שינוי ציר הזמן באותות מוסיקליים

יזהר לבנר
שחף גרופינ

נקודות עיקריות

- הקדמה : הבעיות, הרעיון המרכזי ואת מי זה מעניין.
- משהו קטן על פורייה והספקטרום.
- הרעיון של OLA, תחילת הפתרון.
- אלגוריתם WSOLA.
- וריאציות פשוטות ל-WSOLA.
- עקרונות בבחירת פרמטרים.
- סיכומון

שינוי ציר הזמן משמעו האצה או האטה של קצב הדיבור או המוסיקה

אפשר פשוט לשנות את קצב שידור האות?

כן, אלא שאז משתנה האיכות הקולית של האות, גוון הקול (timbre),
וכן, כפי שנראה, שינוי ציר הזמן בשיטה הנאיבית מכתיב שינוי בציר התדר.

פרטים

למה?

- שיטות לשינוי ציר הזמן באותות קול משמשות כבר היום אולפני הקלטות (אודיו ווידיו), סרטים, לצורך הסינכרון בין קטעי קול שונים ובינם לבין המרכיב הויזואלי.
- לימוד שפה זרה באמצעות שליטה על קצב הדיבור.
- שינוי קצב דיבור מוקלט לאנשים כבדי שמיעה : נמצא כי הארכה משך התנועות מסייעת לכבדי שמיעה להבין את תוכן הדיבור.
- יישום של יצירת קול סינטטי מתוך טקסט כתוב.

- אלגוריתמים שונים הוצעו על מנת ששינוי ציר הזמן לא יפגעו באיכות הקולית של האות הדיבור.
- קיים מספר קטן של אלגוריתמים בסיסיים ומספר רב של גישות שונות המבוססות על האלגוריתמים הבסיסיים.
- נתמקד במשפחה של אלגוריתמים המבוססים על הרעיון של התמרת פורייה לזמן קצר (STFT –Short Time FouriereTransform).
  ה STFT הוא, בעצם, התמרת פוריה של קטע מהאות.

נקודות עיקריות

- הקדמה : הבעיות, הרעיון המרכזי ואת מי זה מעניין.
- משהו קטן על פורייה והספקטרום.
- הרעיון של OLA, תחילת הפתרון.
- אלגוריתם WSOLA.
- וריאציות פשוטות ל-WSOLA.
- עקרונות בבחירת פרמטרים.
- סיכומון

הקדמה הכרחית (אולי)

- הרעיון של פורייה :

לכל פונקציה מחזורית $g(t) = g(t+T_0)$

ניתן למצוא טור של קוסינוסים כך ש :

$$g(t) = a_0 + \sum_{k=1}^{\infty} a_k \cdot \cos\left(2\pi\left(\frac{1}{T_0}\right)k \cdot t + \phi_k\right)$$

$$G\left(\omega = 2\pi\left(\frac{1}{T_0}\right)k\right) = a_k \angle \phi_k$$

התמרת פוריה בדידה :

$$X(\omega) = \int x(t) \cdot e^{-j\omega t} \cdot dt \qquad X(\omega) = \sum_{m=-\infty}^{\infty} x(m) \cdot e^{-j\omega m}$$

התמרת פוריה לזמן קצר :

$$X(\omega,n) = \sum_{m=-\infty}^{\infty} x(m) \cdot w(m-n) \cdot e^{-j\omega m}$$

כאשר $w(m)$ פונקציית חלון בעלת תמך סופי וסימטרית סביב האפס. הגדרה דומה :

$$X(\omega,n) = \sum_{m=-\infty}^{\infty} x(m-n) \cdot w(m) \cdot e^{-j\omega m}$$

STFT מאפשרת (פחות או יותר) לברר את הרכב התדרים של האות בסביבת n.

נקודות עיקריות

- הקדמה : הבעיות, הרעיון המרכזי ואת מי זה מעניין.
- משהו קטן על פורייה והספקטרום.
- הרעיון של OLA, תחילת הפתרון.
- אלגוריתם WSOLA.
- וריאציות פשוטות ל-WSOLA.
- עקרונות בבחירת פרמטרים.
- סיכומון

נניח שמעוניינים ל"מתוח" את האות המקורי לפי a מאורכו.

נגדיר פונקציית מיפוי מתאימה: $\tau(n)=a\cdot n$

היינו רוצים שהרכב התדרים של אות היציאה (האות המתוח) יהיה זהה לזה של האות המקורי בסביבות המתאימות לפי פונקציית המיפוי:

$$\forall n,\omega \quad Y\left(\omega,\tau(n)\right)=X\left(\omega,n\right)$$

חבל רק, שהמציאות אומרת: אין דבר כזה!

אז מה לעשות?

מחפשים מרחק מינימלי בין התמרות פוריה של האותות:

$$D=\sum_{n=-\infty}^{\infty}\left(\int_{-\pi}^{\pi}\left|Y\left(\omega,\tau(n)\right)-X\left(\omega,n\right)\right|^{2}\cdot d\omega\right)$$

כיוון ש-n אינו רציף דורשים:

$$D=\sum_{k=-\infty}^{\infty}\left(\int_{-\pi}^{\pi}\left|Y\left(\omega,S_2\cdot k\right)-X\left(\omega,S_1\cdot k\right)\right|^{2}\cdot d\omega\right)$$

$$S_1,S_2\in\mathbb{Z} \quad S_2=\tau\left(S_1\right)$$

לדוגמית, נניח שמעוניינים ל"מתוח" את האות המקורי להיות באורך פי a מאורכו המקורי.

