



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.

תוכן הענינים

I.....	תמצית התקציר המורחב.....	1.
II.....	מבוא.....	2.
VI.....	נקודת המוצא לעבודה זו.....	3.
VIII.....	תרומות מקוריות.....	4.
IX.....	המודל המוצע לזמן ארוך.....	5.
XI.....	שערוך סבירות מירבית של הפרמטרים.....	5.1.
XII.....	שלב 1 - מסנן מלבין רעש.....	
XII.....	שלב 2 - שערוך פונקצית עוות הזמן.....	
XIII.....	שלב 3 - שערוך צורות הגל אב-טיפוס.....	
XIV.....	שלב 4 - שערוך פונקצית השקלול של צורות הגל אב-טיפוס.....	
XV.....	שלב 5 - שערוך פונקצית ההגבר.....	
XV.....	הערות.....	
XV.....	סגמנטציה.....	5.2.
XVI.....	הפרדת דבור באמצעות המודל לזמן ארוך.....	6.
XVI.....	הפרדת אותות באמצעות אלגוריתם ה-EM.....	6.1.
XXII.....	סגמנטציה ומעקב אחר דובר.....	6.2.
XXIII.....	סכום.....	7.

תקציר מורחב

1. תמצית התקציר המורחב

עבודת מחקר זו עוסקת במודל לא-סטציונרי לזמן ארוך לדבור קולי ככלי שימושי בבעיות כגון הפרדת דוברים מתחרים בערוץ משותף בהינתן פונקציות מדגם בודדת של אות הסכום, סינתזה של דבור, וקידוד דבור.

מאחר שאות הדבור נוצר ע"י מערכת פיזיולוגית, הדינמיקה של מאפייניו חייבת לציית לאילוצים פיסיקליים מסוימים המגבילים את קצב השנויים. ניתן לפיכך להניח בקרוב ראשון שהדבור סטציונרי לזמן קצר. מסיבה זו רובן המכריע של השיטות לניתוח ועבוד דבור מחלקות את האות למסגרות קצרות, ומניחות מודל סטציונרי בו אות עירור מחזורי עובר דרך מערכת לינארית וקבועה בזמן. הסינתזה נעשית עפ"ר ע"י שרשרת קטעים סטציונריים באופן שיבטיח את רציפות האות. למרות שגישות אלה מספקות במגוון רחב של בעיות ויישומים, עולה לעתים צורך במודל לא-סטציונרי לזמן ארוך הכולל גם את הדינמיקה של המאפיינים במגבלות האילוצים הפיסיקליים.

בבעיית הפרדת דוברים, לדוגמה, כושר ההפרדה משתפר עם אורך המסגרת כל עוד הנחת הסטציונריות תקפה. עם דבור אמיתי, הארכת יתר תפגע בכושר ההפרדה, אלא אם כן נשתמש במודל לא סטציונרי. למודל כזה יש גם פוטנציאל לשמש לסינתזה ולקידוד ביישומים בהם מותרת השהייה גדולה.

בעבודה זאת מוצג מודל לא-סטציונרי לזמן ארוך לדבור קולי. ע"פ המודל, הברה מיוצגת ע"י אות ערור העובר עוות הפיך של ציר הזמן, סינון לינארי קבוע בזמן, ולבסוף הכפלה בפונקציה הגבר משתנה בזמן. אות העירור הוא שקלול משתנה בזמן של שני אותות בעלי מחזורים זהים אך צורות גל שונות במקצת. שיטה זו, המכונה "אינטרפולציה של צורות-גל אב-טיפוס" - (Prototype Waveform Interpolation - PWI), מסוגלת לתאר שינויים בתמסורת המעבר הקולי. עוות ציר הזמן מתאר את שינוי התדר היסודי, והגבר מתאר שינויים בעוצמת הקול.

למודל 6 פרמטרים וקטוריים: פונקציות עוות הזמן, מקדמי פוריה (או צורת גל) של מחזורים בודדים של צורות-גל אב-טיפוס, פונקציות השקלול שלהם, מקדמי החיזוי הלינארי של המסנן צובע הרעש, ופונקציה הגבר המשתנה בזמן. כל הפרמטרים מאולצים באופן הדוק ע"י סט אילוצים, המהווה חלק אינטגרלי של המודל.

בעיית השערוך הסימולטני של פונקצית עוות הזמן ומקדמי פוריה היא בעיית אופטימיזציה רב-מימדית, הניתנת תחת הנחות מסוימות להצגה כבעיית חשבון וריאציות וקטורית. פתרון אפשרי הוא תכנות דינמי וקטורי, אך עדיין הסיבוכיות גבוהה ביותר. לפיכך מוצע אלגוריתם איטרטיבי הכולל תכנות דינמי להתאמת זמנים בשיטה חדשה בשם *Multi-Dimensional Dynamic Time Warping (MD-DTW)*, בשלוב עם סינון מסרק (*Comb Filtering*), ומובאת הוכחת התכנסות. מציאת פונקצית השקלול ופונקצית ההגבר האופטימליות תחת האילוצים נעשית גם כן באמצעות תכנות דינמי רב-ממדי.

בשלב ראשון מודגמת מערכת אנליזה-סינתזה המבוססת על המודל. הדבור המסונתז כמעט שאינו ניתן להבחנה מהדבור המקורי.

בשלב שני, המודל מיושם לפתרון בעיית הפרדת דוברים, תוך שימוש במסגרות ארוכות מאד. בעיית שערוך הפרמטרים של שני דוברים בו זמנית נפתרת בגישה איטרטיבית, המבוססת על אלגוריתם ה- *Estimate Maximize (EM)*, ומאפשרת לשערך את הפרמטרים של כל דובר בנפרד.

תוצאות מראות כי הגישה האיטרטיבית מתכנסת במהירות, וכי למערכת המבוססת על המודל הלא-סטציונרי לזמן ארוך כושר הפרדה טוב משל אלגוריתמים סטציונריים לזמנים קצרים יותר.

2. מבוא

מאחר ודבור היא צורת התקשורת הטבעית ביותר בין אנשים, קיים עניין רב בניתוח ועבוד דבור עבור מגוון יישומים, כמו זיהוי, דחיסה, שנוי, סינתזה, ושפור מובנות ו/או איכות.

הבעייתיות הקיימת בעיבוד אותות דבור נובעת במידה רבה מהעדר מודל פיסיקלי מדויק למנגנון יצירת האות, ומאי הסטציונריות שלו. בתחום עבוד דבור מקובל לעקוף את בעיית האי-הסטציונריות ע"י חלוקת האות לקטעי זמן קצרים - מסגרות - בהם ניתן להניח סטציונריות של המאפיינים המעניינים. בהעדר מודל פיסיקלי מדויק, וכן מסיבות של נוחיות וסיבוכיות חשוב, נפוץ השימוש במודלים סטציונריים פשוטים יחסית לזמנים קצרים, כמו למשל מודל חזוי לינארי עם מקדמי חזוי קבועים, בו מערכת בעלת קטבים בלבד מעוררת ע"י אות מחזורי כאשר הדבור קולי, או רעש לבן כאשר הדבור לא-קולי. בתחומים הדורשים שחזור באיכות גבוהה של אות הדבור מתוך המודל, כמו קידוד או סינתזה, מקובלות שתי גישות עיקריות. בגישה הראשונה נעשה שימוש במודלים ספציפיים וקומפקטיים כמו *AR* או *ARMA*, לעתים בשלוב עם חזוי התדר היסודי של הערוך. מאחר ומודלים אלה אינם מסוגלים לתאר במדויק את אות הדבור, יש צורך בקידוד אות השארית - שגיאת המודל. בגישה זו נוקטים מקודדי אות השארית כמו *MultiPulse CELP* ועוד. בגישה השנייה נעשה

שימוש במודלים כלליים יותר וקומפקטיים פחות, המסוגלים במקרים רבים לתאר נאמנה גם אותות שאינם אותות דבור. דוגמאות אחדות הן קידוד בפטי מעבר Sub Band Coding - SBC ומקודדי התמרה למיניהם כמו Adaptive Transform Coder - ATC, מקודד ההתמרה הסינוסית Sinusoidal Transform Coder - STC, המקודד ההרמוני Harmonic Coder ועוד. במקודדים אלה אין צורך בקידוד אות שארית כי המודל כללי מספיק - ובד"כ גם כללי מדי - לתיאור מדויק של אות הדבור. בשתי הגישות לשחזור דבור באיכות גבוהה שתוארו לעיל, מקובל לנתח קטעים קצרים תוך הנחת סטציונריות של הפרמטריים המשוערכים. בסיתתה נהוג לשמור על רציפות האות ומאפייניו ע"י אינטרפולציה כלשהיא בין מסגרות עוקבות.

