

הרצאה מס' 2: רקע פיזיקלי ופיזיולוגי

מוטיבציה

ניתוח הבעיה.

ניסיון ללמוד ולחקות את מנגנון השמיעה האנושי.

ננסה לענות על השאלות הבאות:

מה זה קול ?

איך נוצר קול ?

איך נוצר דיבור ?

מבוא לניתוח ספקטרולי

נתונה סידרה של מספרים $a_0 \dots a_{n-1}$.

n חזקה שלמה של 2.

נתייחס לסדרת המספרים כווקטור במרחב \mathbb{R}^n .

הייצוג של הווקטור עפ"י הבסיס הסטנדרטי הוא לפיכך $(a_0, a_1, \dots, a_{n-1})$.

כל אוסף של n ווקטורים בת"ל מהווה בסיס.

נבחר בסיס כדלקמן:

$$(e_0, \dots, e_{n-1})_j, (f_0, \dots, f_{n-1})_j$$

$$(e_i)_j = \cos(2\pi i * j/n) \quad 0 \leq j \leq n/2$$

$$(f_i)_j = \sin(2\pi i * j/n) \quad 0 < j < n/2$$

כלומר הבסיס הוא אוסף של קוסינוסים וסינוסים בתדירויות שונות.

כדי לייצג ווקטור עפ"י הבסיס החדש, נטיל את הווקטור על כל אחד מאברי הבסיס.

$$e_j = \sum_{i=0}^{n-1} a_i \cdot \cos \frac{2\pi i j}{n} \quad f_j = \sum_{i=0}^{n-1} a_i \cdot \sin \frac{2\pi i j}{n}$$

מבוא לניתוח ספקטרלי (2)

לדוגמא: (עבור $n=4$)

$$e_1 = (1, 1, 1, 1)$$

$$e_2 = (1, 0, -1, 0)$$

$$e_3 = (1, -1, 1, -1)$$

$$f_1 = (0, 1, 0, -1)$$

ייצוג של פונקציה (או סידרה) באמצעות בסיס הרמוני קרוי פירוק פורייה.

משמעות פיזיקלית/מתמטית: תופעות שונות בטבע פועלות בציר התדר.

לדוגמא: תמסורת של מעגלים חשמליים.

בהינתן אות כניסה (עוצמה כפונקציה של זמן), צ"ל את אות היציאה מן המעגל.

ידועה השפעת המעגל על אותות הרמוניים.

הפתרון: נבצע פירוק פורייה לאות הכניסה.

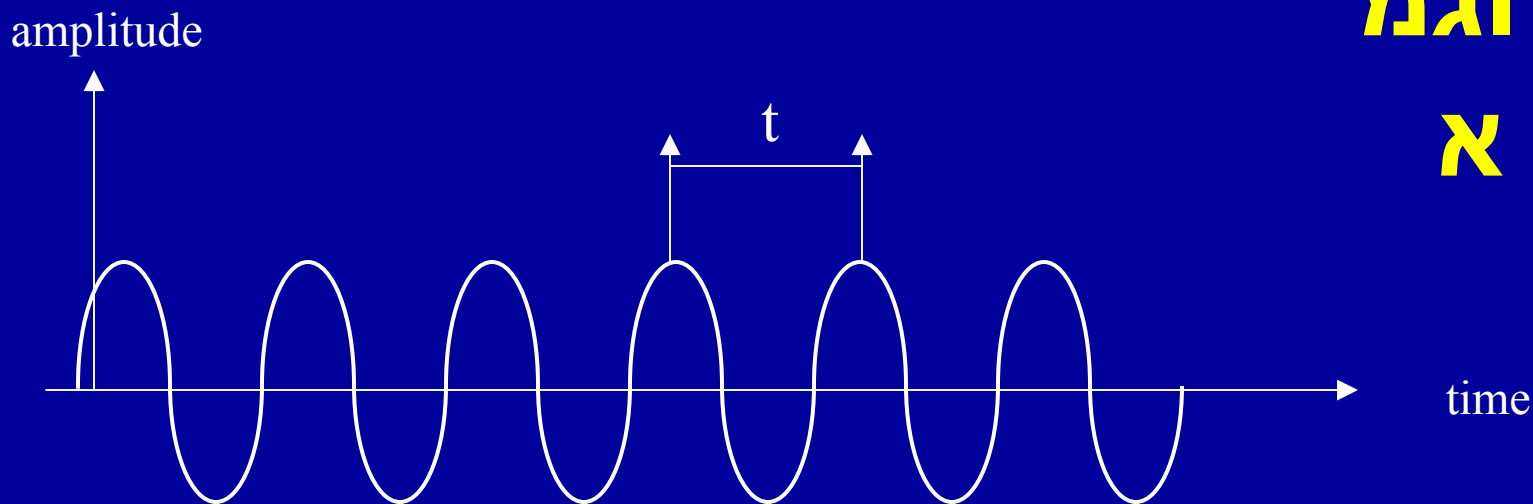
נמצא את השפעת המעגל על רכיבי האות השונים.

נבצע פירוק פורייה הפוך (המרה לבסיס הסטנדרטי) ונמצא את אות היציאה.

הערה: קיים אלגוריתם מהיר ($O(n \log n)$) לחישוב פירוק פורייה (FFT).

