הפקולטה להנדסה

המעבדה לעיבוד אותות

סינון רעשים על ידי בחירת המיקרופונים היעילים ביותר ברשת

אביה חבה

יאיר וולטרס

פרויקט שנה ד' לקראת תואר ראשון בהנדסה

מנחה: ד"ר עופר שוורץ

מנחה אקדמי: פרופ' שרון גנות

אוקטובר 2024

תוכן עניינים

תוכן עניינים 2
תקציר 3
רקע תאורטי 5
הצגת הבעיה 12
שיטות לפתרון הבעיה 15
יצירת בסיס הנתונים לבחינת השיטות 23
מימוש האלגוריתמים 29
תוצאות הניסוי ומסקנות 34
רעיונות להמשך 46
תודות 47

תקציר

מערכי מיקרופונים כמו מכשירי שמיעה, תקשורת טלפונית, אפליקציות זיהוי דיבור ועוד משמשים למטרות רבות של עיבוד דיבור. שימוש במערכי מיקרופונים יכול להביא לביצועים טובים יותר מאשר עבודה עם מיקרופון בודד מכיוון שבמערכי מיקרופונים אות הדיבור נדגם גם בזמן וגם במרחב. (שימוש לדוגמא הוא ה- MVDR beamformer בו נעסוק בהמשך.)

למערכי מיקרופונים יש כמה מגבלות; בדרך כלל, למערכי מיקרופונים יש יחידת עיבוד מרכזית אחת שמתחברת פיזית למיקרופונים. לכן, לא ניתן לשנות את מיקום המיקרופונים במערך ולא ניתן לעבוד עם מערכת כזאת באופן דינאמי. בנוסף, יתכן והדובר ממוקם רחוק מהמערך, וכתוצאה מכך נקבל יחס האות לרעש (SNR) נמוך יחסית.

על מנת לפתור את הבעיות האלה, פיתחו בשנים האחרונות רשתות חיישנים אקוסטיים אלחוטיים (WASN) בעלי פיזור רחב. ב-WASN, המיקרופונים מתקשרים עם המיקרופונים השכנים שלהם ועם יחידת העיבוד באמצעות חיבור אלחוטי. השימוש ב-WASN יכול לפתור את הבעיות שהזכרנו לגבי המערכים הלא - אלחוטיים.

אחת מהבעיות המרכזיות ב-WASN היא צריכת האנרגיה שלה.
שידור נתונים בין מיקרופון אחד לשני ובין המיקרופונים ליחידת העיבוד משפיע על חיי הסוללה של המיקרופון. למרות שכל המיקרופונים ב-WASN יתרמו לשיפור ה-SNR, לא כולם יתרמו לכך באופן משמעותי. יש לקחת בחשבון האם

השימוש בכל אחד מהמיקרופונים ברשת מועיל אל מול המחיר של ההפעלה שלו במובן של צריכת האנרגיה ומשך החיים שלו.

בפרויקט זה אנו נעסוק בשיטות מבוססות MVDR לבחירת תת-קבוצה של מיקרופונים בעלת צריכת ההספק הנמוכה ביותר, תחת האילוץ של שמירה על רמת ביצועים (במובן של SNR) מספקת.

אנחנו נתבסס בעיקר על שיטות שמבואות במאמר הבא:

"Microphone Subset Selection for MVDR Beamformer Based Noise Reduction"

By Jie Zhang, Sundeep Prabhakar Chepuri, Richard C. Hendriks, and Richard Heusdens

ונציע דרכים לשפר שיטות אלו. לאחר שנציג את הפיתוח המתמטי והאלגוריתמי של השיטות, נסביר על האופן שבו השוונו בין ביצועי השיטות הן בסימולציה והן בניסוי במעבדה.

רקע תאורטי

<u>מודל האות</u>:

נתאר את המערכת שלנו כרשת של M מיקרופונים המונחים במרחב. המיקרופונים קולטים את אותות הקול ומעבירים אותם ל-FC- fusion center (יחידת העיבוד המרכזית). שיטות הפחתת הרעש שאנחנו מתבססים עליהן מבוצעות במישור ה-*STFT* ולכן נסמן פריים של זמן ב-l ופס תדר נסמן ב-ω.

תחת ההנחה שישנו אות דיבור רצוי יחיד ומקורות רעש רבים במרחב, התמרת ה-*STFT* של האות הרועש שמתקבל במיקרופון ה-k, $y_k(\omega, l)$, מקיימת:

$$y_k(\omega, l) = x_k(\omega, l) + n_k(\omega, l)$$

כאשר $x_k(\omega, l)$ הוא אות הדיבור שנקלט במיקרופון ו $n_k(\omega, l)$ הוא הרעש המצטבר שנקלט במיקרופון. כמו כן, ניתן לומר ש-

$$x_k(\omega, l) = a_k(\omega) \cdot s(\omega, l)$$

כאשר $s(\omega, l)$ זה אות הדיבור המקורי בנקודת ייחוס מסוימת ו-$a_k(\omega)$ הוא פונקציית התמסורת האקוסטית (*ATF*) של האות הנקלט במיקרופון.

מעתה והלאה נשמיט את האינדקסים של הזמן והתדר מטעמי נוחות ונתאר את הבעיה בצורה וקטורית: וקטור האות הרועש של M המיקרופונים יסומן ב- $\boldsymbol{y} = [y_1, \dots, y_M]^H \in \mathbb{C}^M$, ו-, $\boldsymbol{n} = [n_1, \dots, n_M]^H \in \mathbb{C}^M$ יסמן את וקטור הרעשים של המיקרופונים. מתקבל כי

$$\boldsymbol{x} = \boldsymbol{a} \cdot s$$

כאשר $\boldsymbol{a} = [a_1, \dots, a_M]^H \in \mathbb{C}^M$ הוא וקטור ה- *ATF* של האות הנקלט במיקרופונים.

לפי ההצגה הוקטורית, ניתן לכתוב את מודל האות באופן הבא:

$$\boldsymbol{y} = \boldsymbol{x} + \boldsymbol{n}$$

בהנחה שאות הדיבור והרעש חסרי קורלציה ושתוחלת הרעש היא 0, מטריצת הקורלציה של $\boldsymbol{y}$ נתונה ע"י:

$$R_{yy} = E[\boldsymbol{y}\boldsymbol{y}^H] = E[\boldsymbol{x}\boldsymbol{x}^H] + E[\boldsymbol{n}\boldsymbol{n}^H] = R_{xx} + R_{nn} \in \mathbb{C}^{MxM} ,$$

כאשר R_{nn} מטריצת הקורלציה של הרעש ו-R_{xx} מטריצת הקורלציה של אות הדיבור שנקלט במיקרופונים. כמו כן,

$$R_{xx} = E[\boldsymbol{a}s^2\boldsymbol{a}^{\boldsymbol{H}}] = E[s^2] \cdot \boldsymbol{a}\boldsymbol{a}^{\boldsymbol{H}} = P_s\boldsymbol{a}\boldsymbol{a}^{\boldsymbol{H}} \in \mathbb{C}^{MxM}$$

כאשר P_s היא צפיפות ההספק הספקטרלי (*PSD*) של s.

MVDR beamformer:

Beamforming זוהי טכניקת עיבוד אותות שמשתמשים בה במערכי חיישנים לקליטה או לשידור של אותות כיווניים. הטכניקה מבוססת על הכפלה של האותות שמתקבלים במערך כך שאות המגיע מכיוון מסוים יעבור התאבכות בונה ואות המגיע מכיוון אחר יעבור התאבכות הורסת וכך ניתן לקלוט את האותות באופן שונה מנקודות שונות במרחב.

במקרה שלנו, ניתן להשתמש ב-*beamforming* על האותות הרועשים שמתקבלים במיקרופונים וליצור אלומה שתהיה שחזור של האות הרצוי בנקודת הייחוס ללא הרעש, ובכך לסנן את האות מהרעש.

בצורה מתמטית, אם נסמן את ה*beamformer* כ-w נקבל במוצא:

$$\tilde{s} = \boldsymbol{w}^{\boldsymbol{H}}\boldsymbol{y} = \boldsymbol{w}^{\boldsymbol{H}}\boldsymbol{x} + \boldsymbol{w}^{\boldsymbol{H}}\boldsymbol{n} = \boldsymbol{w}^{\boldsymbol{H}}\boldsymbol{a}s + \boldsymbol{w}^{\boldsymbol{H}}\boldsymbol{n}$$

נרצה שהאות הרצוי ישמר, כלומר שהמשערך יהיה חסר הטיה:

$$E[\tilde{s} - s] = 0 => E[\boldsymbol{w^H a} s + 0 - s] = 0 => \ \boldsymbol{w^H a} = 1$$

כמו כן, נרצה שעוצמת הרעש במוצא תהיה מינימלית, כלומר נרצה למצוא $\boldsymbol{w}$ שמביא למינימום את הביטוי הבא:

$$\boldsymbol{w^H} R_{nn} \boldsymbol{w}$$

ניתן להשתמש בכופלי לגראנז' כדי למצוא מינימום לפונקציה תחת אילוץ:

$$F = \ \boldsymbol{w^H} R_{nn} \boldsymbol{w} + \lambda^H (\boldsymbol{w^H a} - 1)^H + \ \lambda(\boldsymbol{w^H a} - 1)$$

ניתן לגזור את F לפי $\boldsymbol{w}$, להציב את האילוץ ולהגיע לפתרון.

דרך אחרת שבה ניתן להסתכל על זה היא לשים לב שמודל האות הוא מודל לינארי גאוסי:

$$\boldsymbol{y} = \boldsymbol{a} s + \boldsymbol{n}$$

כך ש $\boldsymbol{n} \sim \mathcal{N}(0, R_{nn})$, ו $\tilde{s} = \boldsymbol{w}^H \boldsymbol{y} \sim \mathcal{N}(\underbrace{\boldsymbol{w}^H \boldsymbol{a} s}_{estimator\ mean}, \underbrace{\boldsymbol{w}^H R_{nn} \boldsymbol{w}}_{estimstor\ variance})$

נשים לב שהאילוץ $\boldsymbol{w^H a} = 1$ שקול למציאת משערך חסר הטייה, כמו שהראנו מקודם. כמו כן, הדרישה להביא את $\boldsymbol{w^H} R_{nn} \boldsymbol{w}$ למינימום שקולה למציאת משערך בעל שונות מינימלית, כך שסך הכול ניתן לומר שמציאת ה- *MVDR beamformer* שקולה למציאת משערך חסר הטיה בעל שונות מינימלית לשערוך s, כלומר את משערך ה-*MVU* של s.

אנו יודעים שעבור המודל הלינארי הגאוסי, ניתן למצוא את משערך ה-*MVU* של s ע"י:

$$\tilde{s} = \underbrace{(\boldsymbol{a^H} R_{nn}^{-1} \boldsymbol{a})^{-1} \boldsymbol{a^H} R_{nn}^{-1}}_{\boldsymbol{w}^H} \boldsymbol{y}$$

(ההוכחה המלאה מובאת במצגת *Lecture 5: The Classical Linear Model* של הקורס "ניתוח סטטיסטי של מידע" של פרופ' אמיר לשם)

לפי זה,

$$\boldsymbol{w^H} = (\boldsymbol{a^H} R_{nn}^{-1} \boldsymbol{a})^{-1} \boldsymbol{a^H} R_{nn}^{-1} = \frac{\boldsymbol{a^H} R_{nn}^{-1}}{\boldsymbol{a^H} R_{nn}^{-1} \boldsymbol{a}}$$

$$\boldsymbol{w} = \frac{R_{nn}^{-1} \boldsymbol{a}}{\boldsymbol{a^H} R_{nn}^{-1} \boldsymbol{a}} \in \mathbb{C}^M$$

עתה, נחשב את ה-*SNR* במוצא:

$$SNR_{out} = \frac{E[\boldsymbol{w^H} R_{xx} \boldsymbol{w}]}{E[\boldsymbol{w^H} R_{nn} \boldsymbol{w}]} = \frac{\boldsymbol{a^H} R_{nn}^{-1} P_s \boldsymbol{a} \boldsymbol{a^H} R_{nn}^{-1} \boldsymbol{a}}{\boldsymbol{a^H} R_{nn}^{-1} R_{nn} R_{nn}^{-1} \boldsymbol{a}} = P_s \boldsymbol{a^H} R_{nn}^{-1} \boldsymbol{a} \frac{\boldsymbol{a^H} R_{nn}^{-1} \boldsymbol{a}}{\boldsymbol{a^H} R_{nn}^{-1} \boldsymbol{a}}$$

$$= P_s \boldsymbol{a^H} R_{nn}^{-1} \boldsymbol{a}$$

Beamformer זה נקרא *Minimum Variance Distortion Response* (*MVDR*).

השיטות שנציג בהמשך יהיו שיטות למציאת קבוצת מיקרופונים מתוך כלל המיקרופונים שה-*MVDR* שלהם מספק במוצא *SNR* רצוי.