מאחר והדבור נוצר ע"י מערכת פיזיולוגית, הדינמיקה של פרמטרי המערכת ושל מאפני אות הדבור הנגזרים מהם מוגבלת ע"י אילוצים פיסיקליים. תיאורטית, עבור המודל המושלם לסינתזה של אות הדבור (ברמה של פונמות וללא תלות בשפה), קיימת התאמה חד חד ערכית (חח"ע) בין תחום ההגדרה של וקטור הפרמטרים, ובין קבוצת כל הפונמות שדובר אנושי כלשהו יכול להשמיע. במילים אחרות, בהינתן וקטור פרמטרים כלשהו העומד באילוצים שהוגדרו - וקטור פרמטרים פיזיבילי - התפוקה היא בהכרח פונמה שדובר אנושי יכול היה להשמיעה. את סט האילוצים יש לראות כחלק ממודל, שאיננו כולל דרגות חופש מיותרות בהגדרה, כי איננו מסוגל להפיק אותות שאינם דבור. למודל אידיאלי כזה היה יכול להיות פוטנציאל משמעותי במערכות לשיפור דבור, מאחר שאין אפשרות לקבל בשערוך וקטור פרמטרים לא פיזיבילי גם בנוכחות רעש. ניתן לבנות מערכות אנליזה-סינתזה בעלות חסינות מסוימת לרעש וע"י כך לשפר את איכות ו/או מובנות הדבור. זהו הרעיון מאחורי אלגוריתמים המנסיים לשפר דבור ע"י שימוש במקודדים [Paliwal 86] [Quatieri 90]. כ"כ למודל כזה יש פוטנציאל לקידוד, כי אין צורך להקצות קודים לקידוד וקטורי פרמטרים שאינם פיזיביליים בנוסף, תהיה למקודד כזה חסינות טובה לרעש.

מהדיון לעיל ברורה המוטיבציה למציאת מודלים דטרמיניסטיים המסוגלים לקרב את אות הדבור למשך פונמות שלמות. מודל טוב חייב לכלול ייצוג פרמטרי לדינמיקה של מאפייני הדבור - כי פונמה נמשכת עשרות עד מאות מ"ש, ובמהלכה ישתנו בד"כ פרמטרי הערוך ותמסורת המעבר הקולי.

נקודת מבט אחרת היא כדלקמן. בסכימות שערוך המתבססות על מודלים סטציונריים לזמן קצר, הפרמטרים משוערכים בכל מסגרת באופן עצמאי, בלי קשר לפרמטרים במסגרות הסמוכות. גישה זו איננה מנצלת את האילוצים הפיסיקליים על הדינמיקה של הפרמטרים לשיפור השערוך. יתירה מזאת - לעתים יש צורך בעיבוד נוסף של תוצאות השערוך בכדי לטפל בפרמטרים החורגים מהאילוצים. לדוגמה, כאשר משערכים pitch במסגרות, מתקבלת סדרה של ערכים דיסקרטיים אותה יש צורך להחליק בכדי לסלק ערכים חריגים. אם לעומת

זאת משוערך מסלול ה - pitch כולו בבת אחת - בכפוף לאילוצים הרלבנטיים - התוצאה אמינה יותר ואין צורך בעיבוד נוסף.

אחת הבעיות הקשות בתחום שפור דבור היא בעיית הפרדת הדוברים, ובה מודל לא סטציונרי לזמן ארוך צפוי לסייע באופן משמעותי. בבעיה זו קיימים שני דוברים מתחרים בערוץ משותף - מצב היכול להתהוות לדוגמה כאשר קיים דובר מתחרה במקור, או עקב ערב דבור (Crosstalk) בין ערוצי תקשורת. המטרה היא להפריד את שני הדוברים באופן שמובנות ואיכות הדבור תשתפרנה, וזאת בהינתן פונקצית מדגם בודדת של אות הסכום, כלומר, מיקרופון יחיד. כאשר שני הדוברים במצב קולי, ההפרדה מתבססת על הבדל בתדרי ה-Pitch שלהם. כאשר דובר אחד במצב קולי והשני במצב א-קולי, אפשר להפריד את אות הסכום למרכיב מחזורי ואות שארית ע"י מסנן מסרק אדפטיבי. כאשר שני הדוברים במצב א-קולי, קשה מאוד להפריד ביניהם. בשני המקרים הראשונים יכולת ההפרדה משתפרת ככל שאורך מסגרת האנליזה גדל, כל עוד האותות סטציונריים. בגישות הפרדה המניחות מודל סטציונרי, הארכת מסגרת האנליזה כאשר הנחת הסטציונריות איננה תקפה עלולה להזיק יותר מאשר להועיל. פתרון אפשרי הוא שימוש במודל לא סטציונרי, המסוגל מחד לתאר את האות בדייקנות לזמנים ארוכים, ומאידך איננו כולל דרגות חופש מיותרות. שימוש במודל כזה, יחד עם מסגרות אנליזה ארוכות, יכול להביא לכושר הפרדה משופר.

בעיה דומה לבעיית הפרדת הדוברים היא בעיית דיכוי דובר מפריע. גם כאן קיימים שני דוברים בערוץ משותף, אלא שאחד מהם מוגדר כדובר "רצוי" שאת מובנותו יש לשפר, בעוד השני מוגדר כדובר "מפריע". המקרה המעניין הוא כאשר אות הדבור הרצוי (Target) חלש יותר מאות הדבור המפריע (Jammer). היחס ביניהם נקרא Target to Jammer Ratio - TJR. המטרה היא לשפר את מובנות הדבור הרצוי ע"י דיכוי הדובר המפריע.

בעבודות שנעשו בתחום חוזרות ומופיעות מספר מסקנות המהוות מוטיבציה לשימוש במודל לא סטציונרי לזמן ארוך לדבור קולי:

1. עיקר הפגיעה במובנות הדובר הרצוי נגרמת ע"י הקטעים הקוליים של הדובר המתחרה.
2. להשגת יכולת הפרדה או דיכוי טובה יש צורך במסגרות אנליזה ארוכות מאוד. מסקנה זו נכונה במיוחד כאשר לדובר המפריע Pitch בתדר נמוך או כאשר לשני הדוברים Pitch בתדרים קרובים.
3. מאידך, כאשר משתמשים במודל סטציונרי, הארכת מסגרת האנליזה מעבר לתחום בו האותות סטציונריים פוגעת בכושר ההפרדה. שנויים ב-Pitch - ובמידה פחותה שינויים בתמסורת המעבר הקולי ובעוצמת הדבור, גורמים לפגיעה בכושר ההפרדה, המחמירה בקצות המסגרת ובהרמוניות בתדרים הגבוהים.

הגישה לפתרון בעיית הפרדת הדוברים שונה מהגישה לפתרון בעיית דיכוי דובר מפריע. באחרונה משערכים את פרמטרי ההפרעה ובונים מסנן אדפטיבי (adaptive comb notch filter), או מחסר ספקטרי לא-קוהרנטי, המדכא אותה. ספקטרוס האות הרצוי בתדרי המסרק מונחת אף הוא. ככל שמסגרת האנליזה קצרה יותר, כך שיני המסרק רחבות יותר, והפגיעה באות הרצוי קשה יותר. השאיפה היא לכן למסגרות ארוכות ככל שניתן. מאידך, אי סטציונריות של הדבור המפריע גורמת לכך שאיננו מחזורי, ולכן ההרמוניות שלו - בייחוד בתדרים הגבוהים - נמרחות ומתרחבות. שימוש במסגרת ארוכה מדי יגרום אז לשיני מסרק צרות מכדי שתדכאנה את כל רוחב הרמוניות האות המפריע.

הגישה לפתרון בעיית הפרדת הדוברים היא גישת אנליזה-סינתזה. בשלב האנליזה משערכים סימולטנית את הפרמטרים של כל הדוברים, ובשלב הסינתזה משחזרים אותם ע"ס הפרמטרים ששוערכו. בעיית השערוך היא בעיית מינימיזציה רב-מימדית הניתנת לתרגום למערכת משוואות לא ליניאריות. הארכת מסגרת האנליזה-בתחום בו הנחות המודל עדיין תקפות, תביא להקטנת הצמוד בין המשוואות, והמערכת תתרחק מצב של ill conditioning. במצב כזה נקבל כושר הפרדה משופר בין האותות. בסכימות איטרטיביות, דוגמת זו המבוססת על אלגוריתם ה-EM ותוצג בהמשך, צפוי שפור בקצב ההתכנסות עם הארכת מסגרת האנליזה. הארכת המסגרת חשובה במיוחד כאשר מערכת המשוואות קרובה ל ill - conditioning - למשל במצב בו תדרי ה- Pitch של הדוברים קרובים מאד. בגלל אי הסטציונריות של אות הדבור, הארכת מסגרת האנליזה חייבת להיות מלווה במודל לא סטציונרי לדבור.

לסכום, קיימים מספר תחומים בעיבוד אותות דבור, כמו הפרדת דוברים, דיכוי דובר מפריע, סינתזה, קידוד ושפור דבור בהם קיימת מוטיבציה להכנסת מודל לא סטציונרי לזמן ארוך לדבור קולי.