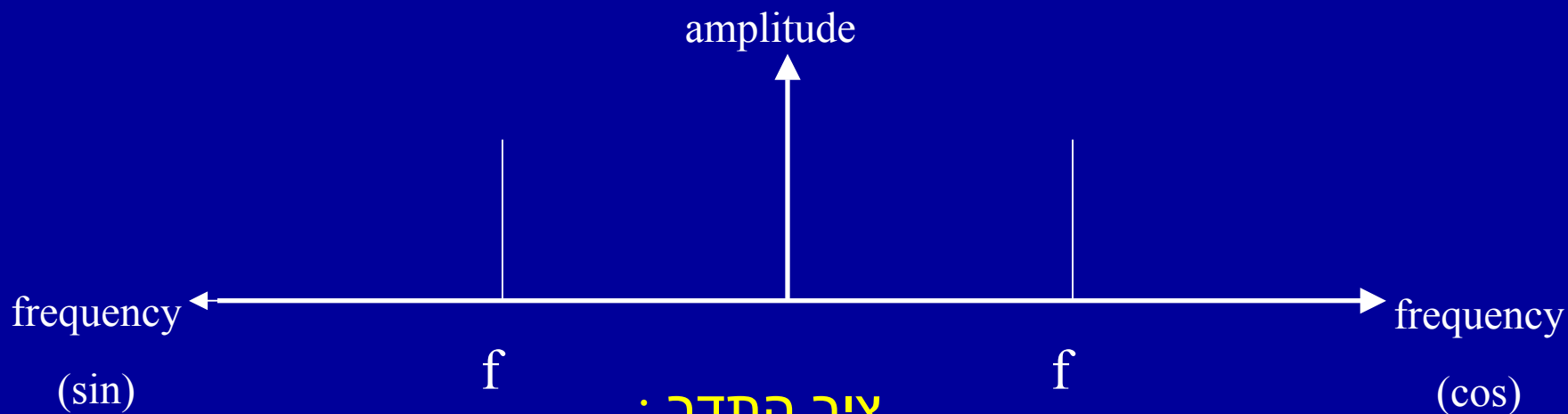
דוגמ

א



ציר הזמן

$$t * f = 1$$



ציר התדר
spectrum

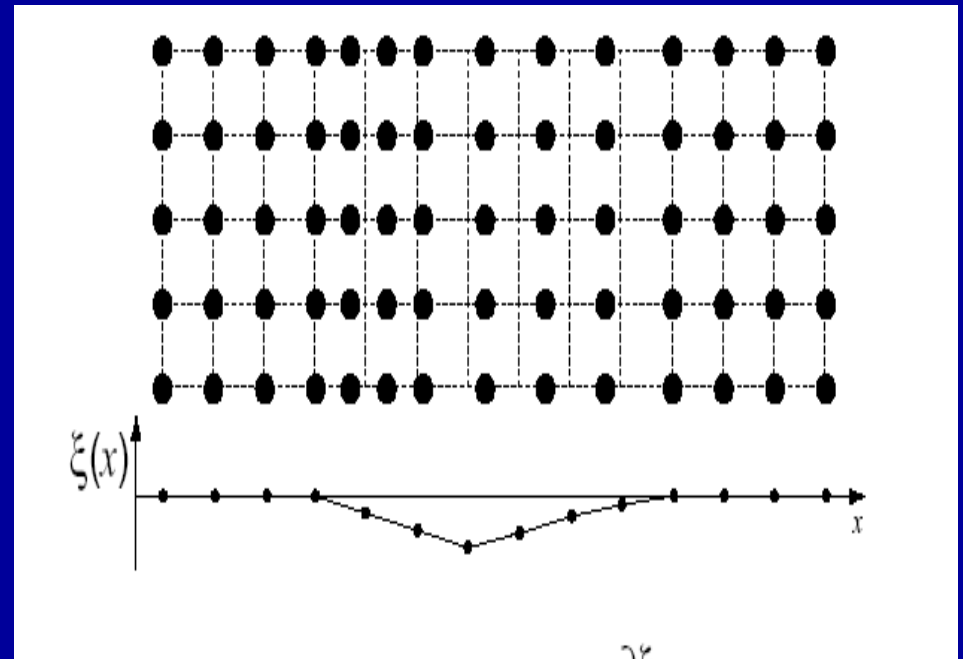
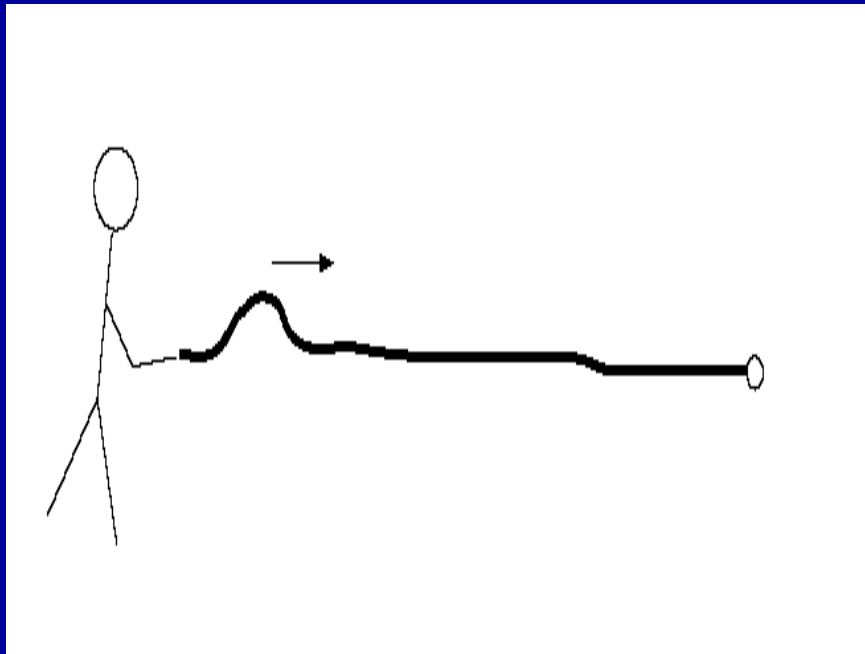
מבוא לאקוסטיקה

קול: תנודות פיזיות של מולקולות האוויר.

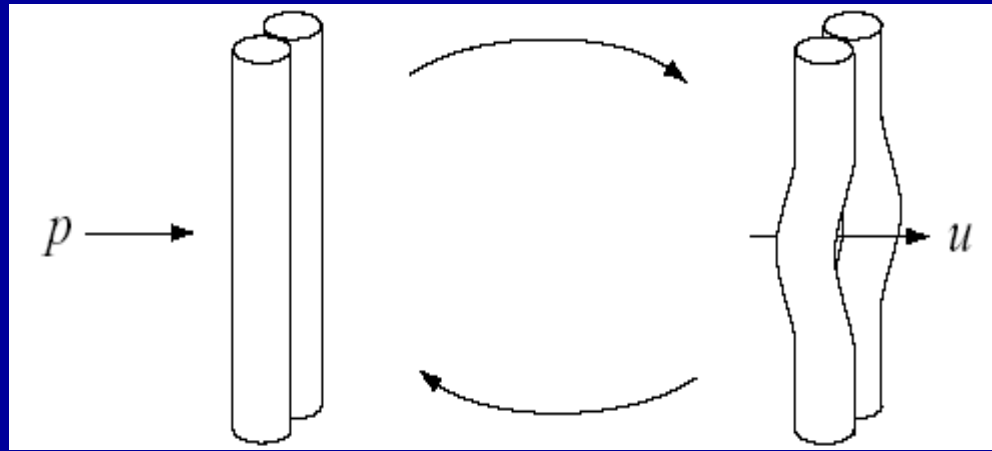
היוצרות קול: ע"י שינויי לחצים.