אופטימיזציה קמורה, LP ו SDP:

בעיית אופטימיזציה קמורה כוללת מציאת ערך מינימום של פונקציה ואת ערך המשתנה שעבורו ערך הפונקציה מינימלית, בתוך תחום מוגדר (כלומר, תחת מערכת אילוצים) שבו היא קמורה (כלומר, בעלת נקודת מינימום גלובלית יחידה)

באופן כללי, ניתן להגדיר מתמטית בעיית אופטימיזציה קמורה כך:

$$\min_{\boldsymbol{x}} f(\boldsymbol{x})$$

$$subject\ to\ \boldsymbol{x} \in D$$

כאשר:

- $\boldsymbol{x} = [x_1, \dots, x_N]^T \in \mathbb{R}^n$ הוא משתנה האופטימיזציה
- $f(x): \mathbb{R}^n \to \mathbb{R}$ היא פונקציית ה-*objective*
- $D \subseteq \mathbb{R}^n$ הוא מרחב ערכי ה $\boldsymbol{x}$ האפשריים

בעיית *Linear Programing (LP)* היא בעיית אופטימיזציה קמורה שפונקציית ה-*objective* שלה וכן מערכת האילוצים תלויה לינארית במשתנה האופטימיזציה, כלומר בעיה מהצורה:

$$\min_{\boldsymbol{x}} \ \boldsymbol{c}^{\boldsymbol{T}}\boldsymbol{x}$$

$$s.t. \ A\boldsymbol{x} \leq \ \boldsymbol{b}$$

$$\boldsymbol{x} \geq \boldsymbol{0}$$

בזכות הלינאריות של בעיות אלו, סיבוכיות חישוב הפתרון שלהן נמוכה יחסית ותוכנת מחשב יכולה לפתור אותן במהירות ויעילות ביחס לבעיות אופטימיזציה קמורה שאינן לינאריות.

בעיית *semidefinite programming (SDP)* היא הכללה של בעיית *LP*.

זוהי בעיה שמשתנה האופטימיזציה שלה הוא <u>המטריצה</u> X, שפונקציית ה-*objective* שלה תלויה לינארית ב-X והאילוצים הינם *Linear Matrix Inequality* (*LMI*) של משתנה האופטימיזציה. בעיית *SDP* סטנדרטית היא מהצורה:

$$\min_{X \in R^{nxn}} Trace(CX)$$

$$s.t.\ Trace(A_i X) \leq\ b_i, \qquad i\ =\ 1, \dots, m,$$

$$X \geq 0.$$

ובדומה ל*LP* ניתן לפותרה ע"י תוכנת מחשב בצורה מהירה יעילה יחסית.

אם הבעיה אינה קמורה, או לחלופין הבעיה קמורה אבל אינה *SDP*, ניתן להפוך אותה לכזאת ע"י לקיחת הנחות – *Relaxations*. יש לקחת הנחות שמצד אחד יהפכו את הבעיה לפשוטה לפתרון ומצד שני עדיין יתארו את הבעיה המקורית בצורה מספיק טובה.

שיטות הפתרון הינן נומריות-איטרטיביות מכיוון שהפתרון האנליטי לבעיות הללו לפעמים הרבה פחות יעיל ולפעמים אינו קיים.

בהמשך נציג את הבעיה שדיברנו עליה בפתיחה בצורה מתמטית ונהפוך אותה ל-*SDP* תחת *Relaxations* מסוימות, כך שנוכל לפתור את הבעיה החדשה ביעילות בתוכנת מחשב (נצטרך להראות שהתוצאות מספיק מדויקות ולכן מערכת המשוואות עם ה-*Relaxations* עדיין מתארת את הבעיה בצורה טובה).

<u>Schur complement של מטריצה</u>:

בהינתן המטריצה $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \in \mathbb{C}^{(p+q)x(p+q)}$ כך ש-

$A \in C^{pxp}, B \in C^{pxq}, C \in C^{qxp}, D \in C^{qxq}$,

אם A מטריצה הפיכה אז ה-Schur complement של בלוק A של מטריצה M הוא:

$M/A = D - CA^{-1}B$.

כמו כן, אם D מטריצה הפיכה אז ה-Schur complement_ של בלוק D של מטריצה M הוא:

$M/D = A - BD^{-1}C$.

במקרה שמטריצה M סימטרית, כלומר $M = \begin{pmatrix} A & B \\ B^H & D \end{pmatrix} \in \mathbb{C}^{(p+q)x(p+q)}$, אם קיים M/A אז -

$M \succeq 0 \;\; iff \; M/A \succeq 0 \;\; AND \; A \succeq 0$.

אם קיים M/D אז-

$M \succeq 0 \;\; iff \; M/D \succeq 0 \;\; AND \; D \succeq 0$.

בשיטות שנביא בהמשך נשתמש ב-Schur complement ובתכונה שלו שהראנו כאן.

הצגת הבעיה

אנו רוצים לבחור את תת-קבוצת המיקרופונים שתביא את עלות העברת המידע בין המיקרופונים ל*FC* למינימום תחת האילוץ שעוצמת הרעש של האות המסונן תהיה נמוכה מערך מסוים.

נגדיר את וקטור הבחירה $\boldsymbol{p} = [p_1, \ldots, p_M]^T \in \{0,1\}^M$ כאשר המשמעות של $p_i = 1$ היא בחירה של המיקרופון ה-i והמשמעות של $p_i = 0$ היא שהמיקרופון ה-i אינו נבחר.

$K = \|\boldsymbol{p}\|_0$ הוא מספר המיקרופונים שנבחרו מתוך כל ה-M מיקרופונים (המשמעות של נורמה-0 היא מספר האיברים בוקטור שערכם אינו 0).

כעת נגדיר את מטריצת הבחירה $\Phi_p \in \{0,1\}^{KxM}$, שמתקבלת ע"י יצירת מטריצה אלכסונית $diag(\boldsymbol{p}) \in \{0,1\}^{MxM}$ שהאיברים על האלכסון שלה הם איברי הוקטור $\boldsymbol{p}$ והורדה ממטריצה זו את שורות האפסים שלה. נשים לב ש-

$$\Phi_p \Phi_p^T = I_K, \qquad \Phi_p^T \Phi_p = diag(\boldsymbol{p})$$

נשתמש בזהויות אלה בהמשך.

המשמעות של Φ_p היא שכאשר מכפילים את וקטור המדידות הרועשות $\boldsymbol{y}$ במטריצה זו, מקבלים את $\boldsymbol{y}_p$, וקטור מדידות של k המיקרופונים שנבחרו בלבד, כלומר:

$$\boldsymbol{y}_p = \Phi_p \boldsymbol{y} = \Phi_p \boldsymbol{x} + \Phi_p \boldsymbol{n} \in \mathbb{C}^K$$

נגדיר את וקטור ה-*ATF* של המיקרופונים הנבחרים $\boldsymbol{a}_p = \Phi_p \boldsymbol{a} \in \mathbb{C}^K$ ונגדיר את מטריצת הקורלציה של הרעש של המיקרופונים הנבחרים

$R_{nn,p} = \Phi_p R_{nn} \Phi_p^T \in \mathbb{C}^{KxK}$ (השורות והטורים ב- R_{nn} שמייצגים מיקרופונים שלא נבחרו הושמטו) ולכן ה- *MVDR beamformer* של

המיקרופונים הנבחרים יהיה:

$$\boldsymbol{w}_p = \frac{R_{nn,p}^{-1}\boldsymbol{a}_p}{\boldsymbol{a}_p^{\boldsymbol{H}} R_{nn,p}^{-1}\boldsymbol{a}_p} \in \mathbb{C}^K$$

אם נמצא וקטור בחירה $\boldsymbol{p}$ אופטימלי (בהתאם לדרישות שנציג בהמשך) אז נוכל לשערך את s ע"י הפעלת *MVDR* על וקטור המדידות של המיקרופונים שנבחרו, כלומר $\widetilde{s_p} = \boldsymbol{w}_p^{\boldsymbol{H}} \boldsymbol{y}_p$.

כעת נגדיר את $\boldsymbol{c} = [c_1, \dots, c_N]^T \in \mathbb{R}^n$ בתור וקטור המחיר של המיקרופונים, כך ש c_i הוא מחיר השימוש במיקרופון ה-i לטובת השיערוך.

נסמן את עוצמת הרעש במוצא ה-*beamformer* של כל M המיקרופונים ב-β ונגדיר קבוע נוסף $\alpha \in (0,1]$ שבעזרתו נגביל את הרעש במוצא ה-*beamformer* של ה-K מיקרופונים ביחס ל-β .

בצורה מתמטית נוכל לנסח את הבעיה כך:

$$\min_{\boldsymbol{w_p}, \boldsymbol{p} \in \{0,1\}^M} \|diag(\boldsymbol{p})\boldsymbol{c}\|_{\boldsymbol{1}}$$

$$s.t. \quad \boldsymbol{w_p^H} R_{nn,p} \boldsymbol{w_p} \leq \frac{\beta}{\alpha},$$

$$\boldsymbol{w_p^H} \boldsymbol{a_p} = 1$$

(P1)

הביטוי $\|diag(\boldsymbol{p})\boldsymbol{c}\|_{\mathbf{1}}$ מייצג את העלות הכוללת של הפעלת K המיקרופונים שנבחרו, האילוץ $\boldsymbol{w_p^H} R_{nn,p} \boldsymbol{w_p} \leq \frac{\beta}{\alpha}$ זוהי דרישה שעוצמת הרעש במוצא תיהיה נמוכה מ-$\frac{\beta}{\alpha}$, (לא נצפה שעוצמת הרעש תהיה נמוכה מ-β כי הרעש במוצא ה*beamformer* של K מיקרופונים בהכרח יהיה גבוה יותר מאשר הרעש במוצא ה-*beamformer* של כל המיקרופונים, לכן הוספנו את הקבוע- $\frac{1}{\alpha}$.), והאילוץ $\boldsymbol{w_p^H a_p} = 1$ זוהי דרישה שהאות במוצא לא יתעוות.

נשים לב שהתיאור המתמטי הנ"ל של הבעיה דומה לתיאור המתמטי של בעיית השערוך המרחבי הכללית שאותה הצגנו ברקע התאורטי. לכן המוטיבציה היא לפתח את הביטוי בעזרת מאפיינים של *MVDR*.

ראינו שעבור *MVDR* של כל המיקרופונים $SNR_{out} = P_s \boldsymbol{a^H} R_{nn}^{-1} \boldsymbol{a}$ ולכן עבור המיקרופונים שנבחרו בלבד במוצא ה-*MVDR* יתקבל

$$SNR_{out,\boldsymbol{p}} = P_s \boldsymbol{a_p^H} R_{nn,\boldsymbol{p}}^{-1} \boldsymbol{a_p}.$$

כעת אפשר לייצג את הבעיה הקודמת באופן אחר:

$$\min_{\boldsymbol{w_p}, \boldsymbol{p} \in \{0,1\}^M} \|diag(\boldsymbol{p})\boldsymbol{c}\|_{\mathbf{1}}$$

$$s.t. \quad P_s \boldsymbol{a_p^H} R_{nn,\boldsymbol{p}}^{-1} \boldsymbol{a_p} \geq \alpha * SNR_M = \frac{\alpha P_s}{\beta},$$

(P2)

החלפנו את שני האילוצים באילוץ יחיד על ה-*SNR*. נשים לב ששני הייצוגים של הבעיה הם שקולים מבחינה מתמטית אבל הם מנקודת מבט שונה:
בעוד הייצוג הראשון מאלץ את עוצמת הרעש להיות נמוכה מערך מסוים מבלי שהאות יתעוות, הייצוג השני מאלץ את ה*SNR* להיות גבוה מערך מסוים.

שיטות לפתרון הבעיה

לאחר שהצגנו את הבעיה באופן מתמטי, נפתור אותה בעזרת *convex relaxation*, כלומר ננסח את הבעיה מחדש כך שתהיה בעיית אופטימיזציה *SDP* תחת הקלות מסוימות שניקח על הבעיה המקורית, ואז יהיה ניתן לפותרה בעזרת תוכנת מחשב בצורה יעילה. נשים לב שלמרות שה-*objective* $\min_{\boldsymbol{w_p},\boldsymbol{p}\in\{0,1\}^M}\left|\left|diag(\boldsymbol{p})\boldsymbol{c}\right|\right|_{\boldsymbol{1}}$ לינארי ב-$\boldsymbol{p}$, האילוץ $P_s\boldsymbol{a_p^H}R_{nn,\boldsymbol{p}}^{-1}\boldsymbol{a_p} \geq \frac{\alpha P_s}{\beta}$ אינו לינארי ב-$\boldsymbol{p}$ ולכן הבעיה אינה בעיית *SDP*. בנוסף, הבעיה אינה קמורה כי $\boldsymbol{p}$ אינו רציף אך לזה נציע פתרון בהמשך.

כעת נציג שתי שיטות שונות לתאר את הבעיה כ*SDP*.
השיטה הראשונה מתבססת על ידיעה של מטריצת הקורלציה של אות הדיבור R_{xx}. השיטה השנייה מתבססת על ידיעה של וקטור מקדמי ה-*ATF* $\boldsymbol{a}$.
גם R_{xx} וגם $\boldsymbol{a}$ אינם ידועים לרוב, אבל בהמשך נציג שיטות לשערך אותם.