רוב האלגוריתמים מבצעים את ההפרדה בתחום התדר ומניחים שהצורה של כל הרמוניה היא התמרת פוריה של החלון הזמני - הנחה שאיננה מדויקת כאשר ה, תדר היסודי משתנה, מאחר שההרמוניות נמרחות. לכן, באלגוריתמים אלה מגבילים בד"כ את אורך המסגרת ל 40 מ"ש, ובכך מגבילים את כושר ההפרדה ביחס לאלגוריתמים שמתמודדים עם האי-סטציונריות ע"י רלקסציה מקומית של הנחת המחזוריות. הרלקסציה יכולה להתייחס לצורה או למקום של כל הרמוניה ביחס לנומינלי. רלקסציות כאלה נוסו רק באלגוריתמים לנחות מפריע, ובד"כ הושג נחות משופר של ההפרעה, אך לרוע המזל לעתים הדובר הרצוי הונחת אף הוא.

ב- [Stettiner & Chazan 89], אי-הסטציונריות של ה pitch- הוטמעה בתוך המודל ע"י הנחת שנוי לינארי של התדר היסודי לכל דובר במהלך המסגרת. האיכות והמובנות של דבור שהופרד ע"י אלגוריתם זה עולה על מה שהושג תוך שימוש באלגוריתם מנוון שהניח pitch קבוע.

עדיין, אפילו עם ידיעה מושלמת של מסלולי ה-pitch, איכות ההפרדה מוגבלת ע"י מידת החפיפה בין ההרמוניות של הדוברים. במקרה הגרוע ביותר, לדוברים יש pitch זהה בדיוק, וכל אלגוריתם הפרדה מבוסס pitch יהיה חסר תועלת. במקרה הכללי, ערכי ה-pitch שונים, ומידת ההפרדה תלויה בעקב במידת החפיפה בין ההרמוניות. בהנחה שהתדרים היסודיים משתנים בזמן, ניתן למזער את מידת החפיפה בין ההרמוניות ע"י הצרתן באמצעות הארכת המסגרות. מאידך, עם דבור אמיתי, שנויים בתדרים היסודיים מורחים את ההרמוניות, והשימוש במסגרות ארוכות גורם יותר נזק מאשר תועלת.

המסקנה היא שאם באמת ניתן לשפר כושר הפרדה ע"י שימוש במסגרות ארוכות יותר, יש לעשות זאת ע"י הטמעת האי-סטציונריות של האות בתוך המודל הבסיסי, תוך שימוש באילוצים הדוקים על הפרמטרים.

בנוסף לשיפור דבור, מודל לזמן ארוך יכול להוות בסיס לקידוד דבור, מאחר ומסגרות ארוכות מקלות על השגת קצבי סיביות נמוכות. האילוצים ההדוקים המוטלים על הפרמטרים עשויים לאפשר קידוד יעיל, למרות ריבוי הפרמטרים. יתירה מכך, יתכן ולמטרות קידוד, בהן לא נדרשת שמירה על צורת האות, ניתן להדק את האילוצים אף יותר מבלי לפגוע באיכות האות המשוחזר.

לסכום, במספר יישומים של עבוד דבור, כמו הפרדת דוברים, דיכוי דובר מפריע, סינתזה, דחיסה ושיפור דבור, ושימוש במודל הלא-סטציונרי לדבור קולי צפוי להיות מועיל.

[Martinelli 86] הציע מודל סינוסי לא-סטציונרי לזמן ארוך, אך לא שערך את התדרים המשתנים בזמן של ההרמוניות. במקום זאת נוון את המודל להרמוניות ברווחים שווים של תדר יסודי קבוע. המקודדים ההרמוניים של [Almeida 82-89, McAulay & Quatieri 86-90], וה- Prototype Waveform Interpolator (PWI) של [Kleijn 91-94], יכולים אמנם לתאר דבור קולי לא סטציונרי, אבל רק למשך פחות מ-30 מ"ש.

בעבודה זו בחרנו ב- *Non-stationary Spectral Model (NSM)* של Almeida ו-Tribolet - המתואר בהמשך - כנקודת המוצא.

3. נקודת המוצא לעבודה זו

שלא כמו בתחומים אחרים של עבוד דבור, הפרדת דוברים דורשת מודל מדויק לצורת האות. הפאזה היא פרמטר קריטי, ושערוך הפרמטרים של המודל הספקטרי הלא-סטציונרי, ובמיוחד פונקציות עוות הזמן, הוא מסובך.

כאשר מיישמים את המודל לזמן ארוך שאנו מציעים לבעיית הפרדת דוברים, יש לשערך בו-זמנית את פרמטרי המודל של שני הדוברים - כולל פונקציות עוות הזמן. לשערוך רובוסטי,

יש לאלץ את פונקציות עוות הזמן באופן הדוק, ע"מ שכל אחת תוכל לעקוב אחרי שנויי התדר היסודי של אחד - ורק אחד - מהדוברים.

המודל הקלאסי לדבור קולי מניח סטציונריות מקומית גם של הערור וגם של תצורת המעבר הקולי. לפיכך דבור קולי נחשב כמחזורי באופן מקומי, עם ספקטרום לזמן קצר בעל מבנה קווי, שלעתים קרובות חורג משמעויות מהספקטרום של דבור קולי אמיתי. בהמשך נציג מודל ספקטרלי לא-סטציונרי שהוצע לראשונה ע"י Almeida & Tribolet [Almeida 83], שימשם כנקודת יציאה לפתוח המודל הלא-סטציונרי לזמן ארוך לדבור קולי המוצע. המודל המוצע לא רק שמסוגל לתאר פונמות שלמות של דבור קולי בכל אורך, אלא גם מאפשר שימוש בסכימות שערור פרמטרים מעשיות.

Almeida & Tribolet הציעו מודל ספקטרלי לא-סטציונרי המהווה הכללה של מודל הקוים הספקטראליים הקלאסי, ומסלק את מגבלת הסטציונריות גם של הערור וגם של תצורת המעבר הקולי. המודל כולל מסנן ליניארי משתנה בזמן המוזן ע"י רכבת הלמים מחזורית שעברה עוות של ציר הזמן באופן המייצג את השתנות התדר היסודי. התוצאה היא מבנה ספקטרלי של קוים מוכללים, הקשורים ישירות לפתוח הדיבור הלא-סטציונרי לטור של "הרמוניות מוכללות". Almeida & Tribolet נמנעו מהבעיה הסבוכה של שערור פונקצית עוות הזמן במישרין, ע"י פתוח המודל לטור Taylor מסדר נמוך והנחת תדר יסודי קבוע למשך כל מסגרת האנליזה. בסופו של דבר הם הגיעו ל- *Harmonic Coder*, שבניגוד למודל המקורי שלהם, אינו יכול לתאר פונמות שלמות של דבור קולי. לעומת זאת, המודל המקורי הלא-מפושט הוא כללי במידה מספקת לתיאור מדויק של פונמות קוליות באורך כלשהו, ומסיבה זו בחרנו בו כנקודת מוצא לעבודתנו. להבא נתייחס למודל זה ולנגזרותיו, המקודד ההרמוני [Almeida 84-89], ה- *Sinusoidal Transform Coder (STC)* [McAulay 87-88], וה- *Prototype Waveform Interpolation (PWI) coder* [Kleijn 91-94], בשם הכולל *המודל ההרמוני המוכלל* - "Generalized Harmonic Model".

הגישה המוצעת

- שנוי המודל הספקטראלי הלא-סטציונרי NSM כך שהמערכת המשתנה בזמן תורכב מאות PWI מוכלל, פונקצית הגבר סקלרית ומסנן קטבים-בלבד משתנה בזמן המשמש לצביעת הרעש.
- פירוק המודל החדש למרכיבים ע"מ להקל בשערור הפרמטרים - כולל שערור ישיר של פונקצית עוות הזמן.
- הטלת אילוצים הדוקים על הפרמטרים ונגזרותיהם.
- פתוח סכימת שערור יעילה ורובוסטית.

4. תרומות מקוריות

בעבודה זאת אנו משנים ומרחיבים את המודל ההרמוני המוכלל המקורי של Almeida & Tribolet, ומציעים סכימה יעילה לשערוך הפרמטרים - כולל פונקציות עוות הזמן. חלק אינטגרלי מהמודל מהווה קבוצה של אילוצים הדוקים על הפרמטרים ונגזרותיהן. יש להיזהר שלא להקשיח את האילוצים יתר על המידה באופן שיפגע ביכולת המודל לתאר במדויק פונמות דבור קולי לזמן ארוך.

במודל המוצע, קומבינציה משוקללת של שתי צורות-גל אב-טיפוס (Protoype Waveforms) מורחבות מחזורית, בעלות אותו זמן מחזור נומינלי אך צורות-גל שונות, עוברת עוות של ציר הזמן. האבולוציה מצורת גל אחת לשנייה, המבוקרת ע"י פונקציות שקלול משתנה בזמן, מתארת את השנויים בעוטפת הספקטרלית, בעוד הנגזרת של פונקציות עוות הזמן עוקבת אחר התדר היסודי הרגעי. בנוסף, האות המתקבל עובר הכפלה בפונקציות הגבר משתנה בזמן - המייצגת את עוצמת הדבור, ולבסוף סינון ליניארי קבוע בזמן - הנועד לצבוע את הרעש.