נגדיר פונקציית מיפוי מתאימה: $\tau(n)=a\cdot n$

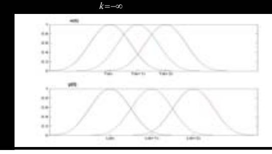
כלומר, היינו רוצים שהרכב התדרים של אות היציאה (האות המתוח) יהיה דומה ככל הניתן לזה של האות המקורי בסביבות המתאימות לפי פונקציית המיפוי.

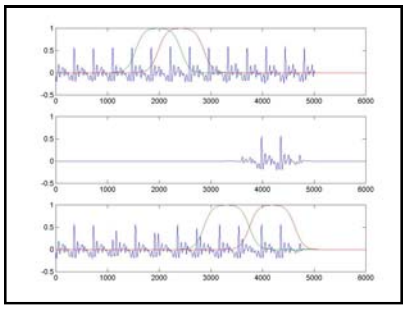
$$D=\sum_{k=-\infty}^{\infty}\left(\int_{-\pi}^{\pi}\left|Y_w\left(\omega,S_2\cdot k\right)-X_w\left(\omega,S_1\cdot k\right)\right|^{2}\cdot d\omega\right)$$

$$S_1,S_2\in\mathbb{Z} \quad S_2=\tau\left(S_1\right)$$

אות יציאה שיביא למינימום את פונקצית המרחק יהיה (לא נוכיח):

$$y(n)=\frac{\sum_{k=-\infty}^{\infty}x\left(n+S_1\cdot k-S_2\cdot k\right)\cdot w^{2}\left(n-S_2\cdot k\right)}{\sum_{k=-\infty}^{\infty}w^{2}\left(n-S_2\cdot k\right)}$$





למה זה נשמע כל כך גרוע? במה לא התחשבתי?

מיקומי החלון שרירותיים מבחינת מחזור ה- pitch על כן סביר להניח חוסר סנכרון, קפיצות פאזה.

נזכר בדרישת המוצא:

$$D=\sum_{k=-\infty}^{\infty}\left(\int_{-\pi}^{\pi}\left|Y\left(\omega,S_2\cdot k\right)-X\left(\omega,S_1\cdot k\right)\right|^{2}\cdot d\omega\right)$$

$$S_1,S_2\in\mathbb{Z} \quad S_2=\tau\left(S_1\right)$$

ניתן להראות שאם האות מחזורי וזמן המחזור ידוע, ניתן לבחור S_1 , S_2 ופונקצית חלון כך שאות יציאה שנבנה לפי נוסחת OLA יהיה מושלם!

נקודות עיקריות

- הקדמה: הבעיות, הרעיון המרכזי ואת מי זה מעניין.
- משהו קטן על פורייה והספקטרום.
- הרעיון של OLA, תחילת הפתרון.
- אלגוריתם WSOLA.
- וריאציות פשוטות ל-WSOLA.
- עקרונות בבחירת פרמטרים.
- סיכומון

לו ידענו מראש את מחזור ה-pitch היינו יכולים להתגבר על בעיות אלו באמצעות בחירת פרמטרים מתאימה.

אותות דיבור אכן "כמעט מחזוריים" רוב הזמן, אלא שזמן המחזור משתנה הפונמה לפונמה.

לכן, "זמן המחזור", להלן pitch, משתנה בזמן ואינו ידוע מראש.

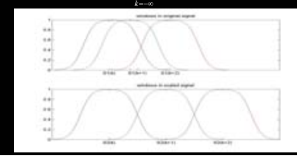
מה מציע אלגוריתם WSOLA?

1) האלגוריתם מציע לבחור את S_2 להיות חצי גודל החלון, כך שהחפיפה בין שני חלונות באות היציאה תהיה בדיוק S_2 .

2) למשקל מראש את פונקציית החלון כך שיתקיים: $\forall n \quad \sum_{k=-\infty}^{\infty}w^{2}\left(n-S_2\cdot k\right)=1$

3) לאפשר מרחב חופש מסוים במיקומי החלונות באות הכניסה על מנת לאפשר בחירת חלונות כאלו שלא יפגעו ביחסי הפאזה.

$$y(n)=\frac{\sum_{k=-\infty}^{\infty} x\left(n+S_1\cdot k-S_2\cdot k+\Delta_k\right)\cdot w^2\left(n-S_2\cdot k\right)}{\sum_{k=-\infty}^{\infty} w^2\left(n-S_2\cdot k\right)}$$



4) $-\Delta_{\max}\le\Delta_k\le\Delta_{\max}$ נבחר כך שהאות שבחלון שיכתיב יהיה דומה ככל האפשר לאות שבחלון האופטימלי.

מהו החלון האופטימלי?