ביטוי פיזיקלי: לחץ כפונקציה של זמן.

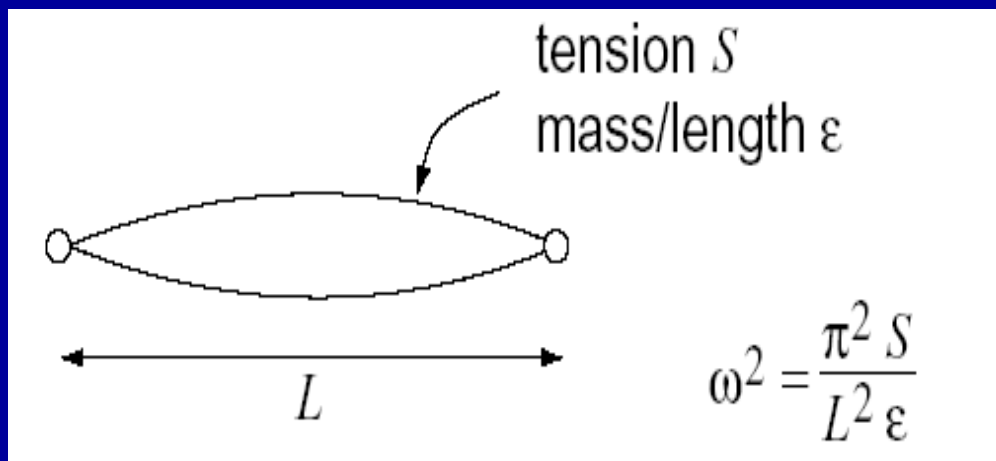
מושגים בסיסיים: מהירות הקול, אורך גל, $v = \lambda * f$



יצירת קול



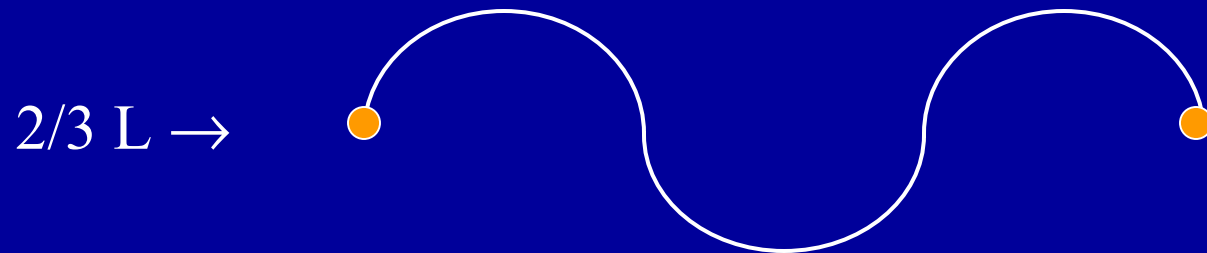
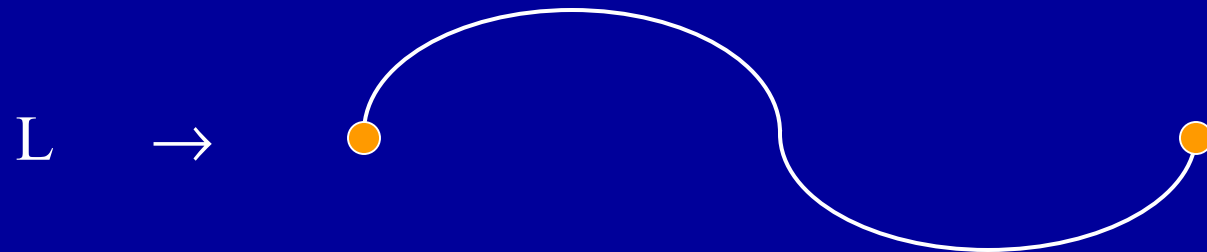
יצירת קול ע"י
מיתרי הקול



יצירת קול ע"י
מיתר:
תדרי תהודה:
תדר בסיסי
והרמוניות

יצירת קול במיתר

במיתר באורך L יכולים להיווצר גלים באורכי הגל הבאים:



ובאופן כללי:

$$\lambda_i = 2L/i \quad i=1,2,3,\dots$$

$$F_i = i * F_1$$

יצירת קול במיתר - המשך

מסקנה:

במיתר נוצר תדר בסיסי ("גובה" הצליל) והרמוניות שלו (כפולות שלמות בתדר).

התדר הבסיסי נקרא גם pitch.

מעטפת ספקטרלית

העוצמה היחסית של התדרים השונים מושפעת מגורמים שונים, והיא קובעת



כלי מיתר:

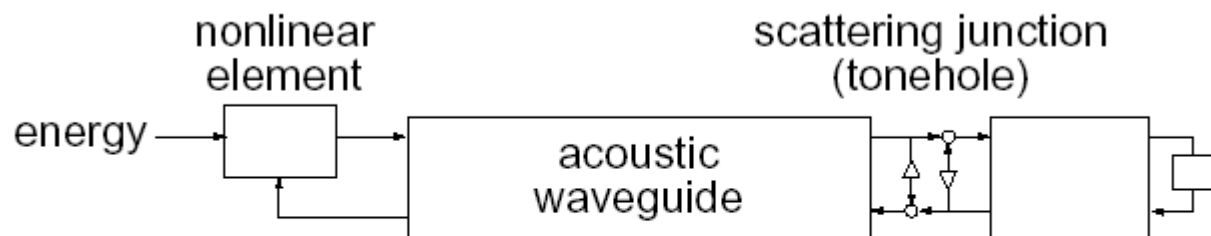
שליטה על גובה ואופי הצליל:

- אורך המיתר
- מתיחות המיתר
- מאסת המיתר
- תיבת התהודה

כלי נשיפה:

שליטה על גובה הצליל:

- האורך האפקטיבי של החלל בין הפה למקום יציאת האוויר.



$$\omega = \frac{\pi c}{2L} \quad (\text{quarter wavelength})$$

קליטת קול ע"י המחשב

אמצעי קלט למחשב:

מיקרופון הממיר את הקול לאותות חשמליים,

ורכיב A-TO-D הממיר את האות החשמלי לאות דיגיטלי.