אנו מפתחים סכימות איטרטיביות יעילות לשערוך 6 הפרמטרים הוקטוריים של המודל: מקדמי פוריה (או צורת-גל) של מחזורים בודדים של צורות-גל אב-טיפוס, פונקציות השקלול שלהם, פונקציות עוות הזמן, מקדמי החיזוי הליניארי של המסנן צובע הרעש, ופונקציות ההגבר המשתנה בזמן [Stettiner, Malah and Chazan 93].

פונקציות עוות הזמן משוערכת ע"י אלגוריתם איטרטיבי שמתחלף בין: (א) מסנן מסרק המשערך את האות המחזורי המתאים ביותר לאות שעבר עוות הופכי של ציר הזמן, ו-(ב) טכניקת *Multi-Dimensional Dynamic Time Warping (MD-DTW)* עם בקרת מסלול ומחיר לא מקומי [Stettiner, Malah and Chazan 94], המשערכת את פונקציות עוות הזמן המתאימה באופן האופטימלי את האות המחזורי האמור לאות המקורי. למיטב ידיעתנו, שערוך ישיר של פונקציות עוות הזמן עבור המודל ההרמוני המוכלל מעולם לא פורסם קודם לכן.

מטלת ה-DTW-MD מבוצעת ע"י כלי כללי לתכנות דינמי רב-מימדי *Multi-Dimensional Dynamic Programming (MD-DP)*, שפותח במיוחד למטרה זאת. אותו כלי בסיסי משמש גם לשערוך פונקציות ההגבר והשקלול. רב-המימדיות מאפשרת לקחת בחשבון אילוצים קשיחים ו/או רכים על הנגזרות הגבוהות (ובמיוחד עקמומיות) של פונקציות המטרה.

צורות הגל אב-טיפוס משוערכות ע"י מסנני מסרק עם חלונות משקול הנוטים קדימה או אחורה בזמן. כאמור פונקציות השקלול שלהם משוערכת ע"י אותו כלי כללי לתכנות דינמי רב-מימדי (MD-DP), עם אילוצים מתאימים.

מערכת אנליזה-סינתזה המבוססת על המודל מפיקה דבור משוחזר שכמעט ואינו ניתן להבחנה מהמקור.

הפרדת דבור בערוץ משותף. כאשר מיישמים את המודל לזמן ארוך המוצע לבעיית הפרדת דבור בערוץ משותף, יש לשערך בו-זמנית את הפרמטרים של שני הדוברים - כולל פונקציית עוות הזמן. מוצעת גישה איטרטיבית המבוססת על אלגוריתם ה-EM, בו הפרמטרים של דובר אחד משוערכים מתוך אות השארית של הדובר השני - המתקבל ע"י חיסור שחזור של הדובר השני מהאות המקורי. ניתן להראות שזהו מקרה פרטי של סכימה להפרדת אותות מסוכמים שהוצעה ע"י פדר ווינשטיין [Feder 88]. פונקציות עוות הזמן חייבות להיות מאולצות באופן הדוק, ע"מ שתוכלנה כ"א לעקוב אחרי שנויי התדר היסודי של דובר אחד - ורק אחד. שאם לא כך, פונקציית עוות זמן יחידה עשויה להתאים עצמה לחילופין לדובר הדומיננטי באותו הרגע, וע"י כך להכשיל את השערוך האיטרטיבי. באופן דומה ניתן להצדיק את הצורך באילוצים הדוקים גם על שאר הפרמטרים. כפועל יוצא, יש לראות את סט האילוצים כחלק בלתי נפרד מהמודל.

בניגוד לאלגוריתמים אחרים, האלגוריתם המוצע יכול לשמש בו-זמנית גם להפרדת דוברים וגם לדיכוי דובר מפריע, מאחר שבכל רגע אפשר לבחור לשמוע או את שחזור הדובר המעניין, או את אות השארית של הדובר המפריע. בנוסף, לא נדרשות החלטות קולי/לא-קולי, מאחר שהאלגוריתם מתאים עצמו באופן אוטומטי למצב.

המובנות והאיכות של דבור בערוץ משותף משתפרים משמעותית לאחר העיבוד. הדבור הקולי הלא-רצוי מונחת באופן משמעותי - ללחישה במקרה של אות השארית, ועוד יותר מכך באות המשוחזר. למעט מקרים בהם ה-TJR נמוך מאד (פחות מ-18 ד"ב), האלגוריתם בד"כ מתפקד טוב יותר כמערכת הפרדה מאשר כמערכת הנחתה.

יכולות האנליזה-סינתזה ושפור הדבור של המודל לזמן ארוך המוצע, לא רק שמאמתות את כשירותו, אלא גם מוכיחות שסט האילוצים שלו הדוק - אך לא יתר על המידה. בתור שכזה, המודל הוא גם בעל פוטנציאל לקידוד ושיפור דבור.

5. המודל המוצע לזמן ארוך

מוצע המודל הבא, כאשר ההנחות והאילוצים הנילוים מהווים חלק אינטגרלי מהמודל.

$$x(t) = s(t) + v(t)$$

$$s(t) = h_w(t) * \left[g(\Phi(t)) \sum_{k=0}^{K(t)} c_k(\Phi(t)) e^{jk\Phi(t)} \right]$$

כאשר:

x	אות הדבור הקולי בפועל
s	אות הדבור הקולי ע"פ המודל
v	רעש גאוסטי אדיטיבי (AGN) - לאו דווקא לבן
Φ	פונקצית עוות זמן. נגזרתה מייצגת את התדר היסודי הרגעי.
h_w	תגובה להלם של מסנן צובע רעש קבוע בזמן. לא נדרש במקרה של AWGN. הצורך בו יתברר מאוחר יותר.
g	פונקצית הגבר סקלרית המייצגת את עוצמת הדבור.
c_k	מקדמי פוריה מוכללים משתנים בזמן (קומפלקסיים), המייצגים את העוטפת הספקטרלית.
k	אינדקס ההרמוניה.
$K(t)$	מספר משתנה בזמן של הרמוניות.

ברור שהמודל איננו יחיד. קריטריונים לייצוג יחיד יובאו בהמשך.

המודל - למרות היותו מיועד לדבור קולי - יכול לייצג גם דבור לא-קולי, בתנאי שדואגים לכך שהתדר היסודי יהיה נמוך דיו כך שהספקטרום ידגם בצפיפות. לפיכך החלטות קולי/לא-קולי אינן דרושות.

הנחות

1. התדר היסודי ותצורת המעבר הקולי משתנים לאט¹.
2. כל פרמטרי המודל גם הם משתנים לאט².
3. המסנן צובע הרעש הוא מסוג קטבים-בלבד, עם מקדמי חזוי ליניארי $\{a\}$.

¹ להגדרה פורמלית ראה 3.2.

² להגדרה פורמלית ראה 3.2.

4. מקדמי פוריה המשתנים בזמן הם מהצורה

$$c_k(\Phi(t)) = \alpha(\Phi(t))c_k^1 + (1 - \alpha(\Phi(t)))c_k^2$$

כאשר α היא פונקצית שקלול סקלרית תחת אילוצים שיוגדרו בהמשך. צורות הגל אב-טיפוס מוגדרות כדלהלן.

$$p_1(\Phi(t)) \triangleq \sum_{k=0}^{K(t)} c_k^1 e^{jk\Phi(t)}$$

$$p_2(\Phi(t)) \triangleq \sum_{k=0}^{K(t)} c_k^2 e^{jk\Phi(t)}$$

ברור שלצורות הגל אב-טיפוס יש בדיוק אותו המחזור, אך צורות שונות. שקלול צורות הגל יכול להיעשות בתחום הזמן או באופן שקול בתחום התדר.

5.1 שערך סבירות מירבית של הפרמטרים

עם רעש אדיטיבי גאוסי, שערך סבירות מרבית (ML) שקול לשערך Weighted Least Squares (WLS). נציג את הבעיה בזמן בדיד. הפרמטרים הם (כ"א הוא וקטור).

$$\theta = [\Phi \quad c^1 \quad c^2 \quad \alpha \quad a \quad g]^T$$

ברצוננו לפתור

$$\underset{\theta}{\text{Min}} (x - s(\theta))^T Q^{-1} (x - s(\theta))$$

כאשר Q היא מטריצת הקווריאנס של הרעש, הכולל גם רעש אדיטיבי וגם שגיאות מודל. המינימיזציה נעשית תחת אילוצים מתאימים המפורטים בתיזה.

הערות

- תחום הקיום של α מאפשר קומבינציות לא-קונווקסיות של צורות הגל אב-טיפוס. הנושא יוצדק בהמשך.
- בנוסף לאילוצים קשיחים, נעשה שימוש באילוצים רכים, המתמחרים סטיות של נגזרות הפרמטרים מאפט.

מאחר ובעיית האופטימיזציה הלא-לינארית לעיל מסובכת למדי, מוצעת גישה רב-שלבית, תת-אופטימלית ואיטרטיבית.