חלון מאות הכניסה שחופף ב-50% לחלון הקודם שנבחר מאות הכניסה:

$$-S_2\le n<S_2 \quad x\left(n+S_1\cdot(k-1)+S_2+\Delta_{k-1}\right)$$

והחלונות האפשריים לפי:

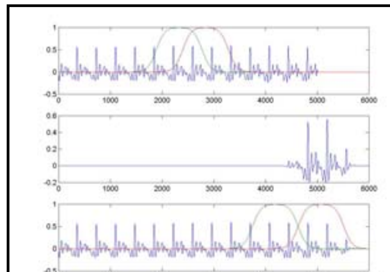
$$-S_2\le n<S_2,-\Delta_{\max}\le\delta\le\Delta_{\max} \quad x\left(n+S_1\cdot k+\delta\right)$$

איך נבדוק את הדמיון בין שני האותות?

היות ומדובר למעשה בשני וקטורים, מחפשים וקטור שהזוית בינו לבין הוקטור האופטימלי, תהיה מינימאלית.

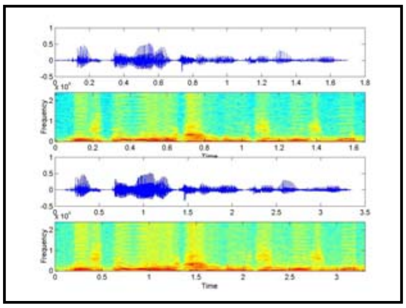
$$c_n(k,\delta)=\frac{\sum_{n=-S_2}^{S_2-1} x\left(n+S_1\cdot(k-1)+S_2+\Delta_{k-1}\right)\cdot x\left(n+S_1\cdot k+\delta\right)}{\sqrt{\sum_{n=-S_2}^{S_2-1} x^2\left(n+S_1\cdot(k-1)+S_2+\Delta_{k-1}\right)}\cdot\sqrt{\sum_{n=-S_2}^{S_2-1} x^2\left(n+S_1\cdot k+\delta\right)}}$$

$$\Delta_k=\max_{\delta}\left\{c_n(k,\delta)\ \middle|\ -\Delta_{\max}\le\delta\le\Delta_{\max}\right\}$$



אז איך זה נשמע?

ועכשיו לאט:

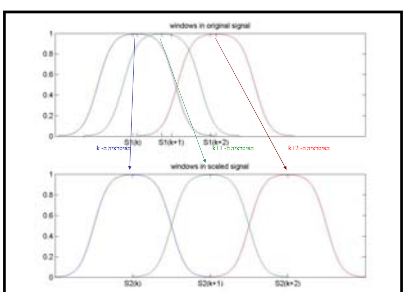


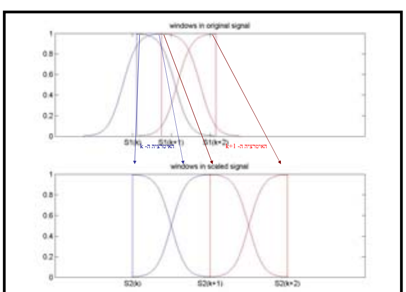
נקודות עיקריות

- הקדמה: הבעיות, הרעיון המרכזי ואת מי זה מעניין.
- משהו קטן על פורייה והספקטרום.
- הרעיון של OLA, תחילת הפתרון.
- אלגוריתם WSOLA.
- וריאציות פשוטות ל-WSOLA.
- עקרונות בבחירת פרמטרים.
- סיכומון

ואריאציות פשוטות ל- wsola

- ישום wsola מתבצע בלולאה על k, אינדקס החלון, כאשר בכל צעד של האלגוריתם מוסיפים מקטע מהאות הנכפל בפוני חלון אחת מאות הכניסה לאות היציאה.
- שיטה שקולה לחלוטין, תוסיף בכל צעד מקטע מהאות הנכפל בחצי הימני של החלון הקודם ומקטע הנכפל בחצי השמאלי של החלון הנוכחי.





נקודות עיקריות

- הקדמה: הבעיות, הרעיון המרכזי ואת מי זה מעניין.
- משהו קטן על פורייה והספקטרום.
- הרעיון של OLA, תחילת הפתרון.
- אלגוריתם WSOLA.
- וריאציות פשוטות ל-WSOLA.
- עקרונות בבחירת פרמטרים.
- סיכומון

עקרונות בבחירת פרמטרים

- **הפרמטרים שאותם עלינו לבחור:**
- סוג פונקצית החלון ואורך התמך שלה $w(n)$
- הקפיצה בין חלון לחלון באות הכניסה (או באות היציאה) S_1, S_2
- דרגת החופש בבחירת מרכזי החלונות מאות הכניסה. $\Delta_{\max}$

עקרונות בבחירת פרמטרים

- **המגבלות:**
- אורך התמך של פונקציית החלון: $2 \cdot S_2$
- היחס בין הקפיצות: $S_2 = \tau\left(S_1\right)$
- דרגת החופש בבחירת מרכזי החלונות מאות הכניסה: $\Delta_{\max} < \frac{S_1}{2}$

עקרונות בבחירת פרמטרים

- **העקרונות:**
- ככל שהקפיצות קטנות יותר, פונקציית השגיאה עליה מסתמכים "תבדוק" את האות ביותר מקומות:

$$D = \sum_{k=-\infty}^{\infty}\left(\int_{-\pi}^{\pi}\left|Y\left(\omega, S_2 \cdot k\right) - X\left(\omega, S_1 \cdot k\right)\right|^2 \cdot d\omega\right)$$

- דרגת החופש צריכה לאפשר סינכרון לפי "זמן המחזור" של האות.
- מקובל לקחת פונקציית חלון בעלת תמך של 4-2 מחזורי pitch.