איכות המיקרופון משפיעה על העיוותים שנגרמים לקול.

סוגי העיוותים:

- רעש אדטיבי
- הנחתת תדרים שונים (תגובת תדר).
- עיוותים לא לינארים.

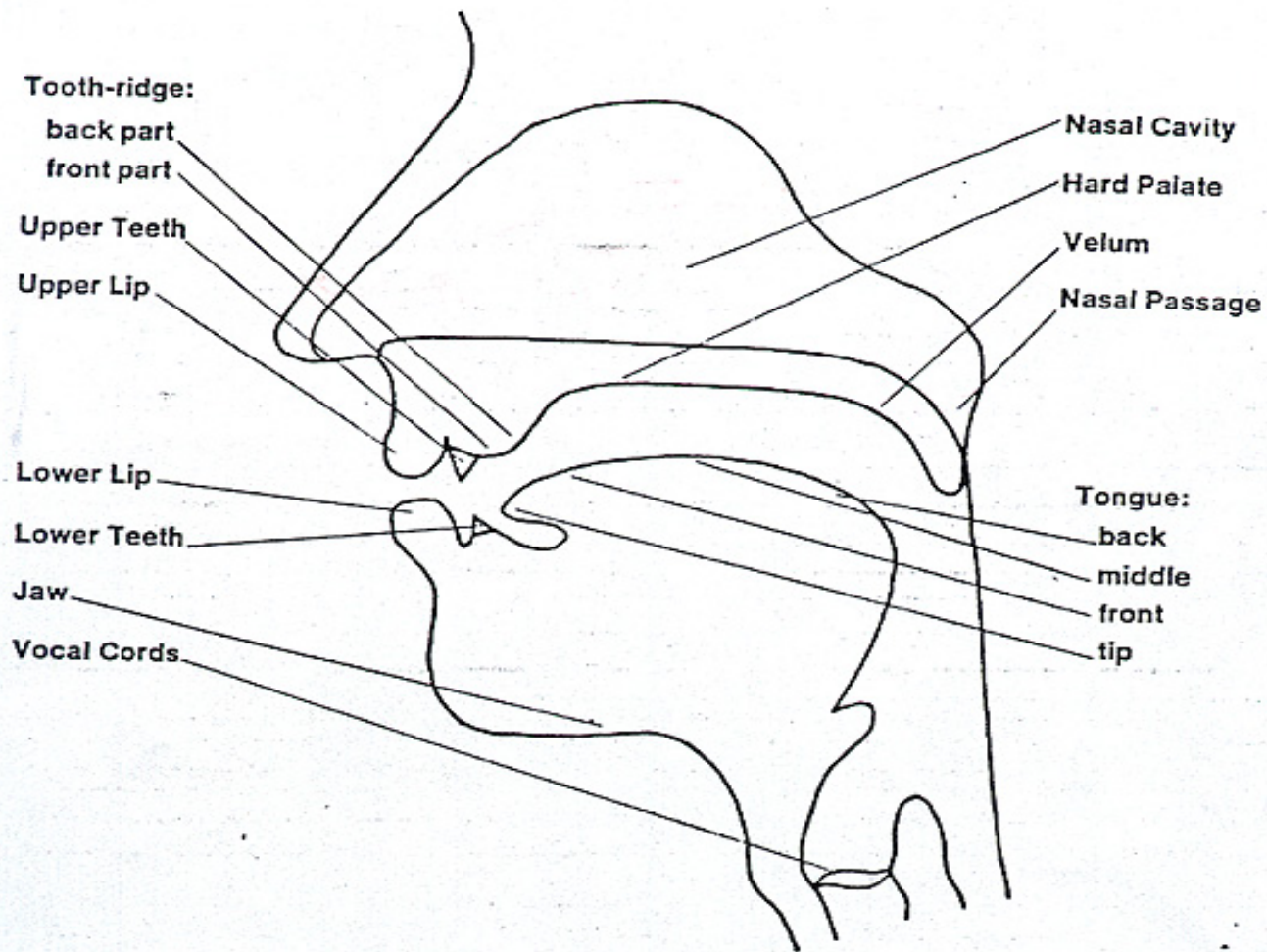
העברת קול בקו טלפון

איכות הקו משפיעה על העיוותים שנגרמים לקול.

סוגי העיוותים:

- רעש אדטיבי
- העברת תחום תדרים צר יחסית: 300-3300 הרץ.
- הנחתת תדרים שונים (תגובת תדר).
- עיוותים לא לינאריים.

The Organs of Speech



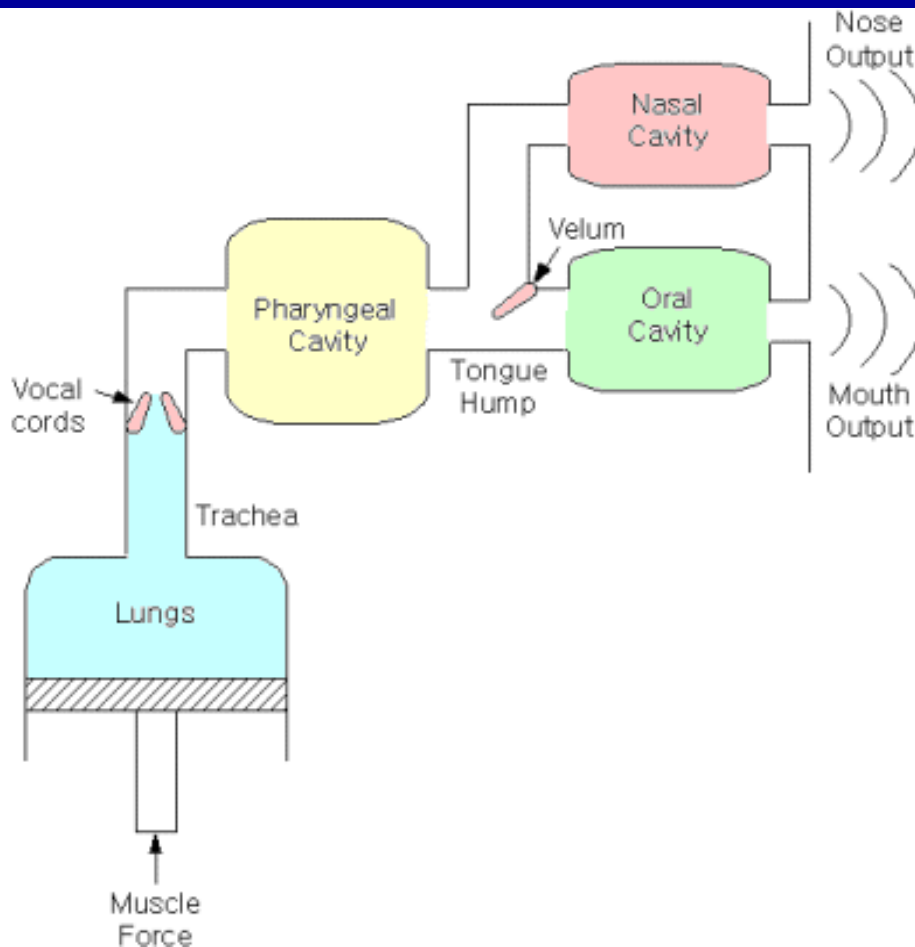
מודל היווצרות הדיבור

מוטיבציה:

הבנת התהליך הפיסיקלי של היווצרות הדיבור עשויה לתרום לפיתוח אלגוריתמיקה לזיהוי אוטומטי של דיבור.

מסלול היווצרות הדיבור:

1. אוויר ננשף מן הריאות.
2. מיתרי הקול מתנדדים (עבור פונמות קוליות בלתי
האוויר (והקול) =
3. תנודות פיזיות באוויר)
עובר דרך חלל הגרון והפ



מודל היווצרות הדיבור - המשך

4. היגוי (חיתוך) מתבצע באחד מן האיברים הבאים:

שפתיים

שיניים

לשון

חך

5. היגוי (חיתוך) עיצורים מתבצע באחד מן האופנים הבאים:

עצירה מלאה (stop) – פוצץ/סותם (p,t,k,b,d,g).

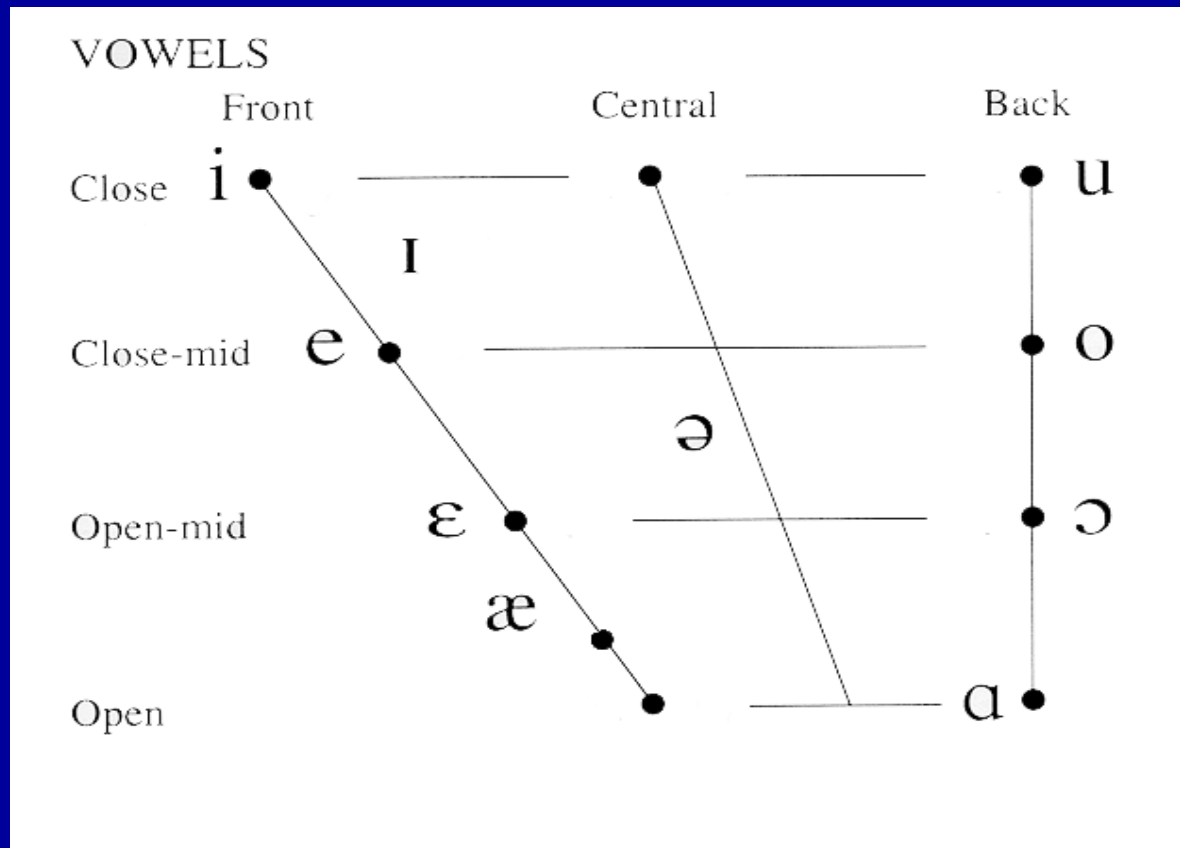
חוכך (fricative) – התקרבות בין שני איברי חיתוך אך לא עצירה מלאה (s,f).