1. <u>*convex relaxation* בהתבסס על R_{xx}:</u>

הביטוי באילוץ $SNR_{out,p} = P_s\boldsymbol{a_p^H}R_{nn,\boldsymbol{p}}^{-1}\boldsymbol{a_p}$ מכיל כמה ביטויים של $\boldsymbol{p}$. נגדיר את $Q = \Phi_p^T R_{nn,\boldsymbol{p}}^{-1}\Phi_p$ כך שיכיל את כל הביטויים של $\boldsymbol{p}$ ולכן $P_s\boldsymbol{a_p^H}R_{nn,\boldsymbol{p}}^{-1}\boldsymbol{a_p} = P_s\boldsymbol{a^H}Q\boldsymbol{a^H}$. כעת נרצה לסדר Q כך ש-$\boldsymbol{p}$ יופיע בו פעם אחת בלבד. כדי לעשות זאת, נפרק את R_{nn} באופן הבא:

$$R_{nn} = G + \lambda I_M$$

כאשר G מטריצה חיובית בהחלט ו-λ סקלר חיובי. מכיוון ש-R_{nn} היא מטריצת קורלציה היא מטריצה חיובית מוחלטת ולכן כל עוד נבחר λ קטן מהערך עצמי הכי נמוך של R_{nn} נקבל G מטריצה חיובית בהחלט (עבור λ גדול מהע"ע הכי קטן של R_{nn} נקבל שלמטריצה $G = R_{nn} - \lambda I$ יש ע"ע שלילי ולכן G לא תהיה מטריצה חיובית).

לכן,

$$R_{nn,p} = \Phi_p\,(\lambda I_M + G)\Phi_p^T = \lambda I_K + \Phi_p\,G\Phi_p^T$$

ולכן

$$Q = \Phi_p^T\left(\lambda I_K + \Phi_p\,G\Phi_p^T\right)^{-1}\Phi_p$$

נשתמש בלמת היפוך המטריצה:

$$C(B^{-1} + C^T A^{-1} C)^{-1} C^T = A - A(A + CBC^T)^{-1}\,A$$

נציב $A = G^{-1}, B = \frac{1}{\lambda} I_K, C = \Phi_p^T$ ונקבל:

$$Q = \Phi_p^T\left(\lambda I_K + \Phi_p\,G\Phi_p^T\right)^{-1}\Phi_p = G^{-1} - G^{-1}\left(G^{-1} + \frac{1}{\lambda}diag(\boldsymbol{p})\right)^{-1} G^{-1}$$

אומנם פישטנו את Q כך ש-$\boldsymbol{p}$ מופיע בו רק פעם אחת, אבל Q עדיין לא תלוי לינארית ב-$\boldsymbol{p}$ (בגלל סימן ההיפוך).

כעת נחזור לביטוי באילוץ:

$$SNR_{out,p} = trace\left(SNR_{out,p}\right) = trace(P_s \boldsymbol{a}^{\boldsymbol{H}} Q \boldsymbol{a}^{\boldsymbol{H}}) = trace(QR_{xx}) = trace(R_{xx}^{\frac{H}{2}} Q R_{xx}^{\frac{1}{2}})$$

המעברים מבוססים על כך שב$trace$ אפשר להחליף את סדר ההכפלה של המטריצות. המוטיבציה במעברים האלו היא שבהמשך נקבל *LMI* של מטריצה סימטרית.

ננסח מחדש את האילוץ של (P2) בעזרת מטריצה חדשה Z :

$$P_s \boldsymbol{a_p^H} R_{nn,\boldsymbol{p}}^{-1} \boldsymbol{a_p} \geq \frac{\alpha P_s}{\beta} \Rightarrow P_s \boldsymbol{a_p^H} R_{nn,\boldsymbol{p}}^{-1} \boldsymbol{a_p} - \frac{\alpha P_s}{\beta} \geq 0$$

$$\Rightarrow trace\left(P_s \boldsymbol{a_p^H} R_{nn,\boldsymbol{p}}^{-1} \boldsymbol{a_p} - \frac{\alpha P_s}{\beta}\right) \geq 0$$

$$\Rightarrow trace\left(R_{xx}^{\frac{H}{2}} Q R_{xx}^{\frac{1}{2}} - \frac{\alpha P_s}{\beta M} I_M\right) \geq 0$$

$$\Rightarrow (i)\ trace\left(Z - \frac{\alpha P_s}{\beta M} I_M\right) \geq 0$$

$$(ii)\ Z = R_{xx}^{\frac{H}{2}} Q R_{xx}^{\frac{1}{2}}$$

החלפנו את האילוץ של (P2) בשני אילוצים חדשים ונוסף לנו משתנה אופטימיזציה חדש Z.

נשים לב שהאילוץ הראשון תלוי לינארית ב-Z אבל האילוץ השני תלוי באופן לא לינארי ב-$\boldsymbol{p}$.

על מנת להפוך אילוץ זה ללינארי נעשה עליו relaxation באופן הבא:

$$(ii)\ R_{xx}^{\frac{H}{2}} Q R_{xx}^{\frac{1}{2}} \succcurlyeq Z$$

נבחין כי גם אחרי ה-relaxation שני האילוצים הם תנאים מספיקים בשביל לקבל את האילוץ המקורי של (P2) ולכן זוהי relaxation טובה.

נחזור ל- (ii):

$$R_{xx}^{\frac{H}{2}} Q R_{xx}^{\frac{1}{2}} \succcurlyeq Z \Rightarrow R_{xx}^{\frac{H}{2}} Q R_{xx}^{\frac{1}{2}} - Z \succcurlyeq 0_M$$

$$\Rightarrow R_{xx}^{\frac{H}{2}} \left(G^{-1} - G^{-1} \left(G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) \right)^{-1} G^{-1} \right) R_{xx}^{\frac{1}{2}} - Z \succcurlyeq 0_M$$

$$\Rightarrow R_{xx}^{\frac{H}{2}} G^{-1} R_{xx}^{\frac{1}{2}} - Z \succcurlyeq R_{xx}^{\frac{H}{2}} G^{-1} \left(G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) \right)^{-1} G^{-1} R_{xx}^{\frac{1}{2}}$$

$$\Rightarrow R_{xx}^{\frac{H}{2}} G^{-1} R_{xx}^{\frac{1}{2}} - Z - R_{xx}^{\frac{H}{2}} G^{-1} \left(G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) \right)^{-1} G^{-1} R_{xx}^{\frac{1}{2}} \succcurlyeq 0_M$$

לסיום, נשתמש ב- *Schur complement* :

נסמן -

$$A = G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p})$$

$$B = G^{-1} R_{xx}^{\frac{1}{2}}$$

$$C = R_{xx}^{\frac{H}{2}} G^{-1}$$

$$D = R_{xx}^{\frac{H}{2}} G^{-1} R_{xx}^{\frac{1}{2}} - Z$$

נקבל כי

$$J/A = D - CA^{-1}B$$

$$= R_{xx}^{\frac{H}{2}} G^{-1} R_{xx}^{\frac{1}{2}} - Z - R_{xx}^{\frac{H}{2}} G^{-1} \left(G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) \right)^{-1} G^{-1} R_{xx}^{\frac{1}{2}}$$

$$J = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) & G^{-1} R_{xx}^{\frac{1}{2}} \\ R_{xx}^{\frac{H}{2}} G^{-1} & R_{xx}^{\frac{H}{2}} G^{-1} R_{xx}^{\frac{1}{2}} - Z \end{pmatrix}$$

לפי המשפט שהבאנו ברקע התיאורטי, מכיוון ש-

$$A = G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) \succcurlyeq 0_M$$

(זה נכון משום שסכום שתי מטריצות חיוביות מוחלטות הוא מטריצה חיובית מוחלטת) ומכיוון ש $J/A \succcurlyeq 0_M$ כמו שהוכחנו אזי

$$J = \begin{pmatrix} G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) & G^{-1} R_{xx}^{\frac{1}{2}} \\ R_{xx}^{\frac{H}{2}} G^{-1} & R_{xx}^{\frac{H}{2}} G^{-1} R_{xx}^{\frac{1}{2}} - Z \end{pmatrix} \succcurlyeq 0_{2M}$$

סך הכול קיבלנו את הבעיה הבאה:

$$\min_{\boldsymbol{p} \in \{0,1\}^M, Z} \|diag(\boldsymbol{p})\boldsymbol{c}\|_{\mathbf{1}}$$

$$s.t. \quad trace\left(Z - \frac{\alpha P_s}{\beta M} I_M\right) \geq 0,$$

$$\begin{pmatrix} G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) & G^{-1} R_{xx}^{\frac{1}{2}} \\ R_{xx}^{\frac{H}{2}} G^{-1} & R_{xx}^{\frac{H}{2}} G^{-1} R_{xx}^{\frac{1}{2}} - Z \end{pmatrix} \succcurlyeq 0_{2M}$$

כעת נבצע relaxation נוספת ונניח ש$\boldsymbol{p}$ הוא וקטור ערכים <u>רציפים</u> בין $[0,1]$ ונקבל :

$$\min_{\boldsymbol{p}, Z} \|diag(\boldsymbol{p})\boldsymbol{c}\|_{\mathbf{1}}$$

$$s.t. \quad trace\left(Z - \frac{\alpha P_s}{\beta M} I_M\right) \geq 0,$$

$$\begin{pmatrix} G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) & G^{-1} R_{xx}^{\frac{1}{2}} \\ R_{xx}^{\frac{H}{2}} G^{-1} & R_{xx}^{\frac{H}{2}} G^{-1} R_{xx}^{\frac{1}{2}} - Z \end{pmatrix} \succcurlyeq 0_{2M,}$$

$$0 \leq p_i \leq 1, \qquad i = 1,2, \ldots\ldots, M$$

נשים לב שפונקציית ה-objective והאילוצים שקיבלנו הינם לינאריים במשתני האופטימיזציה $\boldsymbol{p}, Z$ ולכן הבעיה היא SDP וניתן לפותרה ביעילות ובמהירות בעזרת תוכנות מחשב. במאמר נקרא לשיטה זו שיטה (1).

בהמשך נראה שיטות כיצד ניתן לתרגם חזרה את וקטור הערכים הרציפים $\boldsymbol{p}$ לווקטור ערכים בוליאניים שמייצג את ההחלטה אילו מיקרופונים להדליק ואילו לכבות.

2. <u>*convex relaxation* בהתבסס על $\boldsymbol{a}$:</u>

ניתן לכתוב את האילוץ של *(P2)* באופן הבא:

$$SNR_{out,p} = P_s \boldsymbol{a}^H Q \boldsymbol{a} \geq \frac{\alpha P_s}{\beta} \Rightarrow \boldsymbol{a}^H Q \boldsymbol{a} \geq \frac{\alpha}{\beta}$$

$$\Rightarrow \boldsymbol{a}^H \left(G^{-1} - G^{-1} \left(G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) \right)^{-1} G^{-1} \right) \boldsymbol{a} \geq \frac{\alpha}{\beta}$$

$$\Rightarrow \boldsymbol{a}^H G^{-1} \boldsymbol{a} - \boldsymbol{a}^H G^{-1} \left(G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) \right)^{-1} G^{-1} \boldsymbol{a} - \frac{\alpha}{\beta} \geq 0_M$$

כעת נשתמש ב- *Schur complement* :

נסמן -

$$A = G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p})$$

$$B = G^{-1} \boldsymbol{a}$$

$$C = \boldsymbol{a}^{\boldsymbol{H}} G^{-1}$$

$$D = \boldsymbol{a}^{\boldsymbol{H}} G^{-1} \boldsymbol{a} - \frac{\alpha}{\beta}$$

נקבל כי

$$J/A = D - CA^{-1}B$$

$$= \boldsymbol{a}^H G^{-1}\boldsymbol{a} - \boldsymbol{a}^H G^{-1}\left(G^{-1} + \frac{1}{\lambda}diag(\boldsymbol{p})\right)^{-1} G^{-1}\boldsymbol{a} - \frac{\alpha}{\beta}$$

$$J = \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} G^{-1} + \frac{1}{\lambda}diag(\boldsymbol{p}) & G^{-1}\boldsymbol{a} \\ \boldsymbol{a}^{\boldsymbol{H}} G^{-1} & \boldsymbol{a}^{\boldsymbol{H}} G^{-1}\boldsymbol{a} - \frac{\alpha}{\beta} \end{pmatrix}$$

לפי המשפט שהבאנו ברקע התיאורטי, מכיוון ש-

$$A = G^{-1} + \frac{1}{\lambda}diag(\boldsymbol{p}) \succcurlyeq 0_M$$

(זה נכון משום שסכום שתי מטריצות חיוביות מוחלטות הוא מטריצה חיובית מוחלטת) ומכיוון ש $J/A \succcurlyeq 0_M$ כמו שהוכחנו אזי

$$J = \begin{pmatrix} G^{-1} + \frac{1}{\lambda}diag(\boldsymbol{p}) & G^{-1}\boldsymbol{a} \\ \boldsymbol{a}^{\boldsymbol{H}} G^{-1} & \boldsymbol{a}^{\boldsymbol{H}} G^{-1}\boldsymbol{a} - \frac{\alpha}{\beta} \end{pmatrix} \succcurlyeq 0_{M+1}$$

סך הכול קיבלנו את הבעיה הבאה:

$$\min_{\boldsymbol{p}\in\{0,1\}^M} \|diag(\boldsymbol{p})\boldsymbol{c}\|_{\boldsymbol{1}}$$

$$s.t.$$

$$\begin{pmatrix} G^{-1} + \frac{1}{\lambda}diag(\boldsymbol{p}) & G^{-1}\boldsymbol{a} \\ \boldsymbol{a}^{\boldsymbol{H}} G^{-1} & \boldsymbol{a}^{\boldsymbol{H}} G^{-1}\boldsymbol{a} - \frac{\alpha}{\beta} \end{pmatrix} \succcurlyeq 0_{M+1}$$

כעת נבצע relaxation נוספת ונניח ש$\boldsymbol{p}$ הוא וקטור ערכים <u>רציפים</u> בין $[0,1]$ ונקבל :

$$\min_{\boldsymbol{p}} \|diag(\boldsymbol{p})\boldsymbol{c}\|_{\boldsymbol{1}}$$

$$s.t.$$

$$\begin{pmatrix} G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) & G^{-1}\boldsymbol{a} \\ \boldsymbol{a}^{\boldsymbol{H}} G^{-1} & \boldsymbol{a}^{\boldsymbol{H}} G^{-1}\boldsymbol{a} - \frac{\alpha}{\beta} \end{pmatrix} \succcurlyeq 0_{M+1}$$

$$0 \leq p_i \leq 1, \qquad i = 1,2, \ldots\ldots, M$$

נשים לב שפונקציית ה-objective והאילוצים שקיבלנו הינם לינאריים במשתנה האופטימיזציה $\boldsymbol{p}$ ולכן הבעיה היא SDP וניתן לפותרה ביעילות ובמהירות בעזרת תוכנות מחשב. במאמר נקרא לשיטה זו שיטה (2).