נקודות עיקריות

- הקדמה: הבעיות, הרעיון המרכזי ואת מי זה מעניין.
- משהו קטן על פורייה והספקטרום.
- הרעיון של OLA, תחילת הפתרון.
- אלגוריתם WSOLA.
- וריאציות פשוטות ל-WSOLA.
- עקרונות בבחירת פרמטרים.
- סיכומון

סיכומון

- אלגוריתם WSOLA נחשב לאלגוריתם מוצלח מאוד לשינוי ציר הזמן באותות דיבור.
- אלגוריתם זה, ואלגוריתמים דומים משמשים את תעשיית המוסיקה ותעשיות הקולנוע והטלויזיה על מנת לסנכרן בין קטעי קול שונים ובינם לבין המרכיב הויזואלי.
- למרבה הצער, שיטות אלו אינן מניבות אות איכותי במיוחד כאשר מדובר באותות מוסיקליים:
- מקורי   לעומת מתוח פי 1.7:

נקודות עיקריות (נגנו עכשיו)

- הקדמה: הבעיות, הרעיון המרכזי ואת מי זה מעניין.
- שיטה למדידת השתנות האות בהתחשב בתכונות האוזן האנושית.
- שיטת דמיון החפיפה והשוואה לשיטה הקודמת.
- שילוב קטעי העתקה בפעולת WSOLA.
- שיטה לבחירת קטעי העתקה.
- פונקצית מיפוי אדפטיבית לחיפוי על קטעי ההעתקה.
- סיכום

מה כל כך מיוחד במוסיקה?

בעיה ראשונה:

רוב כלי נגינה מפיקים צלילים "כמעט-מחזוריים" (מלבד כלי הקשה), ממש כמו הפונמות הקוליות.

בעוד באותות דיבור ה-pitch יהיה 4-20 msec באותות מוסיקליים ל-pitch האפשרי טווח גדול בהרבה ולפעמים אין pitch.

בעיה שניה:

באותות דיבור שינויי האנרגיה בתחילת ובסוף הפונמה איטיים יחסית לשינויי אנרגיה בכלי נגינה.
לקטעים קצרים אלו השפעה משמעותית על האיכות הקולית של האות.

באופן כללי, בכל שני חלונות שנחפפים, מתבצע "מיצוע" של הספקטרום ולכן קטעים בהם קיימים שינויים חדים בהרכב התדרים של האות "נמרחים".

לכן, באלגוריתם המוצע, נעשה ניסיון לאתר קטעים בהם קיימת השתנות ספקטראלית משמעותית לאוזן האנושית ולהימנע מחפיפת החלונות, כלומר, להעתיק קטעים אלו ללא שינוי.

זהירות:

העתקת קטעים ארוכים מידי עלולה לגרום שינויי קצב מורגשים, בעיה שעלולה להיות אף חמורה מהבעיה אותה באנו לפתור!

נקודות עיקריות (נגנו עכשיו)

- הקדמה: הבעיות, הרעיון המרכזי ואת מי זה מעניין.
- שיטה למדידת השתנות האות בהתחשב בתכונות האוזן האנושית.
- שיטת דמיון החפיפה והשוואה לשיטה הקודמת.
- שילוב קטעי העתקה בפעולת WSOLA.
- שיטה לבחירת קטעי העתקה.
- פונקצית מיפוי אדפטיבית לחיפוי על קטעי ההעתקה.
- סיכום

לא לכל שינוי בהרכב הספקטראלי של האות, רגישה האוזן האנושית במידה שווה.

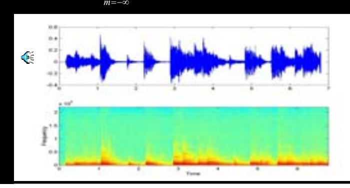
1) הבחנת האוזן האנושית, בשינויים באנרגיה של האות נמדדת בסקאלה לוגריתמית.

2) היכולת להבחין בין תדרים שונים יורדת ככל שהתדר גבוה.

3) האנרגיה הנדרשת מאות על מנת שנוכל להבחין בו עולה ככל שעולה התדר.