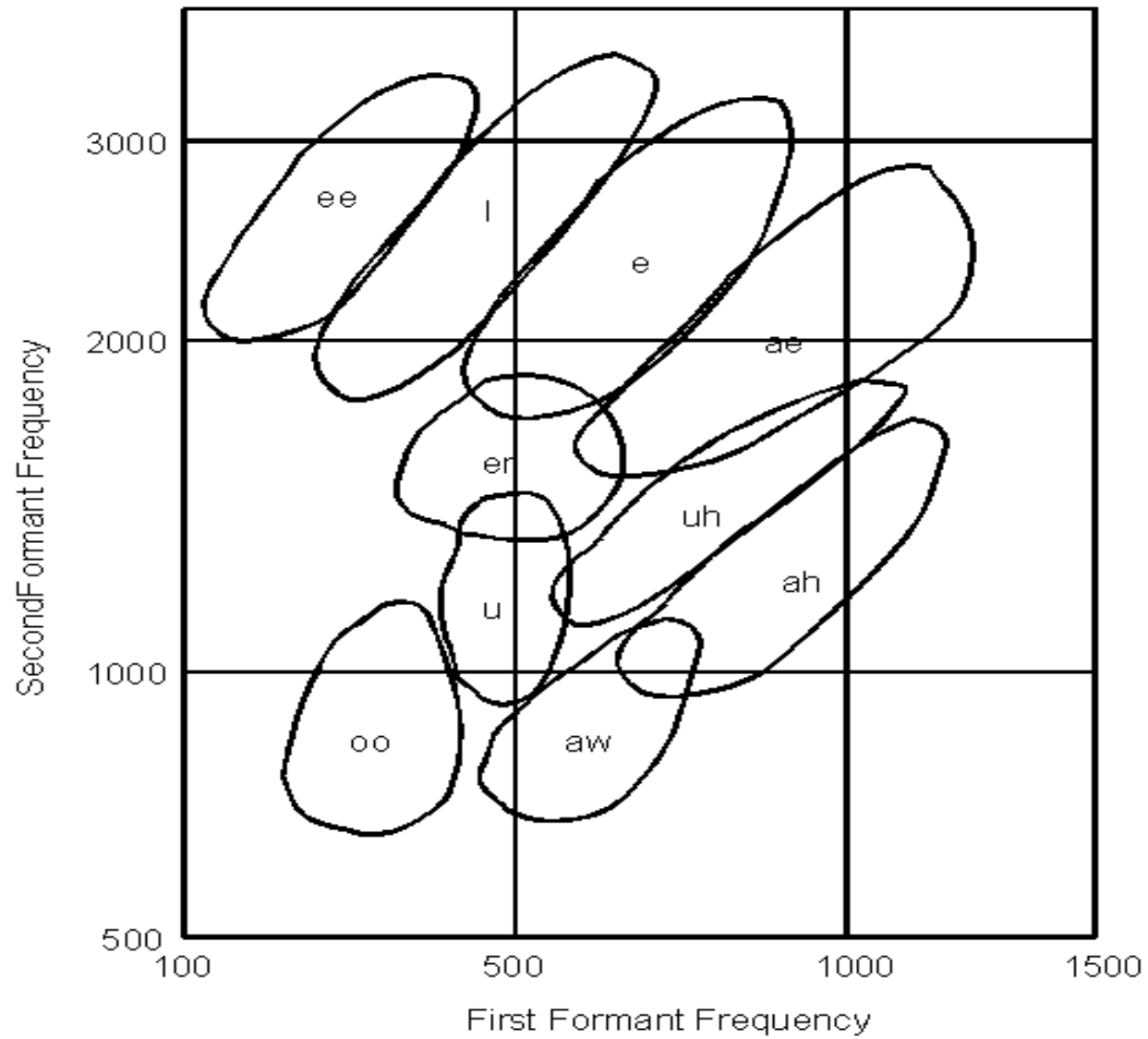
אפי (nasal) – מעבר אוויר דרך האף (m,n).

מודל היוצרות הדיבור - המשך

6. אופן היגוי תנועות:

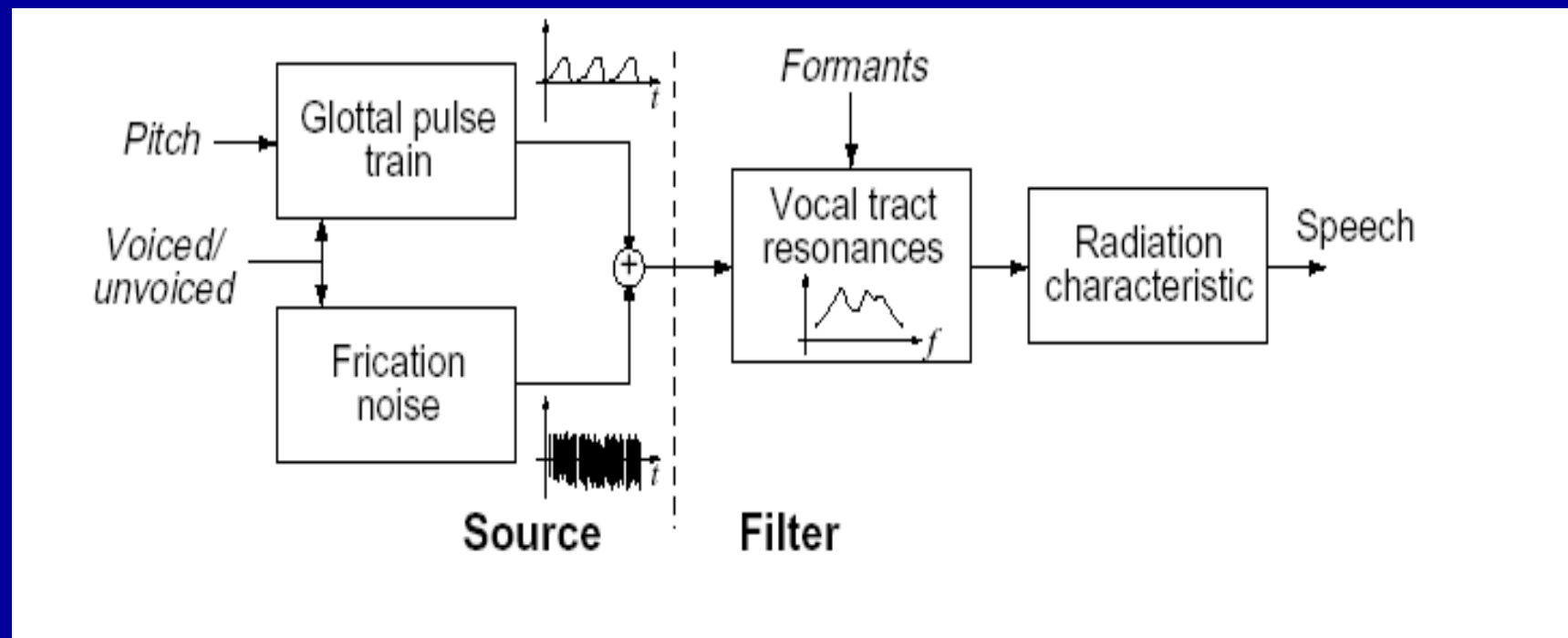
- גובה הלשון
- מיקום הלשון קדימה / אחורה
- עיגול השפתיים



