAT(IX-LB)*SLP)
CONTINUE
KF=LX-LB-1
DO 20 IX=2,KF
20  IY2(IX)=INT(FLOAT(LB)+0.99+FLOAT(IX)*SLP)
CONTINUE
KF1=LX-LB
DO 30 IX=KF1,LX
30  IY2(IX)=LY
ISUM=0
DO 40 IX=2,LX
40  ID=IY2(IX)-IY1(IX)+1
ISUM=ISUM+ID
CONTINUE
RETURN
END
```

4.ב. שגרה לחישוב ערכי SIM לשם קבלת מידת הדמיון בין שתי הגיות. SBWL

הפרמטרים:

- SKL - מידת הדמיון בין שני קטעי זמן לפי פונקציית המרחק של צ'ביצ'ף, המתקבלת משגרת הפונקציה SKL (ראה ב.5).
- SIM - מטריצה, ערכיה - מידות הדמיון החלקיות, מחושבים בשגרה בתהליך של תכנות דינמי.
- IY1 - מערך כניסה, פלט של השגרה BAL, ראה ב.3.
- IY2 - מערך כניסה, פלט של השגרה BAL, ראה ב.3.
- N1 - 1 + מספר קטעי הזמן המיוצגים בוקטור ההגיה הנבדקת.
- M1 - 1 + מספר קטעי הזמן המיוצגים בוקטור יחוס.
- REF - מערך כניסה, מכיל את וקטור הפרמטרים של אחת מעשר הגיות היחוס.
- TEST - מערך כניסה, מכיל את וקטור הפרמטרים של ההגיה הנבדקת.
- NK - מספר מקדמי ה-PARCOR שילקחו לצורך חישוב מידת הדמיון בין שני קטעי זמן, ראה הערה מס' 1 ב-2.
- DNR - גורם נירמול עבור מידת המרחק בין שני קטעי זמן, בכדי שמידת הדמיון המתאימה לה תהיה בין 0 ל-1. מחושב בתכנית הראשית WRDPK.

הטבר:

תחילה מאפסים את SIM. אחר כך, עבור I - מציין קטעי הזמן של הגיית יחוס J - מציין קטעי הזמן של ההגיה הנבדקת, כאשר  $IY1(J) \leq I \leq IY2(J)$ , כאשר  $2 \leq J \leq N1$ , מחשבים את  $SIM(I, J)$ . החישוב הוא תהליך רקורסיבי לפי הבטוי (7.2), כאשר  $SKL = SKL( )$  הוא  $s_{i,j}$  המופיע ב- (7.2).  
 ב-  $SKL( )$  מציבים כפרמטרים את I-1, J-1 ולא את I, J, בגלל העובדה שהשעורים (0,0), מופיעים בשגרה (עקב מגבלות שפת FORTRAN) כ- (1,1).  
 הערך של  $SIM(M1, N1)$  הוא מידת הדמיון בין שתי ההגיות המתאימה למסלול האופטימלי S או  $S_{M,N}$  בסעיף 7.2.

הדפסת השגרה בעמוד הבא.

השגרה לחישוב ערכי SIM לשם קבלת מידת הדמיון בין שתי הגיות SBWL

```
511 SUBROUTINE SBWL(SKL,SIM,IY1,IY2,N1,M1,REF,TEST,NK,DNR)
510 DIMENSION SIM(63,63),REF(12,62),TEST(12,62)
DIMENSION IY1(63),IY2(63)
DOUBLE PRECISION DNR,SK1
DO 510 J=1,N1
DO 511 I=1,M1
SIM(I,J)=0.
CONTINUE
DO 512 J=2,N1
IB=IY1(J)
IE=IY2(J)
DO 503 I=IE,IE
IF(I.GT.M1) GOTO 512
SK1=SKL(I-1,J-1,REF,TEST,NK,DNR)
503 SIM(I,J)=DMAX1(SIM(I-1,J),SIM(I,J-1),(SIM(I-1,J-1)+SK1))
512 CONTINUE
CONTINUE
RETURN
END
```

5.ב. שגרת פונקציה לחישוב מידת הדמיון בין שני קטעי זמן (צ'ביצ'ף) SKL

הפרמטרים:

II - מציין קטע הזמן של הגיית יחוס.

JJ - מציין קטע הזמן של ההגייה הנבדקת.