שלב 1 - מסנן מלבין רעש

תפקיד המסנן הוא ללכסן את מטריצת הקווריאנס Q . היתרון יובהר בהמשך. המסנן הוא אפסים בלבד, כאשר האפסים הם הקטבים של מסנן חזוי לינארי מסדר נמוך. את מקדמי החזוי ניתן לשערך מקטעים המכילים רעש בלבד.

האות שסונן דרך המסנן מלבין הרעש נתון ע"י

$$x_j(t) = \underbrace{g(\Phi(t)) \sum_{k=0}^{K(t)} c_k(\Phi(t)) e^{jk\Phi(t)}}_{s_j(t)} + v_j(t)$$

כאשר v הוא בקרוב לבן, ו- Q_1 מטריצת קווריאנס מסוג Diagonally Dominant.

שלב 2 - שערור פונקצית עוות הזמן

למטרת שערור פונקצית עוות הזמן, נניח כדלקמן:

$$g = \text{constant} \quad 1.$$

$$c^1 = c^2 = c^0 \quad 2.$$

$$\alpha \text{ לא רלבנטי} \quad 3.$$

מתקבלת בעיית השערור הבאה, בייצוג בזמן בדיד,

$$\text{Min}_{\Phi, c^0, g} \left(x_1 - g \sum_{k=0}^K c_k^0 e^{jk\Phi} \right)^T Q_1^{-1} \left(x_1 - g \sum_{k=0}^K c_k^0 e^{jk\Phi} \right)$$

בכפוף לאילוצים

³ בהקשר של בעיית הפרדת הדוברים בערוץ משותף, לא ניתן למצוא את פונקצית ההגבר לפני שמפרידים בין הדוברים, ולכן יש להתחיל בשערור פונקצית עוות הזמן. לאחר מציאת פונקצית ההגבר, אפשר לשוב ולשערך את פונקצית עוות הזמן.

$$0 < \dot{\Phi}_{\min} < \dot{\Phi} < \dot{\Phi}_{\max}$$

$$0 < \ddot{\Phi}_{\min} < \ddot{\Phi} < \ddot{\Phi}_{\max}$$

$$\|c^0\| = \|c^1\| = \|c^2\| = 1$$

$$g = \text{constant}$$

זוהי עדיין בעיית שערך סבוכה. במקומה, אנו מציעים גישה איטרטיבית, המבצעת לסירוגין:

• מינימיזציה עבור Φ בהינתן c^0

• מינימיזציה עבור c^0 בהינתן Φ

הסכימה המתקבלת מפעילה לסירוגין מסנן מסרק פשוט, וטכניקת שפותחה במסגרת העבודה בשם *Multi-Dimensional Dynamic Time Warping - (MD-DTW)*.

אחת מההנחות הנדרשות היא שמטריצת הקווריאנס Q_1 היא אלכסונית, ולכן נדרשת הלבנת הרעש כעיבוד מקדים.

שלב 3 - שערך צורות הגל אב-טיפוס

שערך צורות הגל אב-טיפוס נעשה ע"י הפעלת מסנן מסרק על מסגרת האנליזה לאחר שעברה עוות זמן הופכי, עם חלון שקלול המדגיש פעם את אזור תחילת המסגרת ופעם את אזור הסוף - ראה ציור למטה. צורות הגל אב-טיפוס המתקבלות מנורמלות לנורמת יחידה, ע"מ לאפשר התייחסות נפרדת לאבולוצית צורת האות ולאבולוצית ההגבר. את מקדמי הפוריה קל לחשב מצורות הגל אב-טיפוס ע"י ה-DFT, למרות שאין בהם כל צורך במימוש מעשי.

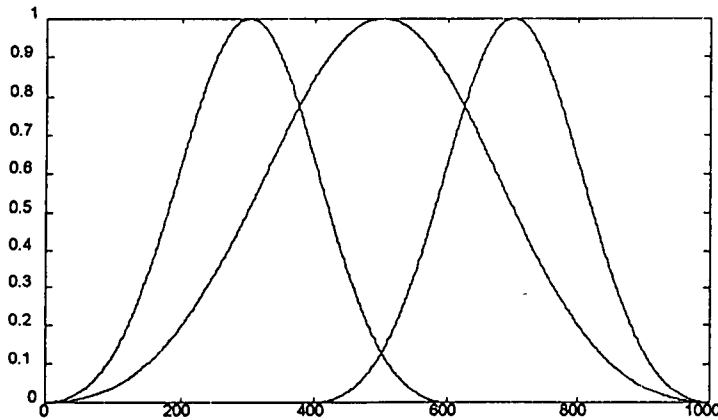


Fig.3-1 - Original window vs. two skewed windows, all Blackman.

ציור 3-1 - חלון מקורי לעומת שני חלונות מוטים.

מאחר שמסנן המסרק פועל בתחום הזמן המעוות והוא בעל מחזור קבוע, מובטח שצורות הגל אב-טיפוס המתקבלות תהיינה בעלות מחזור זהה ומותאמות בזמן, דבר המאפשר שקלול ישיר שלהן בתחום הזמן ללא צורך בכוונונים מיוחדים.

שלב 4 - שערך פונקצית השקלול של צורות הגל אב-טיפוס

בהינתן השערוך הנוכחי של פונקצית ההגבר g , יש לפתור את הבעיה הבאה בתחום הזמן המעוות u ,

$$\text{Min}_{\alpha} \int_{-\infty}^{\infty} \tilde{w}(u, \Phi) \left| \tilde{x}_1(u, \Phi) - g(u) \sum_{k=0}^{K(u)} [\alpha(u)c_k^1 + (1-\alpha(u))c_k^2] e^{jku} \right|^2 du$$

s. t.

$$\alpha_{\min} < \alpha < \alpha_{\max} \quad \alpha_{\min} < 0 \quad \alpha_{\max} > 1$$

$$\dot{\alpha}_{\min} < \dot{\alpha} < \dot{\alpha}_{\max}$$

$$\ddot{\alpha}_{\min} < \ddot{\alpha} < \ddot{\alpha}_{\max}$$

כאשר α היא פונקצית השקלול של צורות הגל אב-טיפוס, והסימון (\sim) מציין פונקציות שעברו עוות זמן הופכי. w הוא האלכסון של מטריצת הקווריאנס Q_1 ניתן להציג את הבעיה כבעיית תכנות דינמי רב-מימדי, ולפתור אותה עם אותו הכלי ששימש לפתרון בעיית ה-MD-DTW, כאשר α תופש את מקום Φ .

שלב 5 - שערור פונקצית ההגבר

בהינתן השערור הנוכחי של פונקצית השקלול של צורות הגל אב-טיפוס α , יש לפתור את הבעיה הבאה בתחום הזמן המעוות u ,

$$\text{Min}_{g(u)} \int_{-\infty}^{\infty} \tilde{w}(u, \Phi) \left| \tilde{x}_1(u, \Phi) - \underbrace{g(u) \sum_{k=0}^{K(u)} [\alpha(u)c_k^1 + (1-\alpha(u))c_k^2]}_{S_{PWI}(u)} e^{jku} \right|^2 du$$

s.t.

$$0 < g_{\min} < g < g_{\max}$$

$$\dot{g}_{\min} < \dot{g} < \dot{g}_{\max}$$

$$\ddot{g}_{\min} < \ddot{g} < \ddot{g}_{\max}$$

כאשר $g(u)$ היא פונקצית ההגבר, והסימון (\sim) מציינ פונקציות שעברו עוות זמן הופכי. ניתן להציג את הבעיה כבעיית תכנות דינמי רב-מימדי, ולפתור אותה עם אותו הכלי ששמש לפתרון בעיית ה-MD-DTW, כאשר g תופש את מקום Φ .

הערות

על שלבים 4 ו-5 יש לחזור 1-3 פעמים עד להתכנסות.

$$S_{PWI}(u) \triangleq g(u) \sum_{k=0}^{K(u)} [\alpha(u)c_k^1 + (1-\alpha(u))c_k^2] e^{jku} \quad \text{נגדיר}$$

ואז ניתן לחזור ולשערך את פונקצית עוות הזמן כאשר S_{PWI}

משמש כייחוס ל-MD-DTW במקום תפוקת מסנן המסרק S_0 .

ניתן לקבל אות משוחזר סינתטי ע"י העברת S_{PWI} דרך עוות זמן.

ואח"כ סינון במסנן צובע הרעש. נסמן את האות המתקבל ב- s_{syn} .

5.2 סגמונטציה

המודל לזמן ארוך מאפשר אמנם שימוש במסגרות ארוכות, אך יש לדאוג למיקום המסגרות בסנכרון לאירועים פונטיים כך שבאופן אידיאלי כל פונמה תמצא במסגרת משלה. לא ברור

כיצד יפעלו טכניקות מקובלות לסגמנטציה פונטית בסביבה מרובת דוברים, ולכן העדפנו להציע גישה המבוססת על המודל עצמו. הרעיון הבסיסי הוא לנסות סגמנטציות שונות, ולבחור בזו שעבורה המודל נותן התאמה מקסימלית לאות לפי קריטריון של יחס אות לאות שארית (*SRR*).