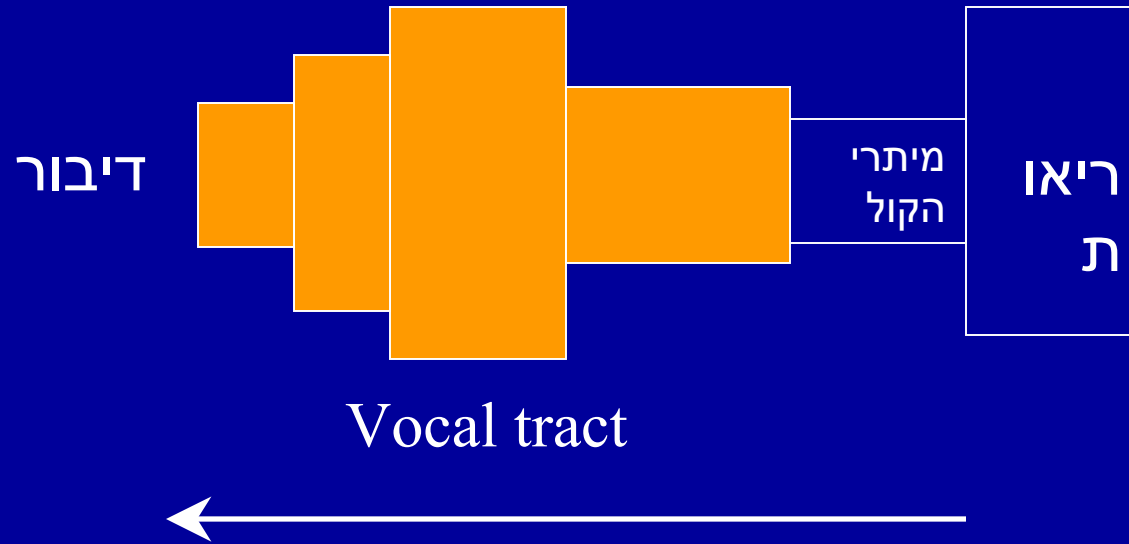


Second versus First Formant Frequency for Some Common Vowels

תיאור סכימטי של מנגנון ייצור הדיבור



מודל פיסיקלי להיווצרות הדיבור



רוחב ואורך הגלילים המרכיבים את ה-vocal tract ותכונת הקוליות (הנקבעת ע"י מיתרי הקול) קובעים את צורת אות הדיבור.

שינוי מיקום ההיגוי ואופן ההיגוי משפיע על צורת (ומספר) הגלילים המרכיבים את ה-vocal tract וכך אדם שולט על הצליל הנוצר, בדומה לנגן כלי נשיפה.

הצגה ויזואלית של קול

ראינו 2 דרכים לייצג דיבור:

- עוצמה כפונקציה של זמן.
- פירוק פורייה (ספקטרום).

נרצה לשלב את שתי השיטות:

לחשב ספקטרום אך לא לאבד את התלות בזמן.

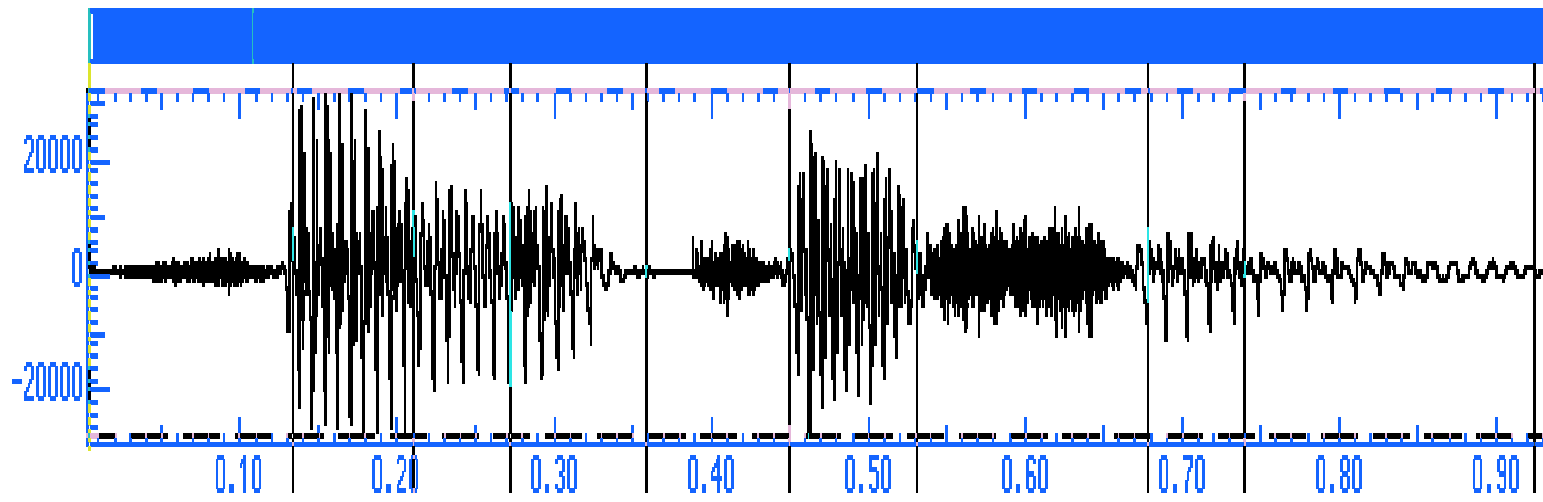
הפתרון: ספקטרוגרמה.

בהינתן אות דיבור ארוך, עבור כל זמן t נחשב ספקטרום עבור סידרת הדגימות $T, ..t+T$.