על מנת לברר את הרכב התדרים של האות משתמשים בהתמרת פוריה לזמן קצר:

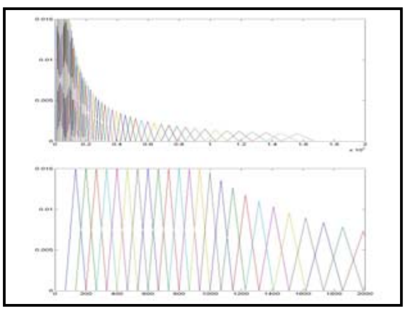
$$X_w(n\cdot J,\omega)=\sum_{m=-\infty}^{\infty} x(m)\cdot w(m-n\cdot J)\cdot e^{-i\omega m}$$



Mel – Scale Filter Bank

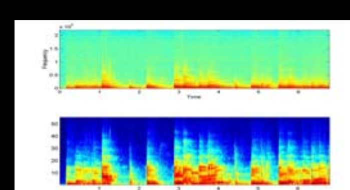
- בכל נקודת זמן, מפעילים על STFT קבוצה של מסננים המדמים (עד כמה שידוע) את תכונות האוזן האנושית.
- כל מסנן מחזיר אנרגיה (ממושקלת) בטווח תדרים מסוים.

$$MFB(n\cdot J,l)=\log_{10}\left(\int_0^{\pi}\left|X_w(\omega,n\cdot J)\cdot H_l(\omega)\right|^2\cdot d\omega\right)$$



את המקדמים המקומיים נסמן: $MFB(n\cdot J,l)$

כאשר $n\cdot J$ מרכז החלון, l מספר המסנן.



נסתכל על התמרת פורייה של האנרגיה הספקטראלית שקיבלנו, בהתחשבותנו בתכונות של האוזן האנושית:

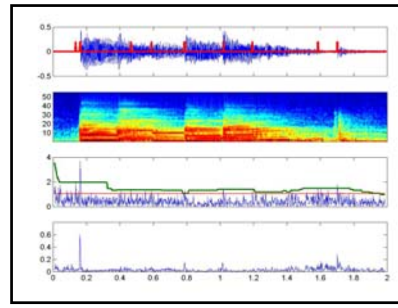
$$MFCC(n\cdot J,k)=\sum_{i=1}^{N_l} MFB(n\cdot J,l)\cdot e^{-ilk}$$

החיה הזו מכונה הקפסטרום של האות. התחשבות במקדמים הראשונים בלבד מאפשרת התעלמות מוצלחת מ"רעשים" קטנים בספקטרום.

ונעריך את השתנות האות בקטע $\left[n\cdot J,(n+1)\cdot J\right]$ ע"י הפונקציה:

$$CD(n\cdot J)=\sum_{i=1}^{K}\left|MFCC(n\cdot J,k)-MFCC((n+1)\cdot J,k)\right|^2$$

כאשר K מייצג את מספר מקדמי ה-MFC בהם בחרנו להתחשב.



נקודות עיקריות (נגנו עכשיו)

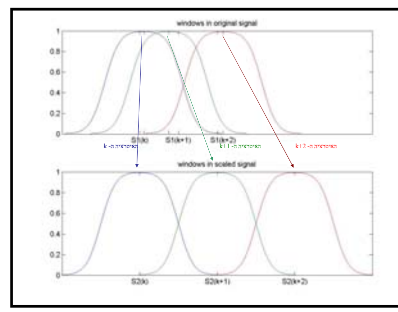
- הקדמה: הבעיות, הרעיון המרכזי ואת מי זה מעניין.
- שיטה למדידת השתנות האות בהתחשב בתכונות האוזן האנושית.
- שיטת דמיון החפיפה והשוואה לשיטה הקודמת.
- שילוב קטעי העתקה בפעולת WSOLA.
- שיטה לבחירת קטעי העתקה.
- פונקצית מיפוי אדפטיבית לחיפוי על קטעי ההעתקה.
- סיכום

שיטת דמיון החפיפה

- נזכר כיצד מציע wsola לבחור את מיקומו המדויק של חלון באות האנליזה (באות הכניסה):
- מחפשים וקטור שהזוית בינו לבין הוקטור האופטימלי, תהיה מינימאלית:

$$c_n(k,\delta)=\frac{\sum_{n=-S_2}^{S_2-1} x\left(n+S_1\cdot(k-1)+S_2+\Delta_{k-1}\right)\cdot x\left(n+S_1\cdot k+\delta\right)}{\sqrt{\sum_{n=-S_2}^{S_2-1} x^2\left(n+S_1\cdot(k-1)+S_2+\Delta_{k-1}\right)}\cdot\sqrt{\sum_{n=-S_2}^{S_2-1} x^2\left(n+S_1\cdot k+\delta\right)}}$$

$$\Delta_k=\max_{\delta}\left\{c_n(k,\delta)\mid -\Delta_{\max}\le\delta\le\Delta_{\max}\right\}$$



שיטת דמיון החפיפה

- אלגוריתם שקול, יחפוף בכל צעד שני חצאי חלון.
- בשיטה זו, ניתן לבחור את ההזזה ע"י דמיון מקסימאלי בין הקטע הנחפף, בגודל חצי חלון:

$$c_n(k,\delta)=\frac{\sum_{n=-S_2}^{-1} x\left(n+S_1\cdot(k-1)+S_2+\Delta_{k-1}\right)\cdot x\left(n+S_1\cdot k+\delta\right)}{\sqrt{\sum_{n=-S_2}^{-1} x^2\left(n+S_1\cdot(k-1)+S_2+\Delta_{k-1}\right)}\cdot\sqrt{\sum_{n=-S_2}^{-1} x^2\left(n+S_1\cdot k+\delta\right)}}$$

$$\Delta_k=\max_{\delta}\left\{c_n(k,\delta)\mid -\Delta_{\max}\le\delta\le\Delta_{\max}\right\}$$

שיטת דמיון החפיפה

- הקורלציה המנורמלת בין קטעי האות שנחפפים, היא למעשה מדד לאיכות אות הסינתזה, בלי חישובים נוספים (נו, כמעט).
- למרבה היגון, גם כאן קשה לקבוע מראש ערך סף כללי. לכן, לאחר הרצה "יבשה", יצרנו פונקצית סף אדפטיבית, ממש כמו עבור המרחק הקפסטראלי.