6. הפרדת דבור באמצעות המודל לזמן ארוך

כאשר מיישמים את המודל לזמן ארוך להפרדת דבור בערוץ משותף, יש לשערך את פרמטרי המודל של שני הדוברים - כולל פונקציות עוות הזמן - בו זמנית. אנו מציעים גישה איטרטיבית המבוססת על אלגוריתם ה-EM [Fessler 93][Chazan, Stettiner & Malah 93][Feder 88], שבה הפרמטרים של דובר אחד משוערכים מתוך אות השארית של הדובר השני.

בתנאים קשים כגון אלה, יש לאלץ את פונקציות עוות הזמן באופן הדוק, כך שכ"א מהן תוכל לעקוב אחרי שנויי התדר היסודי של אחד הדוברים, ורק אחד. גם שאר הפרמטרים דורשים אילוצים הדוקים, מסיבות דומות.

בהמשך הדיון נתרכז במקרה של שני דוברים בלבד, ונניח שהרעש הוא לבן.

6.1 הפרדת אותות באמצעות אלגוריתם ה-EM

אלגוריתם ה-*Estimate-Maximize* [Dempster-77] הוא שיטה כללית איטרטיבית לפתרון סוגים מסוימים של בעיות שערוך *Maximum Aposteriori* ו-*Maximum Likelihood (ML)* (MAP)- בהינתן מידע חלקי⁴. תחת תנאים מסוימים מובטחת הגדלה של הסבירות בכל איטרציה עד להתכנסות. במקרים מסוימים האלגוריתם מאפשר לפרק בעיות רב-מימדיות לסדרת בעיות ממימדים נמוכים יותר, ומוביל לפתרון איטרטיבי יעיל תוך הבטחת התכנסות לנקודה סטציונרית. בנספח A מובאים השיקולים המובילים לאלגוריתם.

פדר ווינשטיין [Feder 88] הציעו שיטה כללית לשערוך פרמטרים של אותות מסוכמים המשתמשת באלגוריתם ה-EM. הבעיה מאופיינת ע"י המודל

$$y(t) = \sum_{i=1}^I s_i(t; \theta_i) + v(t)$$

⁴ בהקשר של דבור בערוץ משותף, המידע החלקי הוא אות הערוץ המשותף, בעוד המידע המלא הוא אוסף צורות הגל הבודדות.

כאשר θ_i הוא וקטור של פרמטרים לא ידועים של מרכיב האות, ו- $v(t)$ הוא רעש אדיטיבי. השיטה של פדר ווינשטיין משערכת בוי"ז באופן יעיל את הפרמטרים. הרעיון הוא לפצל את האות הנתון $y(t)$ למרכיביו, ולשערך את הפרמטרים של כל מרכיב בנפרד. האלגוריתם מבצע איטרציות ומשתמש בשערוך הפרמטרים הנוכחי לפיצול האות הנתון, וע"י כך משפר את שערוך הפרמטרים הבא. תחת תנאים רגולריים מסוימים מובטחת התכנסות האלגוריתם לנקודה סטציונרית של פונקציית הסבירות של הפרמטרים המשווערכים.

תחת ההנחות הבאות:

א. $\{s_i(t; \theta_i)\}_{i=1}^I$ אותות דטרמיניסטיים הידועים עד כדי הפרמטרים $\{\theta_i\}_{i=1}^I$.

ב. הפונקציות $\{s_i(t; \theta_i)\}_{i=1}^I$ רציפות בפרמטרים $\{\theta_i\}_{i=1}^I$ בהתאמה.

ג. $v(t)$ הוא תהליך גאוסי וקטורי בעל תוחלת אפס ומטריצת קווריאנס $E[v(t)v^*(\sigma)] = Q\delta(t-\sigma)$.

אלגוריתם ה-EM מקבל את הצורה הבאה:

E step

For $i=1,2,\dots,I$ compute

$$x_i^{(n)}(t) = s_i(t; \hat{\theta}_i^{(n)}) + \beta_i \underbrace{\left(y(t) - \sum_{m=1}^I s_m(t; \hat{\theta}_m^{(n)}) \right)}_{v^{(n)}(t)}$$

M step

For $i=1,2,\dots,I$ compute

$$\text{Min}_{\theta_i} \int_T \left(x_i^{(n)}(t) - s_i(t; \theta_i^{(n)}) \right)^* Q^{-1} \left(x_i^{(n)}(t) - s_i(t; \theta_i^{(n)}) \right) dt \Rightarrow \hat{\theta}_i^{(n+1)}$$

הערות

רעש כן שסכומם הוא v . המודל מקבל את הצורה הבאה

$$y(t) = \sum_{i=1}^I \left(s_i(t; \theta_i) + \beta_i v(t) \right)$$

אלגוריתם ה-EM לקח את בעיית האופטימיזציה הרב-מימדית ופירק אותה ל- I בעיות אופטימיזציה נפרדות, אחת עבור כל מרכיב אות. בהנחה שכל מרכיבי האות רציפים בפרמטרים שלהם, מובטחת הגדלה של הסבירות בכל איטרציה עד להתכנסות. את האלגוריתם ניתן לתמצת בסכימת המלבנים המופיעה להלן.

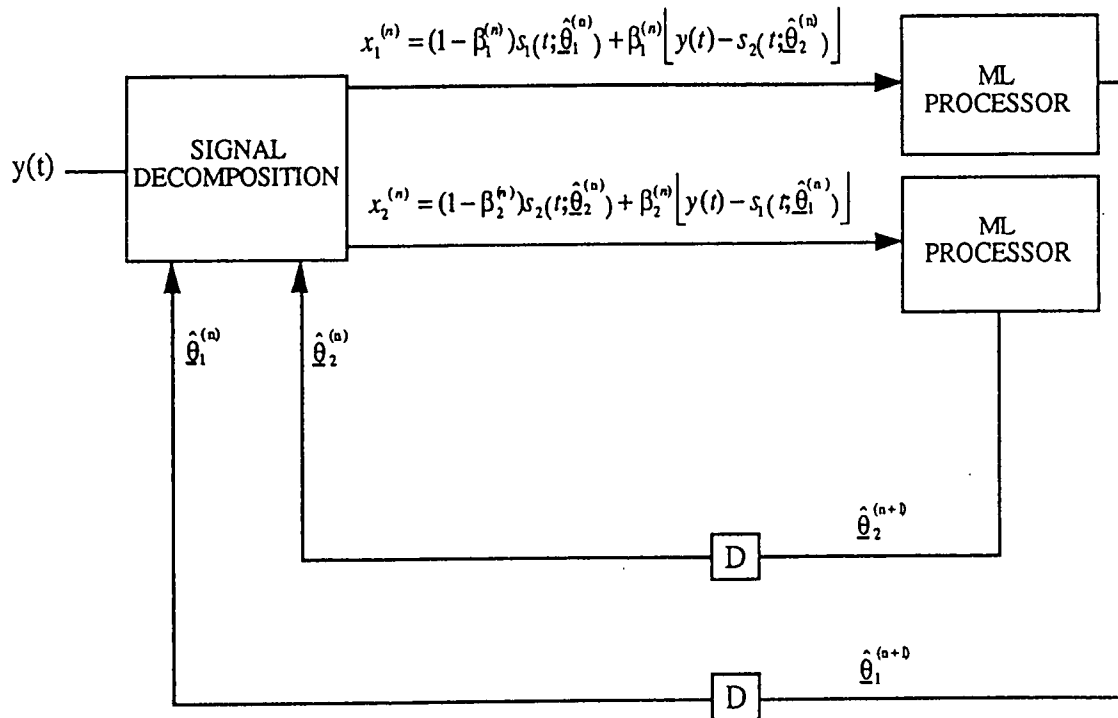


Fig. 6-1 The signal decomposition scheme of Feder and Weinstein

ציור 6-1 - סכימת הפרדת האותות של פדר ווינסטיין

גישתנו היא מקרה מיוחד של הסכימה שהוצגה לעיל⁵. ע"פ גישתנו, הפרמטרים של כל דובר משוערכים מאות שארית, המתקבל ע"י חיסור גרסה של הדובר השני המסונמת על סמך הפרמטרים האחרונים הידועים, מאות הערוץ המשותף המקורי. שימוש ראשון בגישה זאת

נעשה כבר בעבודה קודמת, העוסקת בגלוי סימולטני של מספר תדרים יסודיים בסביבה מרובת דוברים [Chazan, Stettiner & Malah 93, see appendix D]. הסכימה הבאה מציגה את גישתנו עבור המקרה של שני דוברים.