יצירת בסיס הנתונים לבחינת השיטות

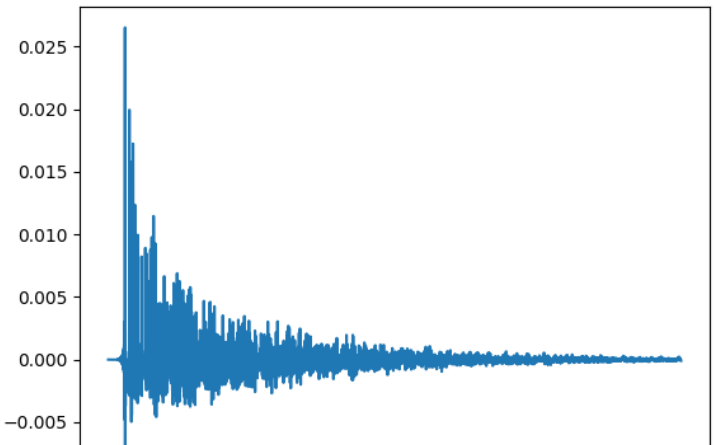
<u>סימולציה:</u>

לפני ההקלטות במעבדה היה חשוב לנו לבדוק מה הם התנאים המתאימים ביותר לסביבת הניסוי. לשם כך, ביצענו סימולציה שמבוססת על RIR (room impulse response) simulator שנכתב בPython.

הסימולטור מקבל את הפרמטרים הבאים:

- תדר הדגימה
- מיקום המיקרופון
- מיקום המקור
- ממדי החדר
- מקדם ההידוד (בשניות)
- מספר הדגימות של התגובה להלם

הסימולטור מחזיר את התוגבה להלם המתאימה כך שהוא מתחשב גם בהדהודים (עד כדי מספר הדגימות הנתון של התגובה להלם).



כתבנו תוכנית שניתן להגדיר בה את מספר המיקרופונים ואת מיקומיהם וכן את מיקום אות הדיבור ומיקום אותות הרעש והתוכנית משתמשת בסימולטור (ע"י לולאות) ויוצרת תגובות להלם עבור אותות שעוברים בכל המיקרופונים.

לאחר מכן, הורדנו הקלטות דיבור והקלטות רעש, טענו אותם בעזרת התוכנה וסיננו את האותות בתגובות ההלם ע"י קונבולוציה. סך הכול יצרנו M הקלטות (כמספר המיקרופונים) עבור כל אות דיבור ועבור כל אות רעש.

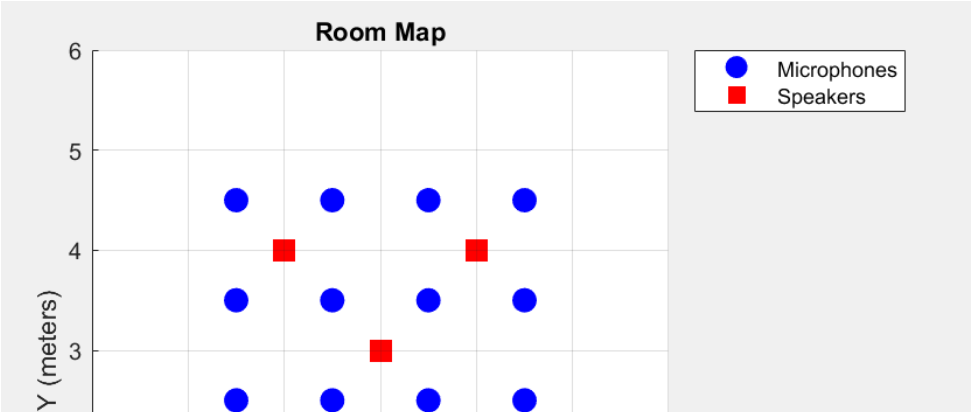
הרצנו סימולציות רבות כך שבין סימולציה לסימולציה שינינו את הפרמטרים הבאים:

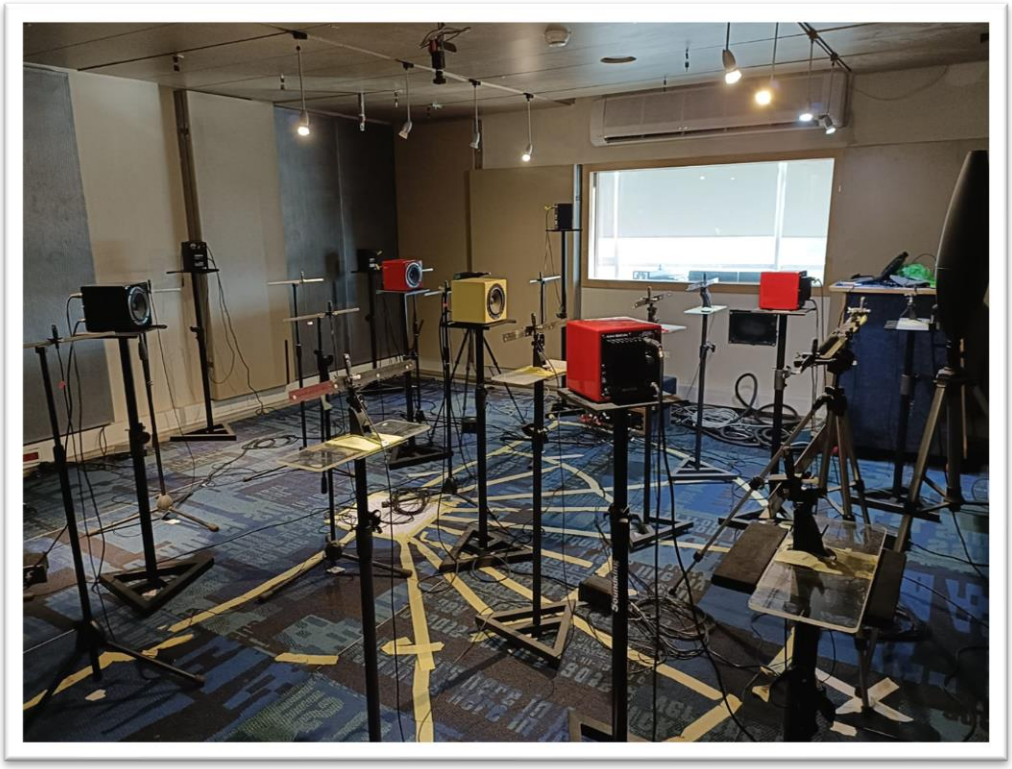
- מספר המיקרופונים
- מיקום המיקרופונים
- מספר מקורות הרעש
- מיקום מקורות הרעש
- מיקום הדובר
- הקלטות שונות של דוברים ושל מקורות רעש

מבין כלל הסימולציות בחרנו את התרחישים שעבורם תוצאות האלגוריתמים היו המעניינות ביותר להיות התרחישים שאותם נבדוק במעבדה.

<u>ניסוי במעבדה:</u>

בחדר בגודל של 6X6 מטר מיקמנו 16 מיקרופונים ו5 רמקולים באופן הבא:





מקדם ההדהוד של החדר הוא $0.4\ sec$.

כדי לאפשר בחירה של תרחישים רבים, הקלטנו בנפרד כל אחד מהאותות (דיבור או רעש) כאשר כל אחד מהאותות מושמע בכל אחד מ-5 הרמקולים ונקלט ב-16 המיקרופונים, כך שעבור כל אות נוצרות 5 הקלטות של 16 ערוצים.

אות הדיבור ששידרנו הוא הקלטה באורך של 2 דקות של דובר יחיד.

שידרנו 3 אותות רעש:

1. מזגן (רעש ורוד, כלומר קבוע בזמן ודועך בתדרים גבוהים).
2. מיקסר (רעש ורוד).
3. גשם (רעש צבעוני, כלומר משתנה בזמן ובתדר).

אורכה של כל אחת מההקלטות של אותות הרעש היא 4 דקות.

לאחר שיצרנו את ההקלטות כתבנו פונקציה ב-Matlab שמקבלת כפרמטרים הקלטת אות דיבור, הקלטות אותות רעש והגברים רצויים והפונקציה יוצרת הקלטה שהיא סכום של אות הדיבור ואותות הרעש עם ההגברים. בהקלטה שנוצרת, ב-2 הדקות הראשונות מוקלט הרעש בלבד ובשתי הדקות שאחריהן מוקלט הדיבור עם הרעש.
בצורת העבודה הזאת אפשר ליצור תרחישים בהם ניתן "למקם" את הדובר ואת מקורות הרעש בכל אחד ממיקומי 5 הרמקולים. כמו כן, ניתן להגביר, להנמיך ולבטל לגמרי את חלק מהרעשים או את כולם.
זה יתרון משמעותי כאשר אנחנו רוצים לבדוק שהאלגוריתמים עובדים בתרחישים שונים זה מזה ולבחון את השוני בהתנהגות האלגוריתמים בין מקרה למקרה.

<u>שערוך R_{xx}, R_{nn} ו-$\boldsymbol{a}$:</u>

האלגוריתמים שהצגנו דורשים ידיעה של R_{xx}, R_{nn} ו-$\boldsymbol{a}$. נפרט את השיטות בהם השתמשנו לשערך את הערכים האלו מתוך ההקלטות שנוצרו בסימולציה או ההקלטות שהוקלטו במעבדה.

על מנת לשערך את R_{nn} ביצענו התמרת STFT על הקטע בהקלטה שבו יש רעש בלבד (2 הדקות הראשונות) בכל אחת מM ההקלטות. עבור כל תדר ω, יצרנו מטריצה A_n בגודל M X L כך שהשורה מייצגת את אינדקס המיקרופון והעמודה מייצגת את האינדקס של פריים הזמן. ע"י הכפלה של A_n בהירמיטי שלה, נקבל $\widetilde{R_{nn}} = A_n A_n^H \in \mathbb{C}^{MXM}$.

בעזרת חישוב דומה על החלק בהקלטה שבו יש דיבור רועש (2 הדקות האחרונות בהקלטה) שיערכנו את R_{yy}.

כפי שהראנו ברקע התאורטי מתקיים הקשר:

$$R_{yy} = R_{xx} + R_{nn}$$

ולכן ניתן לשערך את R_{xx} ע"י:

$$\widetilde{R_{xx}} = \widetilde{R_{yy}} - \widetilde{R_{nn}}$$

על מנת לשערך את $\boldsymbol{a}$ ניסינו שתי שיטות:

1. ניתן לקבוע את מיקום אחד המיקרופונים כנקודת ייחוס (במיקום זה האות הוא s) ולמדוד את הפרשי הפאזות של האותות שעוברים בשאר המיקרופונים ביחס אליו באופן הבא:

נגדיר וקטור $\boldsymbol{d} = [d_1, \dots, d_M]^T \in \mathbb{C}^M$ כך ש-d_i הוא המרחק בין המיקרופון ה-i לבין הדובר. לצורך ההסבר נניח שמיקרופון הייחוס הוא המיקרופון ה-j. כעת כדי לקבל את הפרש הזמנים של הגעת האות לכל אחד מהמיקרופונים ביחס לזמן הגעת האות למיקרופון הייחוס j, נחשב:

$$\boldsymbol{t}_{delta} = \frac{\boldsymbol{d} - d_j}{v_{voice}}$$

כלומר נחסר את מרחק המיקרופון ה-j מכל איבר ב-$\boldsymbol{d}$ ונחלק במהירות האות, כלומר במהירות הקול. נשים לב שאם המרחק נמדד ביחידות של מטרים, והמהירות ביחידות של מטר לשנייה, אז $\boldsymbol{t}_{delta}$ יחושב ביחידות של שנייה.

לכן, כדי לקבל את הפרש הזמנים ביחידות של מספר דגימות, נכפיל בתדר הדגימה f_s, ונקבל $[\delta(n - \boldsymbol{t}_{delta,1} f_s), \dots, \delta(n - \boldsymbol{t}_{delta,M} f_s)]$.