יתרונות מדד החפיפה לעומת מדד ההשתנות המקומית $CD(n\cdot J)$:

- חישוב מדד החפיפה לא מגדיל את זמן הריצה של האלגוריתם.
- מדד החפיפה נובע מהקשר בין האות המקורי לאות המסונתז.
- המדד מתריע על חפיפה בעייתית גם במקרה והאות יציב ספקטרלית.

יתרונות מדד ההשתנות המקומית לעומת מדד החפיפה:

- מדד ההשתנות המקומית נבנה תוך מתן עדיפות לשינויים המורגשים ע"י האוזן האנושית.
- מדד החפיפה עלול להתעלם משינויים בתדרים גבוהים.
- מדד ההשתנות נועד למנוע חפיפת קטעים שבהם השתנות ספקטרלית משמעותית, גם אם יש בניהם דמיון רב.

איזה מדד הועדף?

- בשיטה המוצעת בחרנו להסתמך על שני המדדים על מנת לקבוע את קטעי ההעתקה.
- עדיפות ניתנה למדד ההשתנות הספקטראלית המקומית.
- העדיפות באה לידי ביטוי בבחירת ערכי סף כאלו שיגרמו להעתקת קטעים מעטים יותר כתוצאה ממדד החפיפה.

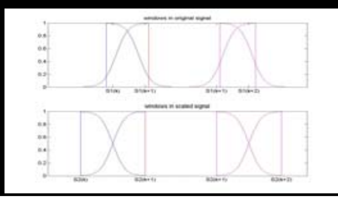
נקודות עיקריות (נגנו עכשיו)

- **הקדמה: הבעיות, הרעיון המרכזי ואת מי זה מעניין.**
- **שיטה למדידת השתנות האות בהתחשב בתכונות האוזן האנושית.**
- **שיטת דמיון החפיפה והשוואה לשיטה הקודמת.**
- **שילוב קטעי העתקה בפעולת WSOLA.**
- **שיטה לבחירת קטעי העתקה.**
- **פונקצית מיפוי אדפטיבית לחיפוי על קטעי ההעתקה.**
- **סיכום**

שילוב קטעי העתקה בפעולת WSOLA

- כזכור, בכל צעד בונים S_2 דגימות לאות היציאה ע״י הוספת מקטע הנכפל בחצי החלון הימני הקודם (מיקומו המדויק נקבע בצעד הקודם) ומקטע הנכפל בחצי החלון השמאלי הנוכחי.
- בהסתמך על שיטה זו, אנו מעוניינים להימנע מחפיפת שני חצאי חלון, המכילים קטע או קטעים בהם משתנה האות באופן חריג.
- כלומר, מעוניינים למצוא ערך סף שיחליט האם ההשתנות באזור מסוים חריגה.

אם אחד מחצאי החלונות המיודעים לחפיפה מכיל קטע בעל השתנות חריגה, האלגוריתם יעתיק J דגימות מאות הכניסה לאות היציאה.

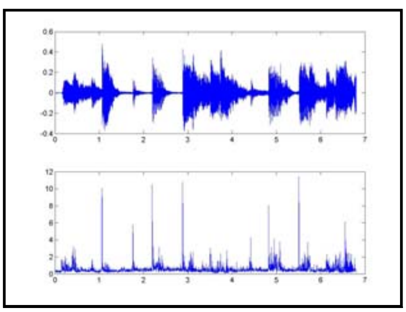


נקודות עיקריות (נגנו עכשיו)

- **הקדמה: הבעיות, הרעיון המרכזי ואת מי זה מעניין.**
- **שיטה למדידת השתנות האות בהתחשב בתכונות האוזן האנושית.**
- **שיטת דמיון החפיפה והשוואה לשיטה הקודמת.**
- **שילוב קטעי העתקה בפעולת WSOLA.**
- **שיטה לבחירת קטעי העתקה.**
- **פונקצית מיפוי אדפטיבית לחיפוי על קטעי ההעתקה.**
- **סיכום**

הבעיות בבחירת ערך הסף

- אנו מעוניינים בהעתקת קטעים קצרים ומשמעותיים ככל הניתן.
- ערך סף שמתאים לאות אחד עלול לא להתאים לכל לאות אחר, כלומר ערך הסף צריך להיות פונקציה של האות הספציפי.
- בחירת ערך סף שיעתיק אחוז סביר מהאות, עלול לגרום שינויי קצב, עקב בחירת קטעי העתקה ארוכים יחסית.