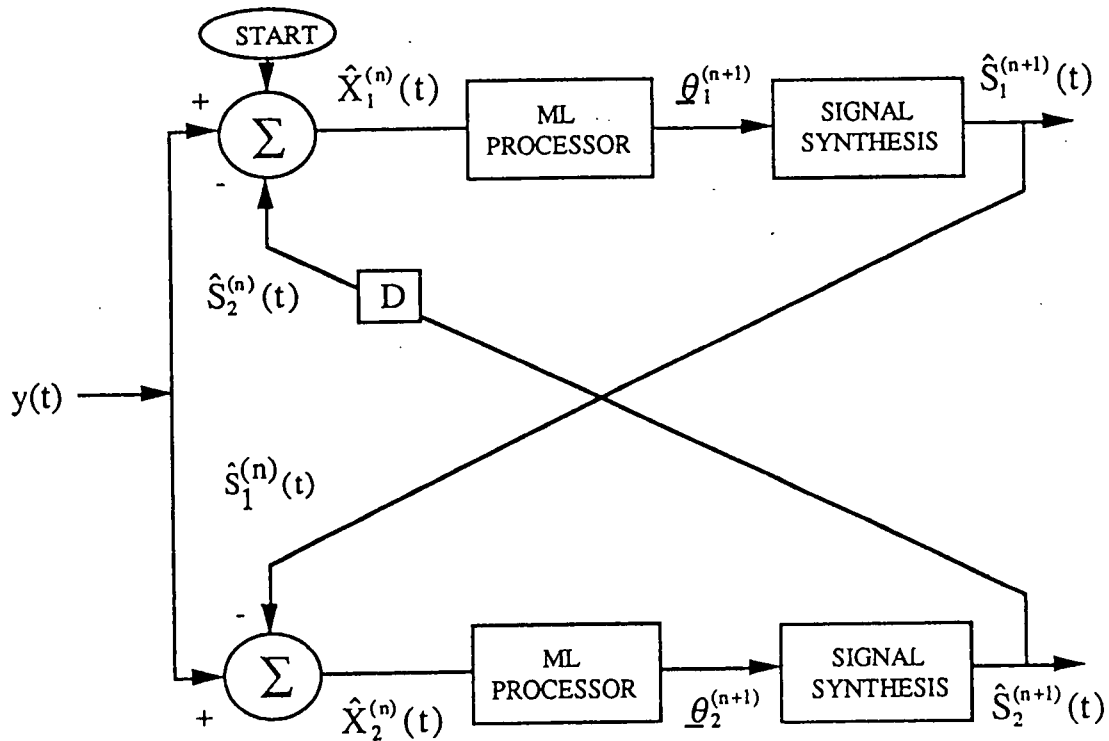


Fig. 6-2 The SAGE scheme for the two speakers case $I=2$

ציור 6-2 - סכימת ה-SAGE עבור המקרה של שני דוברים

תכונה מעניינת של סכימת ההפרדה המוצעת היא יכולת ההתאמה שלה לכל המצבים האפשריים של קוליות של שני הדוברים, המבטלת את הצורך בקבלת החלטות בנושא. עבודות קודמות כבר הראו שניתן לתאר דבור לא-קולי בצורה טובה ע"י מודל הרמוני בתנאי שההרמוניות מספיק צפופות. האופן בו מחומש מעבד ה-ML בסכימת ההפרדה גורם לנטייה טבעית לשייך תדרים יסודיים נמוכים לדבור לא-קולי, וכך נמנע הצורך בקבלת החלטות קולי/לא-קולי/שקט.

באיטריציה הראשונה של סכימת ה-EM, מעבד ה-ML ינעל קרוב לודאי על הדובר הקולי (אם קיים דובר קולי) הדומיננטי באותה המסגרת. מעבד ה-ML כבר יפעל על אות השארית של הדובר הראשון, וממילא ינעל על הדובר השני אם אף הוא קולי - ואם איננו, יסתפק בדגימה של ספקטרום אות השארית. כעת מעבד ה-ML הראשון יפעל על אות השארית של השני, וינעל שוב על הדובר הקולי (אם קיים) הדומיננטי באותה המסגרת, וכן הלאה. אם שני

הדוברים אינם במצב קולי, אזי שני מעבדי ה-ML ינעלו על תדרים יסודיים נמוכים כלשהם וידגמו בצפיפות את הספקטרום הלא-קולי המשותף, כאשר בפועל לא תושג הפרדה של הדוברים.

עם דבור אמיתי, התכנסות מושגת תוך 2-3 איטרציות. כאשר התדרים היסודיים קרובים מדי, עדיין מושגת התכנסות, אך ההפרדה גרועה.

כללית, ניתן לקבל שתי גרסאות לכל דובר: כאות מסוננת ע"פ פרמטרי המודל, או כאות שארית של הדובר השני. ניתן לראות לכן את הסכימה המוצעת גם כמערכת להפרדת דוברים, וגם כמערכת לדיכוי דובר מפריע. בד"כ עדיף האות המסוננת, אך כאשר ה-TJR נמוך מאד, מתחת ל-15 ד"ב, יש ואיכות ו/או מובנות אות השארית טובים יותר.

6.2. סגמנטציה ומעקב אחר דובר

הסגמנטציה כאן גם כן מבוססת מודל ודומה למקרה של דובר יחיד למעט מספר הבדלים, שהעיקרי שבהם הוא הצורך לעקוב אחרי דובר מסוים לפני שניתן לסנתז שחזור שלו. הסכימה אמנם מפרידה כל מסגרת לשני מרכיבי אות, אך לא מזהה אותם כמושמעים ע"י דובר מסוים. יתירה מזאת, מאחר והסכימה ננעלת קודם על הדובר הקולי הדומיננטי, הרי שבהעדר תיקון מתאים, הדוברים במוצא מסוים של הסכימה יתחלפו מדי פעם, בהתאם ל-TJR הרגעי ולקוליות. הבעיה נפתרה בזמנו באופן חלקי בעבודת המגיסטר [Stettiner 89], ואיננה מטופלת במסגרת עבודה זו.

7. טכום

אנו מציעים בעבודה זו מודל לא-סטציונרי לזמן ארוך לדבור קולי, ומיישמים אותו לבעית הפרדת דבור בערוץ משותף. קטע של דבור קולי

הוצג מודל לא-סטציונרי לזמן ארוך לדבור קולי. ע"פ המודל, הברה מיוצגת ע"י אות ערוור העובר עוות הפיך של ציר הזמן, סינון לינארי קבוע בזמן, ולבסוף הכפלה בפונקצית הגבר משתנה בזמן. אות העירור הוא שקלול משתנה בזמן של שני אותות בעלי מחזוריים זהים אך צורות גל שונות במקצת. שיטה זו, המכונה "אינטרפולציה של צורות-גל אב-טיפוס" - (Prototype Waveform Interpolation - PWI), מסוגלת לתאר שינויים בתמסורת המעבר הקולי. עוות ציר הזמן מתאר את שינוי התדר היסודי, וההגבר מתאר שינויים בעוצמת הקול.

למודל 6 פרמטרים וקטוריים: פונקצית עוות הזמן, מקדמי פוריה (או צורת גל) של מחזוריים בודדים של צורות-גל אב-טיפוס, פונקצית השקלול שלהם, מקדמי החיזוי הלינארי של המסנן

צובע הרעש, ופונקצית ההגבר המשתנה בזמן. כל הפרמטרים מאולצים באופן הדוק ע"י סט אילוצים, המהווה חלק אינטגרלי של המודל.

המודל מייצג שינויים בתדר היסודי, במעטפת הספקטרלית ובבעוצמה. מקדמי פוריה ופונקצית השקלול שלהם מייצגים שינויים במעטפת הספקטרלית, פונקצית עוות הזמן מייצגת שינויים בתדר היסודי, ופונקצית ההגבר אחראית לשינויי עוצמה.

המודל אינו יחיד, אך נבנה כך שיאפשר סכימות שערורך יעילות ורובוסטיות. בעיית השערורך הסימולטני של פונקצית עוות הזמן ומקדמי פוריה היא בעיית אופטימיזציה רב-מימדית, הניתנת תחת הנחות מסוימות להצגה כבעיית חשבון וריאציות וקטורית. הוצע אלגוריתם איטרטיבי הכולל תכנות דינמי להתאמת זמנים בשיטה חדשה בשם **Multi-Dimensional Dynamic Time Warping (MD-DTW)** - המאפשרת בקרת מסלול ופונקצית מחיר לא-מקומית - בשלוב עם סינון מסרק (Comb Filtering), והובאה הוכחת התכנסות. ע"פ סימולציות, התכנסות מושגת תוך 2-3 איטרציות. מציאת פונקצית השקלול ופונקצית ההגבר האופטימליות תחת האילוצים נעשית גם כן באמצעות תכנות דינמי רב-ממדי.

כאשר משתמשים במודל במערכת אנליזה-סינתזה, סימולציות עם קטעי דבור אמיתיים מראות כי המודל מסוגל לתאר קטעי דבור קולי באורך של פונמות שלמות בדיוק רב. הדבור המסונתז הוא באופן עקבי באיכות גבוהה מאד ונשמע טבעי אפילו בקטעים לא-קוליים.