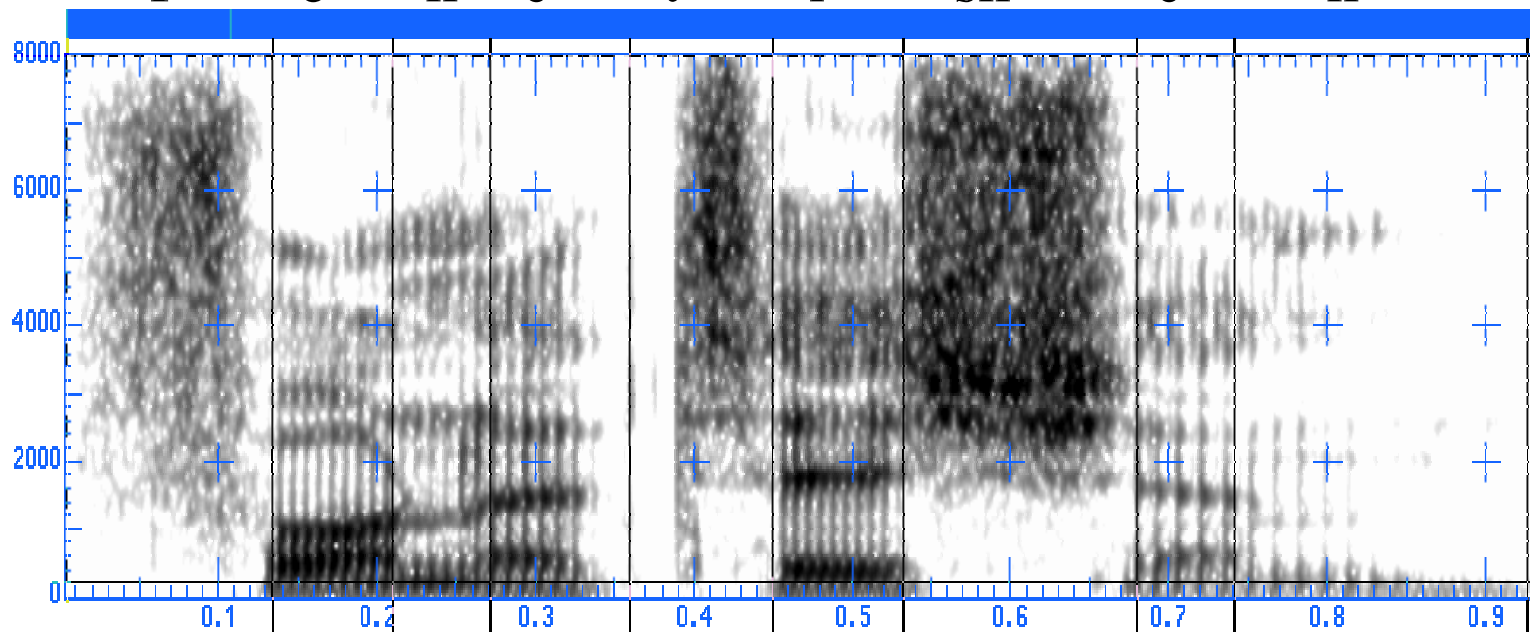
כיצד נציג זאת?

דרושים 3 מימדים:

ממד הזמן (ציר x), ממד התדר (ציר y), וממד העוצמה (צבע / דרגת אפור).



f o n e t i s h e n

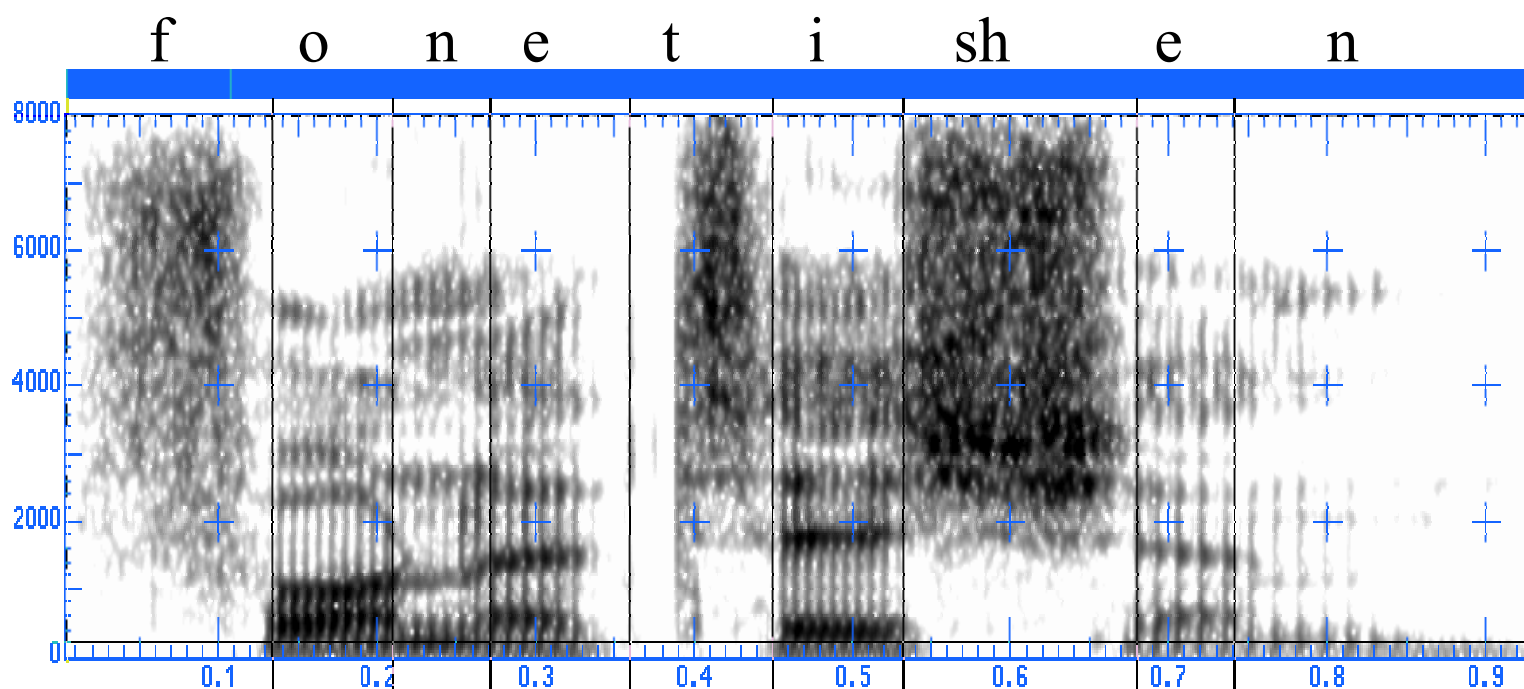


Unvoiced fricatives: אנרגיה (דמוית רעש) מרוכזת בתדרים הגבוהים.

Unvoiced plosives: שקט ואז פרץ אנרגיה בכל התדרים.

Voiced sounds: שיאים ברורים בפורמנטים (בגלל ה-pitch).

מיקום הפורמנטים תלוי בפונמה.



לסיכום

מושגים שלמדנו:

ספקטרום, ספקטרוגרמה.

Pitch, formants, spectral envelope.

ראינו שניתן להבחין (בצורה סבירה) בין פונמות שונות ע"י ניתוח
ספקטרוגרמות

(רק בלשנים מאומנים).

ניסיונות לחקות את אופן קריאת הספקטרוגרמות של בלשנים לא הצליחו

עד היום.

מודל השמיעה האנושי