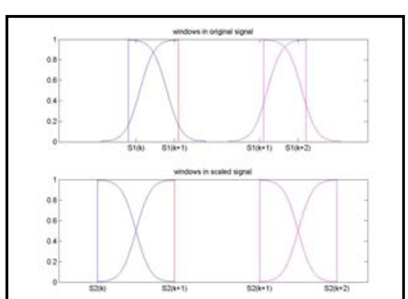


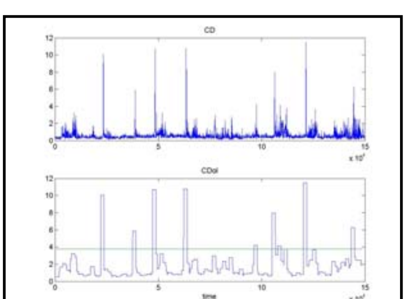
הפתרון: פונקציית סף

- נגדיר פונקציית סף $T_{LCD}(n\cdot J)$ הנקבעת מקומית ע״י פונקציית ההשתנות.
- על מנת לבחור בכל אזור את הקטעים החשובים ביותר להעתקה, נקבע את ערך הסף המקומי לפי אחוזון מהפונקציה:

$$CD_{ol}(n\cdot J)=\max\left\{CD\left((n+i)\cdot J\right)\ \middle|\ 0\le i\le\frac{\hat{N}_{ol}}{J}\right\}$$

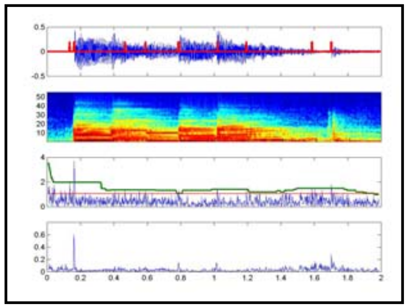
כאשר $\hat{N}_{ol}$ מציין את מספר הדגימות הממוצע שיועתק כתוצאה מקטע השתנות משמעותית בודד.





שילוב פונקציית סף מקומית וגלובלית

- בעיה : איך להמנע מהעתקת קטעים שנבחרו ע״י הסף המקומי בגלל אזור ארוך בעל השתנות מינורית.
- על מנת להתגבר על בעיה זו, מוצאים ערך סף גלובלי (שוב כאחוזון מ- $CD_{ol}(n \cdot J)$).
- קטע נבחר להעתקה רק אם ערך ההשתנות בו גדול משני ערכי הסף : ערך הסף המקומי וערך הסף הגלובלי.



נקודות עיקריות (נגנו עכשיו)

- **הקדמה : הבעיות, הרעיון המרכזי ואת מי זה מעניין.**
- **שיטה למדידת השתנות האות בהתחשב בתכונות האוזן האנושית.**
- **שיטת דמיון החפיפה והשוואה לשיטה הקודמת.**
- **שילוב קטעי העתקה בפעולת WSOLA.**
- **שיטה לבחירת קטעי העתקה.**
- **פונקצית מיפוי אדפטיבית לחיפוי על קטעי ההעתקה.**
- **סיכום.**

בעיית שינוי פונקצית המיפוי

- נניח שנתבקשנו למתוח קטע לפי 2 מאורכו המקורי ואכן מתחנו את כל הקטע פי 2, מלבד 10% ממנו שהוגדרו כקטעי ״אל תיגע בי״
- בפועל אורכו של אות היציאה יהיה :

$$2 \cdot X_{len} \cdot (0.9) + X_{len} \cdot (0.1) = 1.9 \cdot X_{len}$$

אין זמן

יחס מתיחה אדפטיבי

- הבעיה : לא יודעים מראש איזה אחוז מהקטעים יועתק בפועל.
- פתרון : שינוי ציר הזמן כך שיחס המתיחה ישתנה תוך כדי פעולת האלגוריתם בהתאם ל״סטיה״ מפונקציית המיפוי הנדרשת.
- לצורך כך נצטרך :
  1) שיטה לשינוי אקטיבי של פונקציית המיפוי.
  2) שיטה למדוד את הסטייה מהדרישה המקורית.

איך ניתן לשנות את יחס המתיחה?

- יחס המתיחה בא לידי ביטוי ביחס שבין S_1 ל- S_2 ולכן שינוי של אחד מהם ישנה את יחס המתיחה.
- שינוי של S_2 יחייב שינוי פונקצית החלון.
- שינוי של S_1 יחייב שינוי של $\Delta_{\max}$.
- השיטה שנבחרה כאן היא שינוי S_1 על מנת להימנע משינוי פונקצית החלון.