נתמיר את הוקטור למישור התדר ונקבל:

$$\widetilde{\boldsymbol{a}} = \left[e^{2\pi f \boldsymbol{t}_{delta,1} f_s}, \dots, e^{2\pi f \boldsymbol{t}_{delta,M} f_s}\right]^T = \left[e^{\frac{2\pi\omega}{W} \boldsymbol{t}_{delta,1} f_s}, \dots, e^{\frac{2\pi\omega}{W} \boldsymbol{t}_{delta,M} f_s}\right]^T$$

כאשר ω הוא התדר ו-W הוא מספר פסי התדר.

2. מכיוון שאנחנו יודעים לשערך את R_{xx} ומכיוון שאנו יודעים ש $R_{xx} = P_s \boldsymbol{a}\boldsymbol{a}^{\boldsymbol{H}}$ ניתן לקבוע את המיקרופון הראשון כמיקרופון ייחוס ולכן האות במיקום שלו הוא s. ניתן לחשב את P_s ע"י $\sum_{l=0}^{L-1} s(l)^2$. נחלק כל אחד מאיברי R_{xx} ב- P_s שחישבנו ונקבל $B = \boldsymbol{a}\boldsymbol{a}^{\boldsymbol{H}}$.
כעת ניקח את הווקטור העצמי שמתאים לערך העצמי הגבוה ביותר של מטריצה B. לאחר שנחלק כל אחד מאיברי הווקטור הזה באיבר הראשון שלו (זהו נרמול על מנת שנקבל שהאיבר הראשון הוא 1) נקבל את ווקטור $\boldsymbol{a}$ המשוערך.
אומנם השיטה הזו מדויקת יותר מהשיטה הראשונה כי היא כוללת גם את ההדהודים אבל היא מסתמכת על ידיעת R_{xx} שזהו חיסרון משמעותי.

השיטה הראשונה יותר אינטואיטיבית וניתן להבין בעזרתה בצורה טובה יותר את המשמעות של מקדמי ה-ATF, וכן אפשר להשתמש בה ללא צורך בידיעה של R_{xx}. אבל, שיטה זו תהיה מדויקת רק כשאין הדהודים, כלומר במרחב פתוח ללא גופים שמחזירים את אותות הקול, ולכן היא פחות מדויקת במקרה שלנו כשמדובר בחדר סגור.
בניגוד לשיטה הראשונה, השיטה השנייה מתחשבת בהדהודים ולכן מדויקת יותר, אך היא מסתמכת על ידיעת R_{xx}. במקרה שלנו ניתן לשערך את R_{xx} בצורה טובה ולכן בחרנו להשתמש בשיטה זו.

נשים לב שכל הערכים ששערכנו תלויים ב- ω ומשוערכים עבור כל תדר בנפרד.

מימוש האלגוריתמים

על מנת לממש את האלגוריתם כתבנו קוד מטלב שמשתמש בחבילת CVX, זהו כלי שפותר בעיית אופטימיזציה קמורה.

את ה-SDP שמבוסס על R_{xx}:

$$\min_{\boldsymbol{p},Z} \|diag(\boldsymbol{p})\boldsymbol{c}\|_{\mathbf{1}}$$

$$s.t. \quad trace\left(Z - \frac{\alpha P_s}{\beta M} I_M\right) \geq 0,$$

$$\begin{pmatrix} G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) & G^{-1} R_{xx}^{\frac{1}{2}} \\ R_{xx}^{\frac{H}{2}} G^{-1} & R_{xx}^{\frac{H}{2}} G^{-1} R_{xx}^{\frac{1}{2}} - Z \end{pmatrix} \succcurlyeq 0_{2M,}$$

$$0 \leq p_i \leq 1, \qquad i = 1,2, \ldots\ldots, M$$

מימשנו באופן הבא:

```
cvx_begin
   variables p(M) Z(M,M)
   minimize(norm(diag(p)*c,1))
   subject to
      trace(Z - (alpha*P_s/(M*beta)).*eye(M)) >= 0;

      s_t_matirx_a = inv(G) + (1/lambda).*diag(p);
      s_t_matirx_b = G\sqrtm(Rxx) ;
      s_t_matirx_c = sqrtm(Rxx)/G ;
      s_t_matirx_d = (sqrtm(Rxx)/G)*sqrtm(Rxx) - Z ;
      s_t_matrix = [s_t_matirx_a s_t_matirx_b ; s_t_matirx_c s_t_matirx_d];
      s_t_matrix == hermitian_semidefinite(2*M);

            for i=1:M
                 0 <= p(i)<= 1;
```

את ה-SDP שמבוסס על $\boldsymbol{a}$:

$$\min_{\boldsymbol{p},Z} \|diag(\boldsymbol{p})\boldsymbol{c}\|_{\boldsymbol{1}}$$

$$s.t.$$

$$\begin{pmatrix} G^{-1} + \frac{1}{\lambda} diag(\boldsymbol{p}) & G^{-1}\boldsymbol{a} \\ \boldsymbol{a}^{\boldsymbol{H}} G^{-1} & \boldsymbol{a}^{\boldsymbol{H}} G^{-1}\boldsymbol{a} - \frac{\alpha}{\beta} \end{pmatrix} \succcurlyeq 0_{M+1}$$

$$0 \le p_i \le 1, \qquad i = 1,2, \ldots\ldots, M$$

מימשנו באופן הבא:

```
cvx_begin
   cvx_quiet true;
   variables p(M)
   minimize(norm(diag(p)*c,1))
   subject to
      s_t_matirx_a = inv(G) + (1/lambda).*diag(p);
      s_t_matirx_b = G\a ;
      s_t_matirx_c = a'/G ;
      s_t_matirx_d = (a'/G)*a - alpha/beta ;
      s_t_matrix = [s_t_matirx_a s_t_matirx_b ; s_t_matirx_c s_t_matirx_d];
      s_t_matrix == hermitian_semidefinite(M+1);

      for i=1:M
           0 <= p(i)<= 1;
      end
cvx_end
```

<u>קביעת הערכים הבוליאניים של $\boldsymbol{p}$</u>

נזכיר שהגדרנו בבעיה את $\boldsymbol{p}$ בתור משתנה אופטימיזציה רציף על מנת שהבעיה תהיה קמורה, אבל בפועל נרצה שהערכים של $\boldsymbol{p}$ יהיו בוליאניים על מנת לקבל החלטה עבור כל מיקרופון.

בחנו שתי שיטות לקביעת $\boldsymbol{p}_{boolean}$ מתוך $\boldsymbol{p}_{con}$:

1. *Randomize rounding*- נגריל בהתפלגות אחידה $[0,1]$ ווקטור בגודל M, נסמן אותו ב- $\boldsymbol{q}$.

   כעת,

$$\boldsymbol{p}_{candidate,i} = \begin{cases} 1, & \boldsymbol{p}_{con,i} > \boldsymbol{q}_i \\ 0, & \boldsymbol{p}_{con,i} < \boldsymbol{q}_i \end{cases}$$

   מבצעים הגרלה זו L פעמים. אם לא התקבל $\boldsymbol{p}_{candidate}$ שעבורו מתקבל $SNR_{\boldsymbol{p}_{\boldsymbol{candidate}}} > \alpha SNR$ מבצעים שוב L הגרלות.

   אחרת, בוחרים מתוך כל ה- $\boldsymbol{p}_{candidate}$ שמקיימים

$$SNR_{\boldsymbol{p}_{\boldsymbol{candidate}}} > \alpha SNR$$

   את ה- $\boldsymbol{p}_{candidate}$ שעבורו $\|diag(\boldsymbol{p}_{\boldsymbol{candidate}})\boldsymbol{c}\|_1$ מינימלי להיות $\boldsymbol{p}_{boolean}$.

   זו השיטה שבה כותבי המאמר השתמשו.

2. נאתחל את $\boldsymbol{p}_{candidate}$ להיות וקטור אפסים. נבדוק מהו האינדקס של האיבר המקסימלי ב- $\boldsymbol{p}_{con}$ ובאינדקס זה נשנה את הערך ב- $\boldsymbol{p}_{candidate}$ ל-1 וב- $\boldsymbol{p}_{con}$ נקבע ל-0.

אם מתקיים $SNR_{\boldsymbol{p}_{candidate}} > \alpha SNR$ אז $\boldsymbol{p}_{boolean} = \boldsymbol{p}_{candidate}$.
אחרת, נחזור על האלגוריתם.

בחרנו בשיטה מס' 2 בגלל כמה סיבות:

- שיטה זו אינה תלויה בסטטיסטיקה בניגוד לשיטה מס' 1
- מספר האיטרציות של שיטה זו חסום ב-M בניגוד לשיטה הראשונה בה מספר האיטרציות אינו חסום
- על מנת לקבל בשיטה מס' 1 דיוק זהה לשיטה מס' 2 יש צורך לבחור ערך L גבוה מאוד שפוגע ביעילות השיטה

<u>בחירת המיקרופונים על סמך כל התדרים</u>

למרות שעבור כל תדר עלול להתקבל $\boldsymbol{p}$ שונה (R_{xx}, R_{nn} ו-$\boldsymbol{a}$ תלויים בתדר), בסופו של דבר צריך לקבל החלטה יחידה עבור כל התדרים.

שיטה נאיבית תהיה לחשב את $\boldsymbol{p}$ עבור כל תדר בנפרד, למצע את ה-$\boldsymbol{p}$ שהתקבלו ולעגל, כלומר:

$$\boldsymbol{p}_{final} = round\left(\sum_{\omega=1}^{\omega=W} \frac{\boldsymbol{p}(\omega)}{W}\right)$$

החיסרון בשיטה הזו שהיא לא לוקחת בחשבון שישנם תדרים משמעותיים יותר שבהם נרצה לתת משקל גבוה יותר להחלטת האלגוריתם, וישנם תדרים פחות משמעותיים שבהם נרצה לתת משקל נמוך יותר.

בעצם, אנחנו מחפשים $c_1, \dots, c_W$ ונקבע את ההחלטה כך:

$$\boldsymbol{p}_{final} = round\left(\sum_{\omega=1}^{\omega=W} \frac{c_\omega \boldsymbol{p}(\omega\,)}{c_\omega}\right)$$

אנחנו בחרנו $c_\omega = trace(R_{xx}R_{nn})$, משום שהציפייה היא שהתדרים בהם עוצמת הרעש ועצמת אות הדיבור חזקים יותר הם תדרים משמעותיים יותר.

<u>ביצוע MVDR beamforming על האותות מהמיקרופונים הנבחרים</u>

לאחר שבחרנו את $\boldsymbol{p}_{final}$, נשערך את הערכים $R_{nn,\boldsymbol{p}_{final}}, R_{xx,\boldsymbol{p}_{final}}$ ו-$\boldsymbol{a}_{\boldsymbol{p}_{final}}$ כמו ששערכנו את R_{xx}, R_{nn} ו-$\boldsymbol{a}$ אך הפעם עבור המיקרופונים הנבחרים בלבד. נחשב גם את $\boldsymbol{y}_{\boldsymbol{p}_{final}}$ ע"י ביצוע התמרת *STFT* על ההקלטות הרועשות של המיקרופונים הרועשים.

כמו שראינו ברקע התאורטי לגבי *MVDR*,

$$\tilde{s}_{\boldsymbol{p}_{\boldsymbol{final}}} = \frac{\boldsymbol{a}^{\boldsymbol{H}}_{\boldsymbol{p}_{final}} R^{-1}_{\boldsymbol{nn},\boldsymbol{p}_{final}}}{\boldsymbol{a}^{\boldsymbol{H}}_{\boldsymbol{p}_{final}} R^{-1}_{\boldsymbol{nn},\boldsymbol{p}_{final}} \boldsymbol{a}_{\boldsymbol{p}_{final}}} \boldsymbol{y}_{\boldsymbol{p}_{final}}$$

ואת החישוב הזה יש לחשב עבור כל פריים של זמן l ועבור כל תדר ω. לאחר מכן נבצע התמרת *ISTFT* ונקבל שערוך של אות הקול המקורי במישור הזמן.

תוצאות הניסוי ומסקנות

על מנת לוודא שהתוצאות משמעותיות ולא מקריות, בדקנו את השיטות שהצגנו על תרחישים רבים ששניים מהם נציג כאן. עבור כל תרחיש ביצענו כמה מבחנים והשוואות. כמו כן, ביצענו את המבחנים ואת ההשוואות האלה על שיטה נוספת שתהווה כשיטת ייחוס וכך נבדוק כמה התוצאות של השיטות שלנו טובות ביחס לתוצאות שיטת הייחוס.