המודל מיושם לבעיית הפרדת דוברים בערוץ משותף תוך שימוש במסגרות ארוכות (60 עד 120 מ"ש, בתלות בסגמנטציה). סכימת שערורך הפרמטרים מתכנסת תוך מספר איטרציות בודדות לנקודה סטציונרית של פונקצית הסבירות, כאשר ההתכנסות מובטחת. יכולת ההפרדה המתקבלת עולה על המושג עם מודלים סטציונריים לזמן קצר. איכות ומובנות אות הדבור הרצוי המסונתז, או אות השארית של הדובר המפריע, עולים על אלה של אות הדבור המקורי בערוץ המשותף.

(ע"פ [Wise 76] [Friedman 77] ו- [Hess 83]).

הבעיה כאן היא שערך של אות מחזורי בלתי ידוע הטכול ברעש לכן גאוסי בעוצמה בלתי ידועה. מאחר והרעש מתפלג גאוסי, משערך ה- ML שקול למשערך רבועים פחותים (LS). נציג נסוח הבעיה בזמן כדוד.

הי \underline{s} חזרה מחזורית של סדרה באורך p , q , כלומר $\underline{s}(k) = \underline{q}(k \bmod p)$. האות הנקלט \underline{r} , באורך K_0 , הוא סכום של \underline{s} ודגימות רעש \underline{n}

$$\underline{r}(k) = \underline{s}(k) + \underline{n}(k) \quad k=0, \dots, K_0$$

המצב האמיתי הוא, כמובן, שיש לשערך את זמן המחזור של אות קוואזי-מחזורי - דבור קולי - אשר הפרמטרים שלו, כולל התדר היסודי, משתנים לאט בזמן. אם נגדיר שעל זמן המחזור המשוערך להתאים לזמן המחזור באמצע מסגרת האנליזה ונניח שמסגרת האנליזה קצרה מספיק כך שפרמטרי האות השתנו במהלכה במעט וניתן לקרב אותם ע"י פונקציה ליניארית של הזמן, הרי שאפשר לראות את הסטיות בצורת האות בין המחזור שבאמצע המסגרת למחזורים הקיצוניים יותר כרעש נוסף בעל פלוג לא אחיד במסגרת, כאשר במרכזה הוא אפס.

נגדיר פונקצית חלון $w(i)$ שהיא זהותית אפס מחוץ לתחום $[0, K_0-1]$. החלון יכול להיות מלבני, או כל אחד מהחלונות המקובלים (למשל Hamming). מטרתנו למצוא את מחזור הסדרה p ואת הסדרה q שמביאים למינימום את השגיאה הרבועית הממוצעת המשוקללת ע"י החלון (WMSE),

$$(4.5.-1) \quad \min_{p, q} \left\{ \frac{1}{K_0} \sum_{k=0}^{K_0-1} w(k) (\underline{r}(k) - \underline{s}(k))^2 \right\}$$

המחזוריות של \underline{s} מאפשרת לכתוב את (4.5.-1) כ-

$$(4.5.-2) \quad \min_{p, q} \left\{ \frac{1}{K_0} \sum_{k=0}^{K_0-1} w(k) (\underline{r}(k) - \underline{q}(k \bmod p))^2 \right\}$$

נמצא כעת את המשערך ל- $\hat{q}(k)$, $\underline{q}(k)$

נגזור לפי $\underline{q}(k)$ ונשווה ל-0

$$(4.5.-3) \quad \frac{\partial (\cdot)}{\partial \underline{q}(k)} = -2 \sum_{i=0}^{K_0-1} \delta((i \bmod p) - k) \underline{w}(i) [\underline{r}(i) - \underline{q}(i \bmod p)] = 0$$

$$(4.5.-4) \quad \underline{q}(\hat{k}) = \frac{\sum_{i=0}^{K_0-1} \delta((i \bmod p) - k) \underline{w}(i) \underline{r}(i)}{\sum_{i=0}^{K_0-1} \delta((i \bmod p) - k) \underline{w}(i)}$$

$$(4.5.-5) \quad \hat{g}(k) = \frac{\sum_{j=-\infty}^{\infty} \underline{w}(k+jp) \underline{r}(k+jp)}{\sum_{j=-\infty}^{\infty} \underline{w}(k+jp)}$$

כאשר $\underline{w}(i)$ שונה מ-0 רק בתחום $[0, K_0-1]$.
 (4.5.-4) ו- (4.5.-5) מגדירות את צורת המשערוך של $\hat{g}(k)$. הסדרה \hat{g} מתקבלת ע"י קונבולוציה של סדרת הכניסה \underline{r} עם סדרת הלמים \underline{h} במרווח p המוכפלת בחלון $\underline{w}(\cdot)$. זוהי למעשה העברת סדרת הכניסה \underline{r} דרך מסנן מסרק בעל מרווח p .
 ע"פ השערוך של מחזור אחד \hat{g} ניתן לבנות את השערוך של אות הכניסה המקורי \underline{s} , ע"י הרחבה מחזורית $\hat{s}(k) = \hat{g}(k \bmod p)$.
 כעת ברצוננו למצוא את המחזור p שיביא את WMSE למינימום,

$$(4.5.-6) \quad \text{WMSE} = \frac{1}{K_0} \sum_{k=0}^{K_0-1} \underline{w}(k) (\underline{r}(k) - \hat{s}(k))^2 =$$

$$(4.5.-7) \quad \frac{1}{K_0} \left[\sum_{k=0}^{K_0-1} \underline{w}(k) \underline{r}^2(k) - 2 \sum_{k=0}^{K_0-1} \underline{w}(k) \underline{r}(k) \hat{s}(k) + \sum_{k=0}^{K_0-1} \underline{w}(k) \hat{s}^2(k) \right]$$

מאחר ו- \underline{s} הוא מחזורי לפי הגדרה, את המחובר השני ב- (4.5.-7) אפשר לכתוב שוב

$$(4.5.-8) \quad -2 \sum_{k=0}^{K_0-1} \underline{w}(k) \underline{r}(k) \hat{s}(k) = -2 \sum_{k=0}^{p-1} \hat{g}(i) \sum_{i=-\infty}^{\infty} \underline{w}(k+ip) \underline{r}(k+ip)$$

$$(4.5.-9) \quad -2 \sum_{k=0}^{p-1} \left[\frac{\sum_{i=-\infty}^{\infty} \underline{w}(k+ip) \underline{r}(k+ip)}{\sum_{i=-\infty}^{\infty} \underline{w}(k+ip)} \right] \hat{g}(k) \sum_{i=-\infty}^{\infty} \underline{w}(k+ip)$$

אך המנה היא בדיוק (4.5.-5) ולכן

$$(4.5.-10) \quad = -2 \sum_{k=0}^{p-1} \sum_{i=-\infty}^{\infty} \hat{g}^2(k) \underline{w}(k+ip)$$

$$(4.5.-11) \quad = -2 \sum_{j=0}^{K_0-1} \hat{s}^2(j) \underline{w}(j)$$

$$\begin{aligned}
 (4.5.-12) \quad & \frac{1}{K_0} \sum_{k=0}^{K_0-1} w(k) (\underline{r}(k) - \underline{\hat{s}}(k))^2 = \\
 & = \frac{1}{K_0} \left\{ \sum_{k=0}^{K_0-1} w(k) \underline{r}^2(k) - 2 \sum_{k=0}^{K_0-1} w(k) \underline{\hat{s}}^2(k) + \sum_{k=0}^{K_0-1} w(k) \underline{\hat{s}}^2(k) \right\} = \\
 & = \sum_{k=0}^{K_0-1} w(k) [\underline{r}^2(k) - \underline{\hat{s}}^2(k)]
 \end{aligned}$$

מאחר ו- \underline{r} , האות הנמדד, אינו תלוי ב- p , הרי שיש למצוא

$$(4.5.-13) \quad \text{Max}_p \sum_{k=0}^{K_0-1} w(k) \underline{\hat{s}}^2(k) \rightarrow \hat{p}$$

כאשר $\underline{\hat{s}}$ הוא הרחבה מחזורית של \hat{g} , המחושב ע"פ (4.5.-5) בתחום בו החלון $w(\cdot)$ גדול מ-0. זוהי מקסימיזציה במימד אחד.

[Wise 76] ו- [Friedman 77] ממשו PDA-ים המבוססים על (4.5.-13), בשנויים מסוימים. כללית, PDA-ים אלו מצטיינים ברובוסטיות גבוהה לרעש אדיטיבי, אך נוטים לסכול משגיאות אוקטבה לא מעטות [Hess 83].

את (4.5.-13) ניתן לפרש כחפוש מקסימום אנרגיה בתפוקה של מסנן מסרק כאשר משנים את מרווחו. את מסנן המסרק ניתן להפעיל בתחום הזמן, ע"י קונבולוציה של האות עם המסנן ע"פ (4.5.-5), או בתחום התדר, או בתחום האוטוקורלציה, כפי שנראה בסעיף הבא (4.6).

Friedman טבע את השם Pseudo-Maximum-Likelihood PDA למסערך זה, משום שההנחה שהרעש גאוסי איננה נכונה במציאות ובאה יותר מסיכות של נוחיות מתמטית. מכאן ואילך יכונה PDA זה בשם PML-PDA.