REF, TEST, NK, DNR - מוגדרים כפי שהוגדרו עבור השגרה SBWL, ראה ב.4.

הסבר:

מידת הדמיון בין קטע הזמן ה-II לבין קטע הזמן ה-JJ, מחושבת לפי פונקציה המרחק של צ'ביצ'ף, בהתאם לביטוי (7.4).

הדפסת השגרה:

```
DOUBLE PRECISION FUNCTION SKL(II,JJ,REF,TEST,NK,DNR)
DIMENSION REF(12,62),TEST(12,62)
DOUBLE PRECISION DPDF,SUMRT2,DNR
SUMRT2=0.0D0
DO 302 L=1,NK
DPDF=DBLE(TEST(L,JJ))-DBLE(REF(L,II))
IF(DPDF.LT.0) DPDF=-DPDF
SUMRT2=SUMRT2+DPDF
302 SKL=1.0D0-DNR*SUMRT2
303 RETURN
307 END
```

REFERENCES

117175

- [1] D.R. Reddy, "Speech Recognition by Machine: A Review," Proc. IEEE, Vol. 64, pp. 501-531, April 1976.
- [2] F. Itakura, "Minimum Prediction Residual Principle Applied to Speech Recognition," IEEE Trans. Acoust. Speech, Signal Processing, Vol. ASSP-23, pp. 67-72, Feb. 1975.
- [3] M.R. Sambur and L.R. Rabiner, "A Speaker Independent Digit Recognition System," Bell Syst. Tech. J., Vol. 54, Jan. 1975.
- [4] L.C.W. Pols, "Real-Time Recognition of Spoken Words," IEEE Trans. Computers, Vol. C-20, pp. 972-978, Sept. 1971.
- [5] T.B. Martin, "Practical Applications of Voice Input to Machines," Proc. IEEE, Vol. 64, pp. 487-501, April 1976.
- [6] G.M. White and R.B. Neely, "Speech Recognition Experiments with Linear Prediction, Bandpass Filtering, and Dynamic Programming," IEEE Trans. Acoust. Speech, Signal Processing, Vol. ASSP-24, pp. 183-188, April 1976.
- [7] D. Malah, "Time Domain Algorithms for Time Scale Variation of Speech Signals," Technion-IIT, EE Pub. No. 280, May 1976.
- [8] M.T. Clark, "Word Recognition by Means of Orthogonal Functions," IEEE Trans. Audio Electroacoust., Vol. AU-18, pp. 304-312, Sept. 1970.
- [9] A. Ichikawa, Y. Nakano, and K. Nakata, "Evaluation of Various Parameter Sets in Spoken Digits Recognition," IEEE Trans. Audio Electroacoust., Vol. AU-21, pp. 202-209, June 1973.
- [10] V.M. Velichko and N.G. Zagoruyko, "Automatic Recognition of 200 Words," Int. J. Man-Machine Studies, Vol. 2, pp. 223-234, 1970.
- [11] M.J. Coker and S.F. Ball, "An Improved Isolated-Word Recognition System Based upon the Linear Prediction Residual," Proc. ASSP-76 conf., pp. 206-209, April 1976.

- [12] G.L. Clapper, "Automatic Word Recognition," IEEE Spectrum, Vol. 8, pp. 57-69, Aug. 1971.
- [13] J.P. Haton, "A Practical Application of a Real-Time Isolated-Word Recognition System Using Syntactic Constraints," IEEE Trans. Acoust. Speech, Signal Processing, Vol. ASSP-22, pp. 416-419, Dec. 1974.
- [14] G.M. White, "Speech Recognition: A Tutorial Overview," IEEE Computer, pp. 40-53, May 1976.
- [15] J. Makhoul, "Linear Prediction: A Tutorial Review," Proc. IEEE, Vol. 63, No. 4, pp. 561-580, April 1975.
- [16] A.H. Gray and J.D. Markel, "Distance Measures for Speech Processing," IEEE Trans. Acoust. Speech, Signal Processing, Vol. ASSP-24, pp. 380-391, Oct. 1976.
- [17] S. Chandra and W.C. Lin, "Experimental Comparison Between Stationary and Nonstationary Formulations of Linear Prediction Applied to Voiced Speech Analysis," IEEE Trans. Acoust. Speech, Signal Processing, Vol. ASSP-22, pp. 403-415, Dec. 1974.