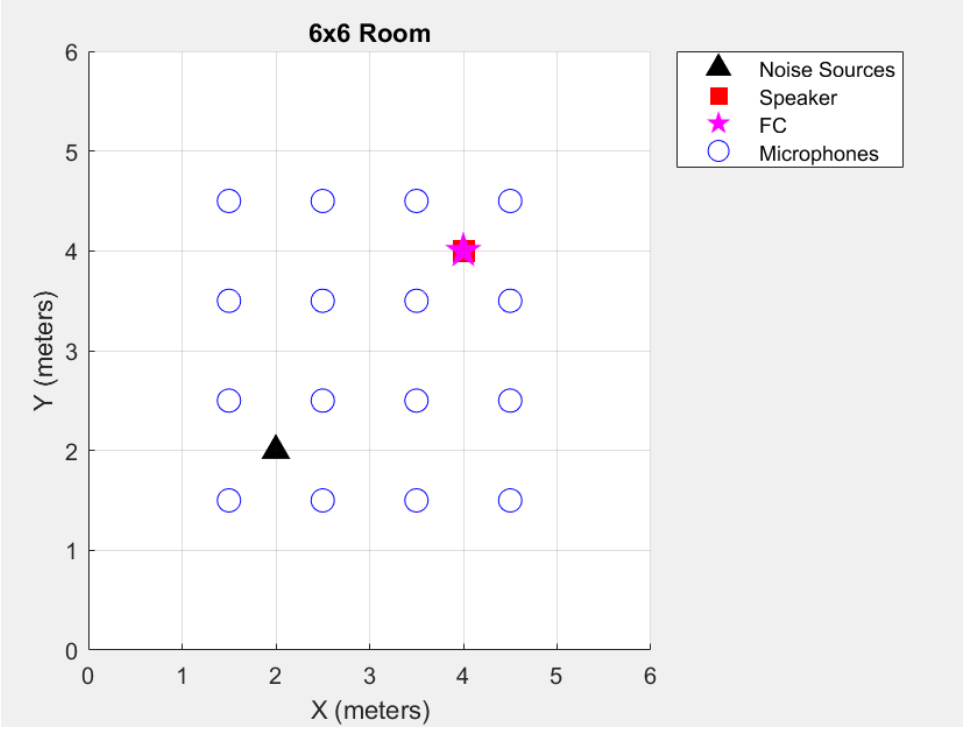
Radius-based method

שיטה זו של בחירת תת קבוצת מיקרופונים היא שיטה שלפיה בוחרים את המיקרופונים שנמצאים במרחק קטן מ-r מה-*fc*, כאשר r הוא פרמטר של השיטה שנקבע לפי הצורך (כמו α בשיטות שלנו). השיטה מבוססת על ההנחה שהמיקרופונים שקרובים יותר ל-*fc* הם מיקרופונים זולים יותר במובן של עלות שידור. היתרון המשמעותי של השיטה הוא שהיא פשוטה מאוד למימוש. החיסרון המשמעותי שלה שהיא לא לוקחת בחשבון את עוצמת הרעש ואת עוצמת אות הדיבור שנקלטים במיקרופונים.

נבחר את השיטה הזאת להיות שיטת הייחוס שלנו ולשם ההוגנות בכל המבחנים וההשוואות שנבצע על השיטות שלנו נקבע את $\boldsymbol{c}$ וקטור המחיר להיות וקטור המרחקים של המיקרופונים מה-*fc*.

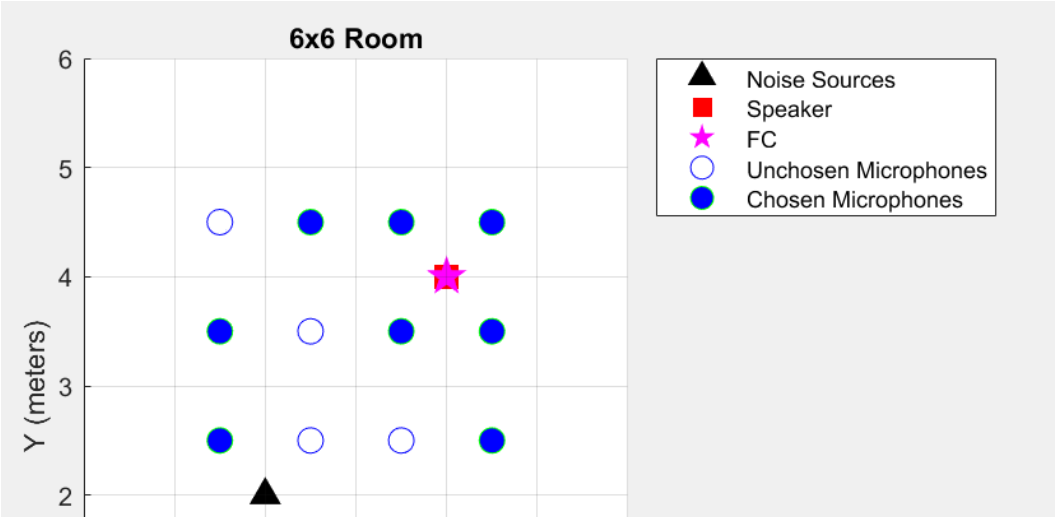
מקור רעש רחוק מה-*fc*, דובר קרוב ל-*fc*

מיקמנו את הדובר ואת ה-*fc* בפינה אחת של החדר, ואת מקור הרעש בפינה השנייה. בחרנו את רעש ה"מיקסר" וקבענו את ההגבר ל-0.5 ביחס לאות הדיבור. גם הדובר וגם מקור הרעש פונים לכיוון מרכז החדר.



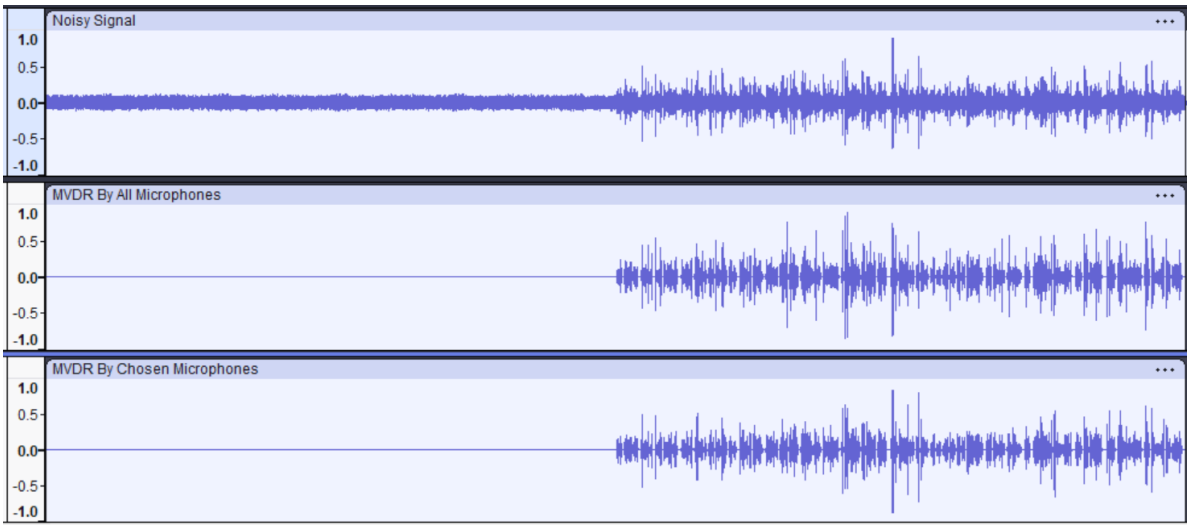
הצפי הוא שבין המיקרופונים שיבחרו יהיו אלו שקרובים ל-*fc* ולדובר מכיוון שהם זולים ואינפורמטיביים יותר ביחס לאחרים. זהו מקרה פשוט יחסית משום שאין *trade-off*.

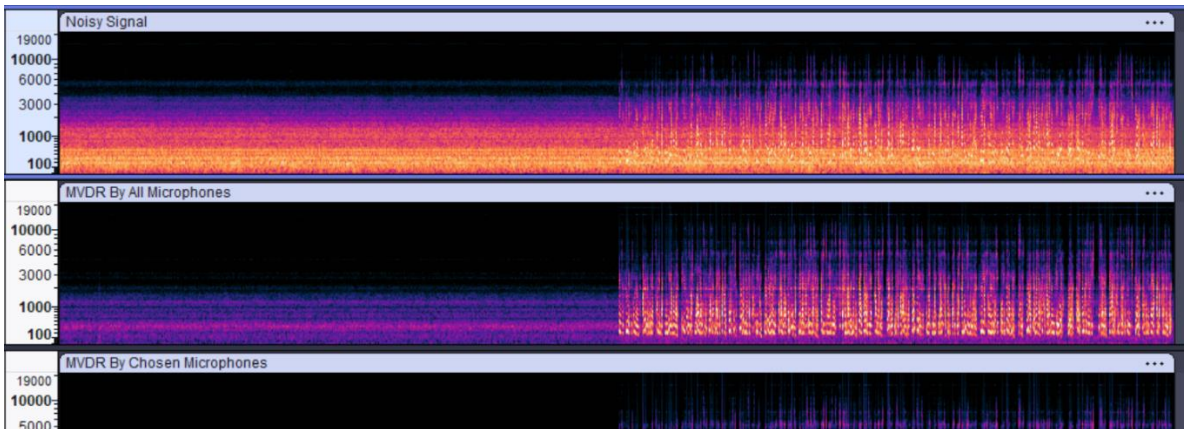
בדקנו את תוצאות שיטה (2) בתרחיש זה עבור $\alpha = 0.7$. אלו המיקרופונים שנבחרו:



נשים לב שארבעת המיקרופונים הקרובים ביותר ל-*fc* נבחרו, זאת משום שהם הזולים ביותר. 3 מתוכם גם קולטים את הדובר בצורה טובה וזו סיבה הגיונית נוספת לבחירתם. כמו כן, מיקרופונים שעוצמת הרעש בהם חלשה (רחוקים ממקור הרעש או שכיוון הרעש הפוך אליהם) נבחרו גם הם. מעניין לראות שהמיקרופונים שלא נבחרו הם בעיקר אלה שעוצמת הרעש בהם חזקה, וזאת למרות שיש מיקרופונים שרחוקים יותר מהם ל-*fc* ולמרות שיש מיקרופונים עם עוצמת אות דיבור חלשה יותר שכן נבחרו.

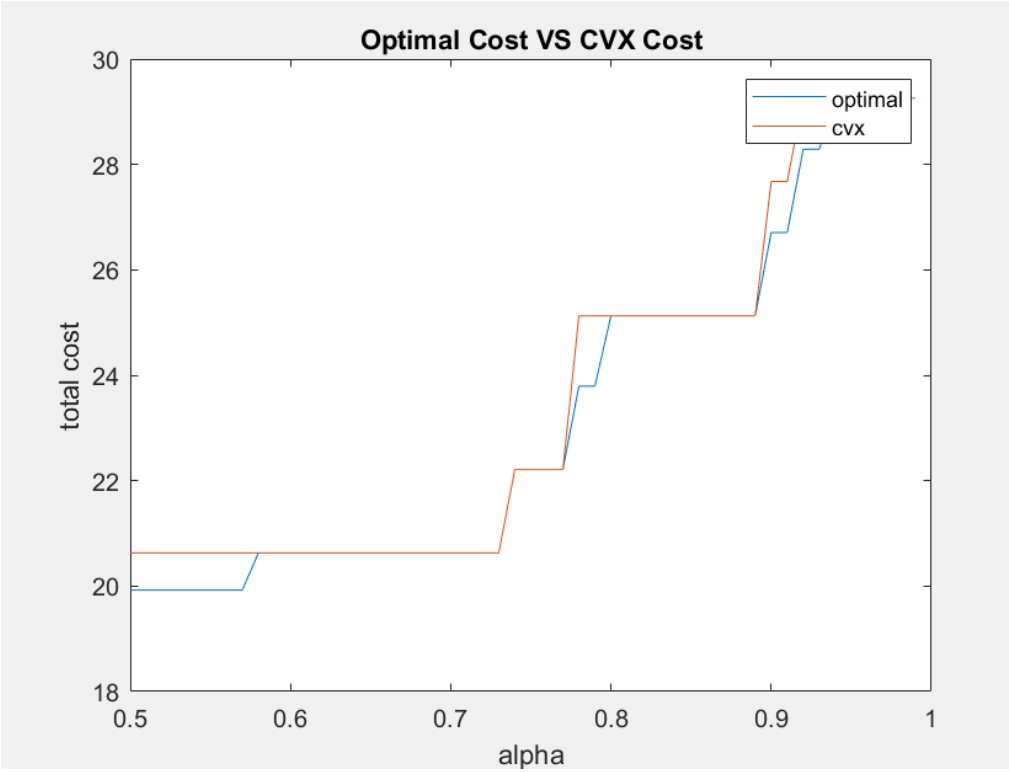
כדי לראות שהאות סונן בצורה טובה למרות הוויתור על חלק מהמיקרופונים נציג את האות לפני הסינון, את האות לאחר *MVDR* ע"י כל המיקרופונים ואת האות לאחר *MVDR* ע"י המיקרופונים הנבחרים:





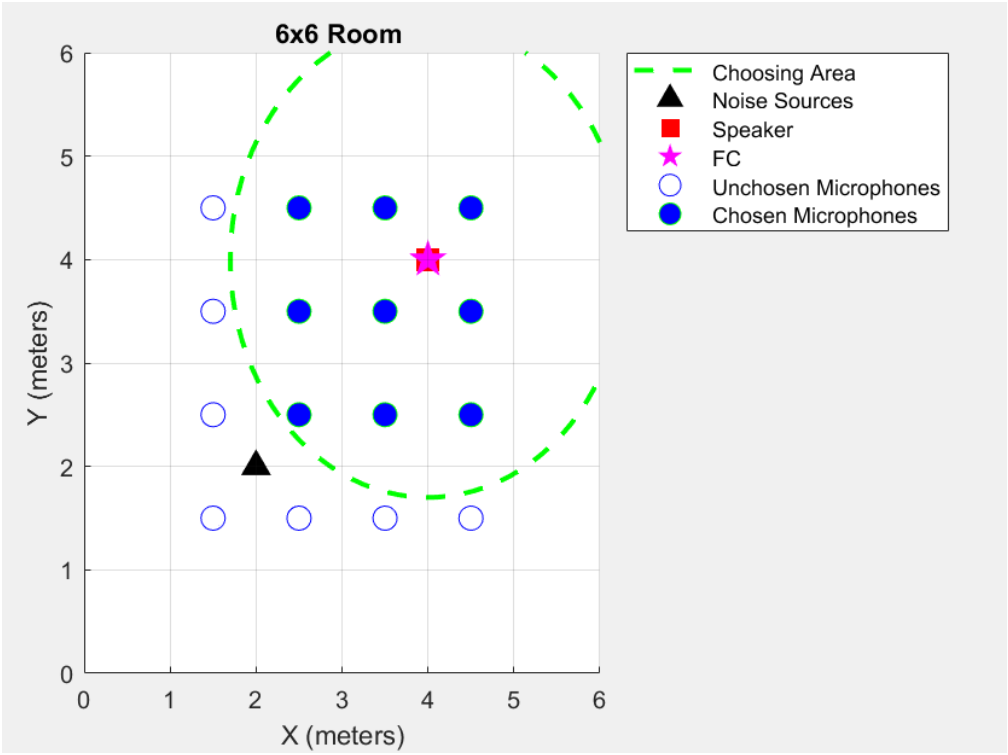
אפשר לראות כי הרעש נוקה בצורה משמעותית גם עבור ה-*MVDR* ע"י המיקרופונים הנבחרים בלבד, ושההבדל ברעש בין *MVDR* ע"י כל המיקרופונים לבין *MVDR* ע"י המיקרופונים הנבחרים קטן (הקשבנו לשתי ההקלטות ובדקנו זאת).

כעת נבדוק את רמת הדיוק של השיטה ביחס לפתרון האופטימלי, אותו נמצא בעזרת בדיקת כל הקומבינציות האפשריות ומציאת תת הקבוצה בעלת המחיר המינימלי מתוך כל הקבוצות שמקיימות את האילוץ $SNR_{\boldsymbol{p}} > \alpha SNR$. נציג את שתי הפתרונות, הפתרון של השיטה שלנו והפתרון הקומבינטורי-אופטימלי עבור ערכי α שונים:



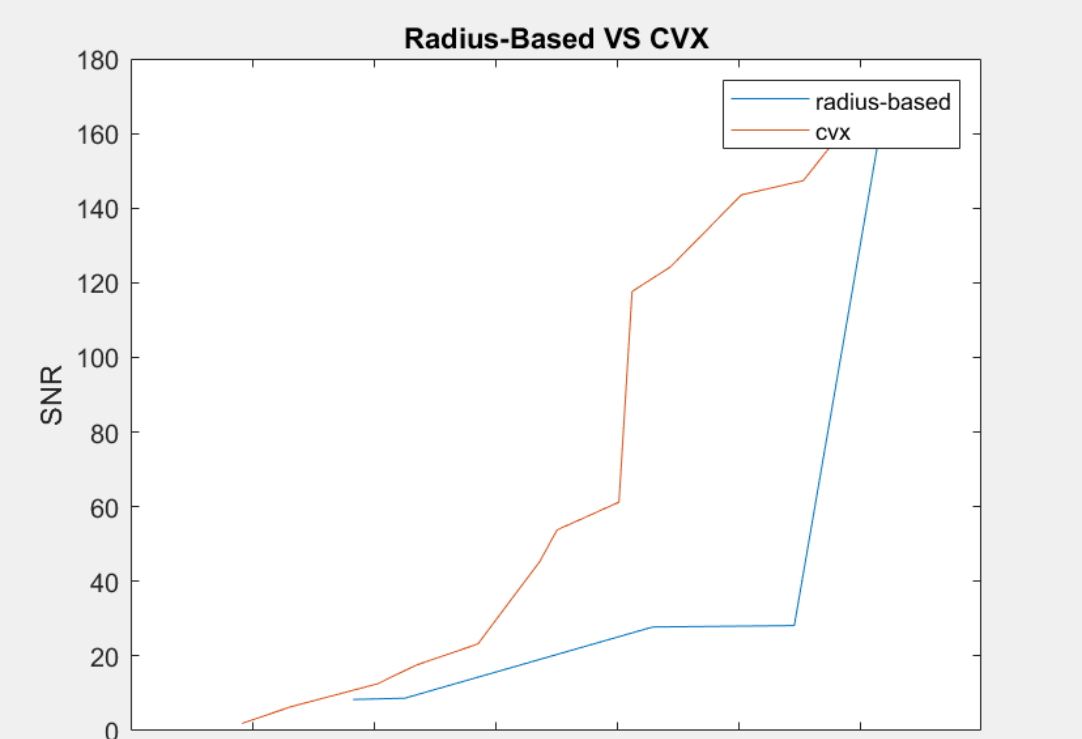
ניתן לראות כי השיטה שהצגנו עוקבת בצורה טובה אחר הפתרון האופטימלי, ועבור ערכי α מסוימים ישנה עקיבה מושלמת.

עבור התרחיש הנ"ל, אם נבחר את המיקרופונים בעזרת שיטת ה-
radius-based ונבחר $r = 2.3m$, יבחרו המיקרופונים הבאים:



נשים לב שהשיטה אומנם בחרה את המיקרופונים הכי זולים, אבל לא בהכרח את המיקרופונים הכי אינפורמטיביים (כמו המיקרופון ב *(2.5, 2.5)* שקרוב מאוד למקור הרעש).

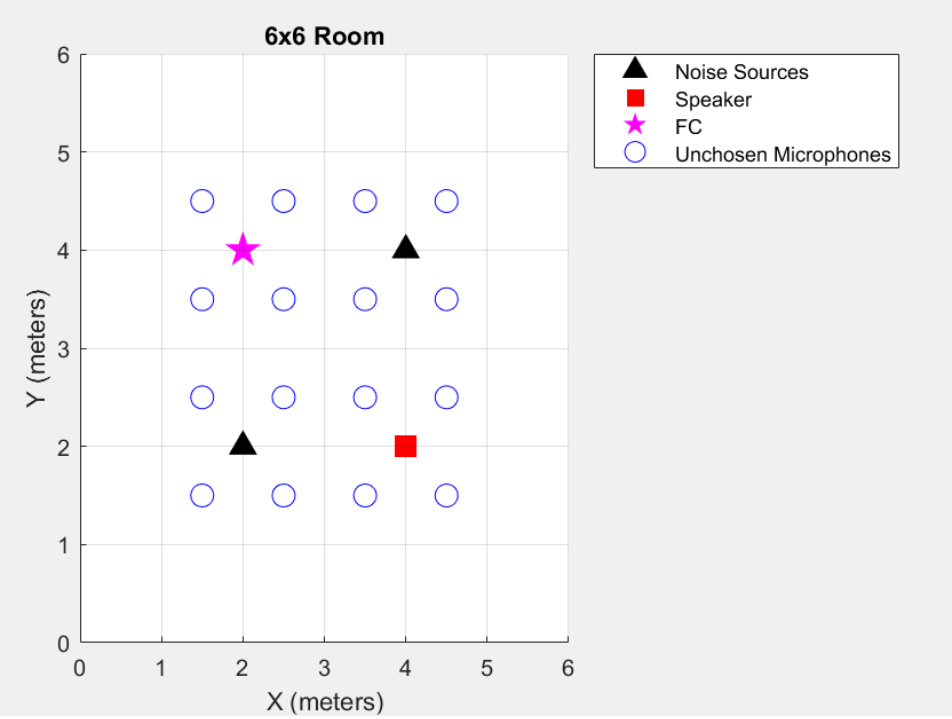
נשווה בין השיטה שהצגנו לבין שיטת ה *radius-based*:



ניתן לראות שעבור אותו מחיר ניתן לקבל *SNR* טוב הרבה יותר בשיטה שהצגנו (כמעט פי 6 עבור $total\ cost \approx 20$) ולכן בתרחיש הזה השיטה שלנו עדיפה על ה-*radius-based* , שזה צפוי כי כאמור ה-*radius-based* לא לוקחת בחשבון את עוצמת האות ואת עוצמת הרעש במרחב.

<u>2 מקורות רעש, דובר ו-*fc* מרוחקים זה מזה</u>

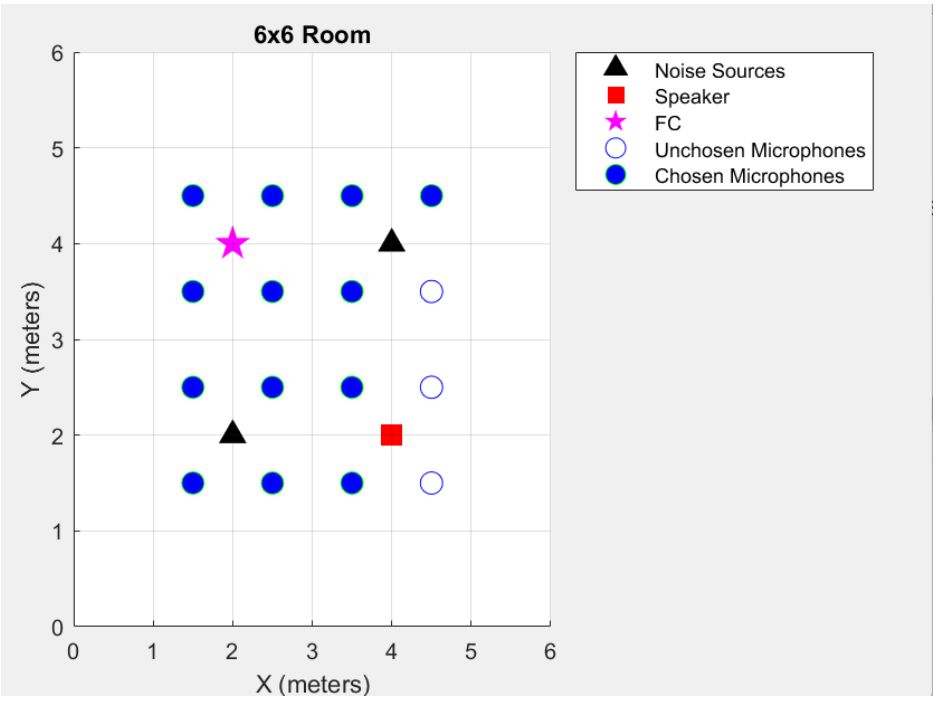
מיקמנו את הדובר, את ה-*fc* ואת שתי מקורות הרעש כל אחד בפינה אחרת של החדר. בחרנו את רעש ה"מיקסר" עם הגבר של 0.5 ביחס לאות הדיבור בתור מקור הרעש הראשון, ואת רעש המזגן עם הגבר של 0.35 ביחס לאות הדיבור בתור מקור הרעש השני. גם הדובר וגם מקורות הרעש פונים לכיוון מרכז החדר.



בתרחיש הזה קשה יותר להחליט אילו מיקרופונים לבחור, מכיוון שהמיקרופונים הזולים יותר הם אלו שקרובים ל-*fc*, אבל המיקרופונים האינפורמטיביים הם אלו שקרובים יותר לדובר. בחרנו את לבחון את המקרה הזה כדי לראות האם השיטה מפיקה תוצאות טובות גם במקרים מסובכים בהם אין בחירה חד-משמעית נראית לעין.

עוד הבדל משמעותי בין התרחיש הזה לתרחיש הקודם זה שבתרחיש זה הוספנו מקור רעש נוסף, זה הופך את R_{nn} למגוונת יותר ואת התרחיש למציאותי יותר ומעניין לראות האם ואיך זה ישפיע על התוצאות.