מדידת הסטייה המקומית

- אם נמתח את האות לפי נוסחת WSOLA, היחס בין מרכזי החלונות (בהתעלם מ- Δ_k) באות הכניסה ובאות היציאה יהיה יחס המתיחה :

$$\frac{Y_{offset}}{X_{offset}} = \frac{S_2 \cdot k}{S_1 \cdot k} = \frac{S_2}{S_1} = \alpha$$

- העתקת קטע, יוצרת סטייה בגודל :

$$Y_{offset} - X_{offset} \cdot \alpha$$

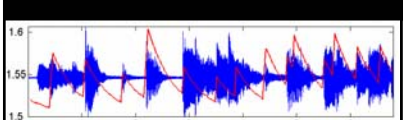
קביעת S_1 כך שיחפה על הסטייה

- נדרוש תיקון הסטייה תוך פרק זמן קבוע מראש (מספר דגימות קבוע מראש) N_{fix}
- על מנת לתקן את השגיאה בפרק זמן זה יחס המתיחה המקומי צריך להיות :

$$\frac{S_2}{S_1} = \frac{\alpha \cdot N_{fix} - \left(Y_{offset} - X_{offset} \cdot \alpha\right)}{N_{fix}}$$

$$S_1 = S_2 \cdot \frac{N_{fix}}{\alpha \cdot N_{fix} - \left(Y_{offset} - X_{offset} \cdot \alpha\right)}$$

הנה, תראו



יחס המתיחה על פני הזמן (והלהקה של ג׳ופלין ברקע)

נקודות עיקריות (נגנו עכשיו)

- **הקדמה: הבעיות, הרעיון המרכזי ואת מי זה מעניין.**
- **שיטה למדידת השתנות האות בהתחשב בתכונות האוזן האנושית.**
- **שיטת דמיון החפיפה והשוואה לשיטה הקודמת.**
- **שילוב קטעי העתקה בפעולת WSOLA.**
- **שיטה לבחירת קטעי העתקה.**
- **פונקצית מיפוי אדפטיבית לחיפוי על קטעי ההעתקה.**
- **סיכום.**

סיכום

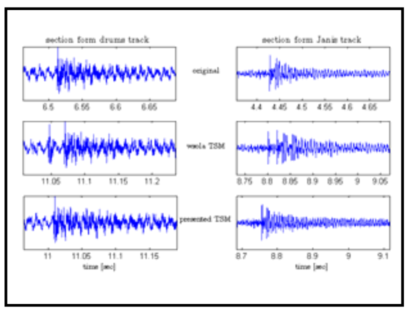
- נוסחת OLA מציעה לבנות אות יציאה, כך שהמרחק בין הספקטרום המקומי שלו, לספקטרום המקומי של אות הכניסה יהיה מינימאלי.
- אלגוריתם WSOLA מציע שיטה לפתור את בעיית הסנכרון ל-pitch.
- האלגוריתם כאן מציע למצוא את המקומות בהם נפגעת איכות האות באופן קיצוני ולהימנע ממתיחתם, תוך חיפוי על הבעיות הנגרמות עקב כך.
- במבחני השמעה שנערכו בקרב מוסיקאים חובבים וטכנאי קול, הועדפה בבירור שיטה זו על פני אחרות.

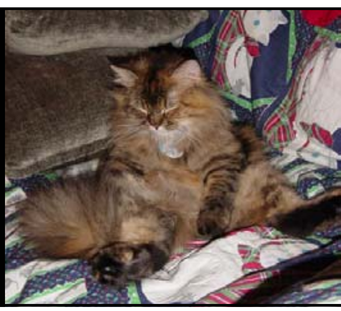
Track name	Original Length [sec]	Stretch ratio	Non-scaled ratio	Present algorithm preferred	Cannot choose	WSOLA preferred	Music excerpt reference
Pipa	8	x 2	11.6 %	7	0	0	Pipa (4-stringed guitar-like planked instrument [7])
Dylan	7	x 2	6.9 %	6	1	0	"Blowin' in the Wind," sung by Bob Dylan alone with guitar and harmonica
Drums	13	x 1.7	11.0 %	6	0	1	"Ghost Train," played by Counting Crows, mostly on drums and bass guitar
Janis	6	x 2	12.7 %	6	0	1	"Trouble in Mind," sung by Janis Joplin, electric guitar
Piano	9	x 2	8.2 %	3	3	1	Track of electric piano

Table 1: Listening tests and results. The fourth column (non-scaled ratio) describes the percentage of transients that were not time scaled by the algorithms.

דוגמיות השמעה

- הנה אות סיני עתיק (כמעט):
- הנה האות מתוח לפי 2 מאורכו המקורי באמצעות WSOLA:
- והנה האות מתוח לפי 2 מאורכו, לפי השיטה המוצעת במאמר:





מה הולך פה?

נסמן את האות המקורי: $x(t)$

ונאט או נאיץ לפי: $y(t) = x(a \cdot t)$

$$X(\omega) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j\omega \cdot t} \cdot dt$$

$$Y(\omega) = \int_{-\infty}^{\infty} y(t) \cdot e^{-i\omega \cdot t} \cdot dt = \int_{-\infty}^{\infty} x(a \cdot t) \cdot e^{-i\omega \cdot t} \cdot dt$$

$$\underset{s = a \cdot t}{=} \int_{-\infty}^{\infty} x(s) \cdot e^{-i\omega \frac{s}{a}} \cdot \frac{ds}{a} = \frac{1}{a} \int_{-\infty}^{\infty} x(s) \cdot e^{-i\frac{\omega}{a}s} \cdot ds = \frac{1}{a} \cdot X\left(\frac{\omega}{a}\right)$$

חזרה