- [18] J. Makhoul, "New Lattice Methods for Linear Prediction," Proc. ASSP-76 Conf., pp. 462-465, April 1976.
- [19] J.D. Markel, "Digital Inverse Filtering - A New Tool for Formant Trajectory Estimation," IEEE Trans. Audio Electroacoust., Vol. AU-20, pp. 129-137, June 1972.
- [20] J.D. Markel and A.H. Gray, JR., "On Autocorrelation Equations as Applied to Speech Analysis," IEEE Trans. Audio Electroacoust., Vol. AU-21, pp. 69-79, April 1973.
- [21] L.R. Rabiner and B. Gold, "Theory and Application of Digital Signal Processing," Englewood Cliffs, N.J.: Prentice Hall, 1975.

A MINICOMPUTER IMPLEMENTATION OF  
A SPEECH RECOGNITION SYSTEM

Project Thesis

Submitted in partial fulfillment of the requirements  
for the degree of Master of Science  
in Electrical Engineering

b y  
Isaac Engel

Submitted to the Senate of the Technion - Israel Institute of Technology  
Tevet 5737 H a i f a January 1977



This research was carried out in the faculty of Electrical Engineering, in the signal processing laboratory, under the supervision of Dr. D. Malah.

I wish to express my deep gratitude to Dr. D. Malah for his excellent guidance and his devoted help throughout the research.

A B S T R A C T

This thesis describes a research for developing and implementing by software of an isolated word recognition system on a NOVA 2 minicomputer.

In a system which is implemented only by software, the duration of the recognition process is longer than in a system which is supported by analogue and/or digital hardware operating in real time. Hence, in this work the reduction of the recognition duration has been emphasised.

In this work a new approach is proposed for significantly reducing the recognition time.

The approach taken was to use a recently developed efficient time domain algorithm for time scale compression of speech signals. The algorithm was used as a means for time scale compression of the speech utterances to the same duration, before the parameters used for classification are extracted.

It should be noted that this compression keeps the frequency band occupied by the signal unchanged. Since the dimension of the parameter vector which represents the utterance is proportional to its length, then after the compression, its dimension is reduced. Since the computation load for extracting the parameters is significantly larger than for compression, a considerable reduction in computation of the parametric representation of the utterance is achieved. The compression causes also a reduction in the duration of the classification phase for the following two reasons. The first, the parameter vector has a dimension which is less than the natural dimension. The second reason is that the reference vectors and the test utterance vector have the same dimension. Therefore, the measure of similarity between the

test utterance vector and each of the reference vectors might be computed directly. This is in contrast to the case where the parameter vectors are of different dimensionality. In such a case it is necessary to perform time normalization which is time consuming, especially when it is implemented by dynamic programming which has been found to yield good results.

The partial correlation coefficients which are derived from the linear prediction model of speech signals, were chosen as a parametric representation for the utterance. These coefficients have proven their efficiency in an existing system, and require less computations than other known parameters.

The measure of similarity between the compared vectors is computed by the Chebyshev distance function. This method is more economical than Euclidean distance while yielding equally good recognition results.

In order to evaluate the new approach, i.e., using time scale compression of the utterance, another isolated word recognition system without compression was developed. This system is the same as the system which uses compression, except that due to the fact that time scale compression was not implemented, the vectors which are computed in this system have different dimensions, and therefore it is necessary to time normalize the compared vectors. This is carried out by the dynamic programming approach.

Comparing the amount of computations, the system which utilized the proposed approach achieved a reduction in computations to 1/3 of the computation load of the other system. In order to compare the recognition scores, recognition tests of the same utterances were performed in the two systems. The vocabulary that was chosen for the tests was the first ten Hebrew digits. It should be noted that

the recognition process is vocabulary independent, but is speaker dependent. The tests were performed under quiet and noisy (S/N = 20dB) environments, using the utterances of three speakers.

In the noisy environment, the two systems achieved the same recognition result - a success of 97.9% in 240 tests. In the quiet environment, the system with the time scale compression achieved a success of 99.1% in 360 tests. This result is slightly better than the result which was achieved by the system without compression - 98.3% in 360 tests. Hence, the time scale compression didn't cause a loss in performance, in spite of the compression of up to a factor of 3. This can be partly explained by the fact that the compression algorithm is based on properly weighting several input segments to produce a single output segment, and not on just discarding portions of the input data.