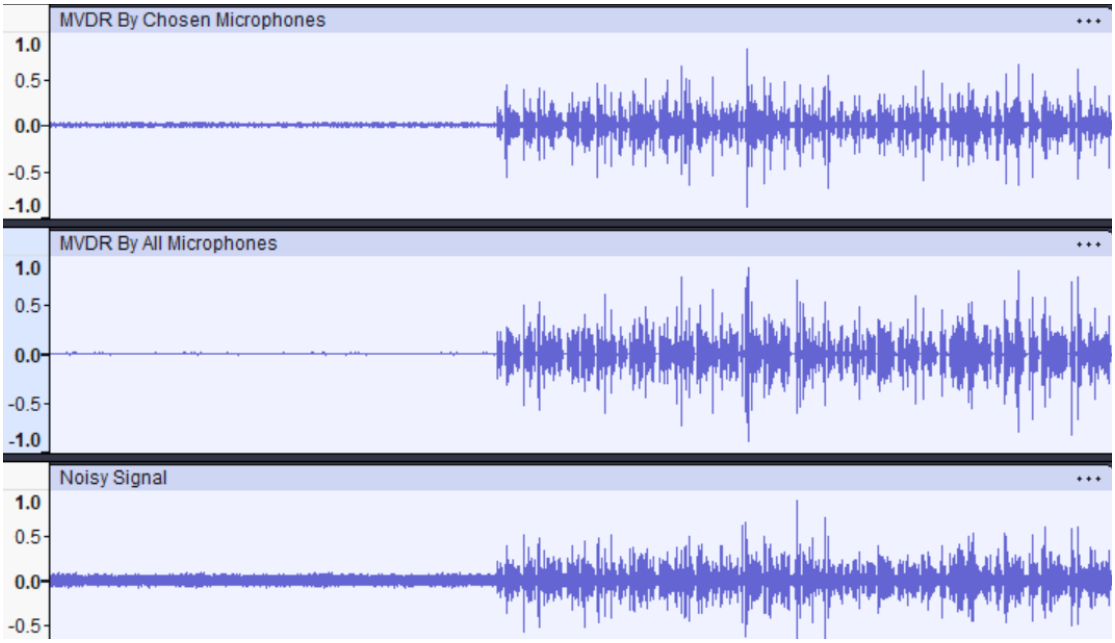
בדקנו את תוצאות שיטה (2) בתרחיש זה עבור $\alpha = 0.7$. אלו המיקרופונים שנבחרו:

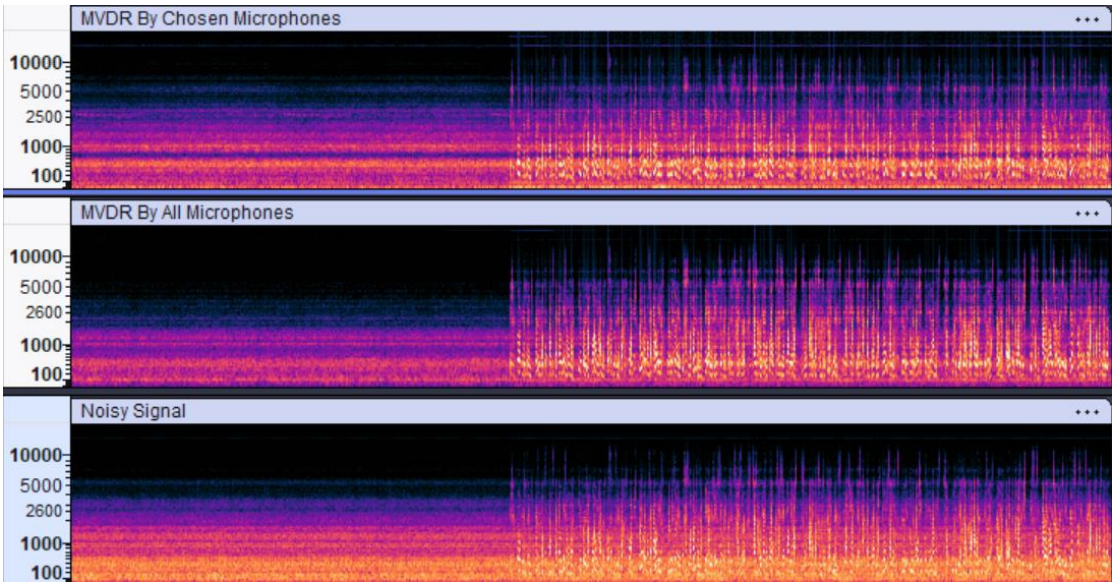


נשים לב שהאלגוריתם ויתר על 2 המיקרופונים היקרים ביותר, כלומר הרחוקים ביותר מה-*fc*. לגבי המיקרופון *(3.5,3.5)* שלא נבחר, ניתן לראות שהוא זול יותר מהמיקרופון *(3.5,1.5)* אבל פחות אינפורמטיבי ממנו ולכן האלגוריתם ויתר דווקא עליו.

סך הכול ניתן לראות שהאלגוריתם פעל במקרה הזה גם תחת שיקולים של מחיר וגם תחת שיקולים של *SNR* כמצופה.

כדי לראות שהאות סונן בצורה טובה למרות הוויתור על חלק מהמיקרופונים נציג את האות לפני הסינון, את האות לאחר *MVDR* ע"י כל המיקרופונים ואת

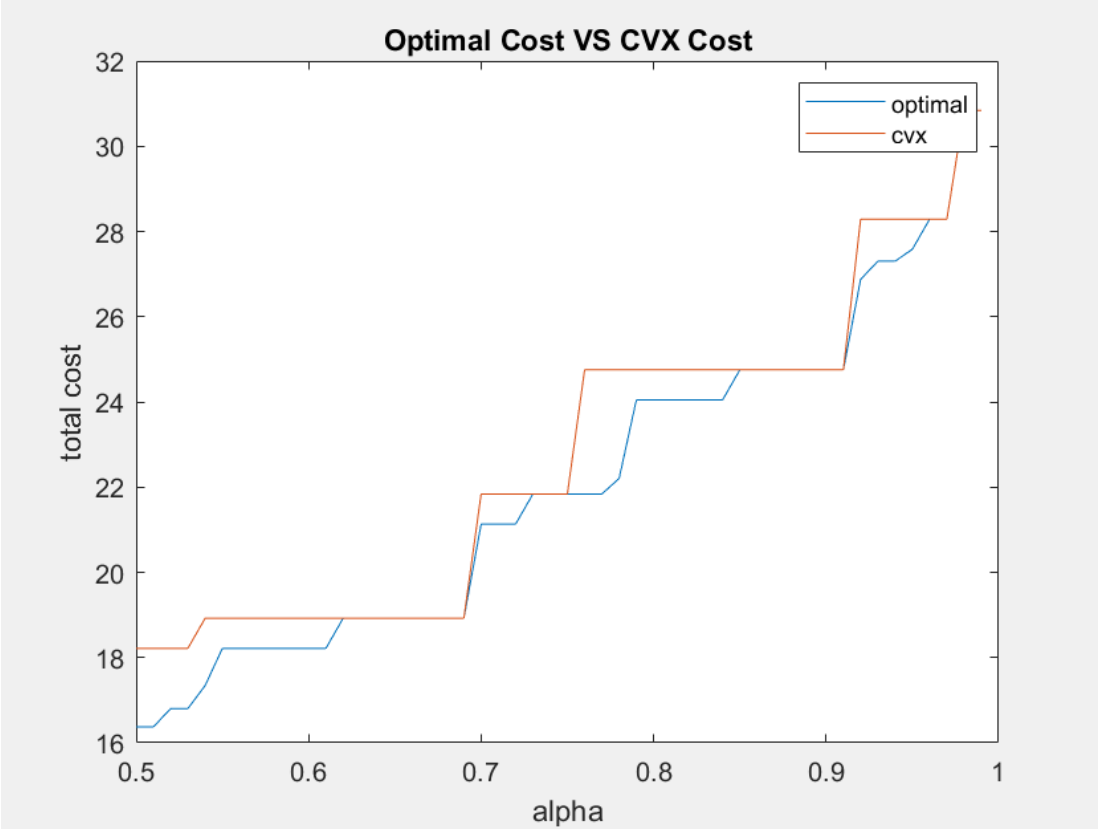




אפשר לראות כי הרעש נוקה בצורה משמעותית גם עבור ה-*MVDR* ע"י המיקרופונים הנבחרים בלבד. אבל, בניגוד לתרחיש הקודם שבדקנו, בתרחיש הזה כן ניתן לראות שיש הבדל ברעש בין ה-*MVDR* ע"י כל המיקרופונים לבין ה-*MVDR* ע"י המיקרופונים הנבחרים (הקשבנו לשתי ההקלטות ובדקנו זאת).

אנו מניחים כי הסיבה לכך היא שהתרחיש הזה יותר מורכב מהתרחיש הקודם כמו שפירטנו.

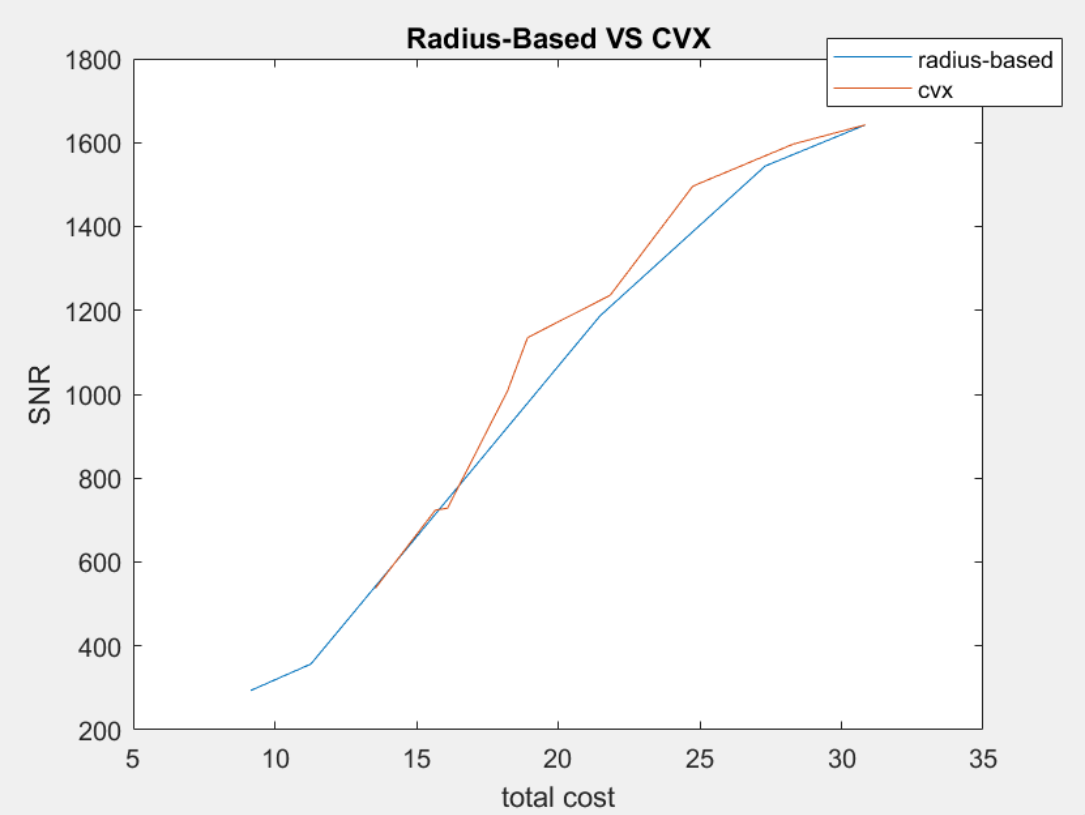
כעת נבדוק את רמת הדיוק של השיטה ביחס לפתרון האופטימלי, אותו נמצא בעזרת בדיקת כל הקומבינציות האפשריות ומציאת תת הקבוצה בעלת המחיר המינימלי מתוך כל הקבוצות שמקיימות את האילוץ $SNR_p > \alpha SNR$. נציג את



ניתן לראות כי השיטה שהצגנו עוקבת בצורה טובה אחר הפתרון האופטימלי, ועבור ערכי α מסוימים ישנה עקיבה מושלמת. אבל, אפשר לשים לב שהעקיבה בתרחיש זה פחות טובה בערכים מסוימים מהעקיבה בתרחיש הקודם, ולדעתנו זה נובע מכך שתרחיש זה מורכב יותר.

לעובדה שהשיטה מפיקה תוצאות קרובות לתוצאות האופטימליות בתרחיש זה יש חשיבות רבה, כיוון שניתן להבין מכך שאפשר לסמוך על השיטה הזו גם בתרחישים שאי אפשר לקבוע בצורה אבסולוטית מהי הבחירה האופטימלית ע"י החלטה אנושית.

נשווה בין שיטה זו לבין שיטת ה *radius-based* בתרחיש זה:



ניתן לראות שבניגוד לתרחיש הקודם, בתרחיש הזה השיטה שהצגנו אינה טובה באופן משמעותי יותר מה-*radius-based* .

זוהי תוצאה מאוד מעניינת שהמשמעות שלה היא שלמרות שהתרחיש הזה מסובך יותר מהתרחיש הקודם, בסופו של דבר השיטה הנאיבית שלפיה בוחרים את המיקרופונים הזולים ביותר מבלי ליחס חשיבות לעוצמת האות והרעש במרחב מביאה תוצאות טובות כמעט כמו השיטה שלנו שכן לוקחת זאת בחשבון.

חשוב לציין שלשיטה שלנו יש ייתרון משמעותי שתקף בכל תרחיש: בשיטה זו ניתן לקבוע מהו ה-SNR הרצוי ע"י בחירת α (בהכרח יתקבל $SNR_p > \alpha SNR$), בעוד שב-*radius-based* לא ניתן לקבוע מה יהיה ה-SNR.

מסקנות מהניסוי

מתוצאות הבדיקות וההשוואות שביצענו עבור שני התרחישים, עולה שהאלגוריתם בוחר תת-קבוצת מיקרופונים הגיונית כך שהמיקרופונים הנבחרים הם זולים ואינפורמטיביים יחסית לאלו שלא נבחרו, וכן האות המסונן במוצא ה-*MVDR* ע"י קבוצת המיקרופונים שנבחרו נוקה בצורה משמעותית מרעש ולא התעוות. עבור ערכי α שנבדקו, הביצועים של האלגוריתם היו קרובים לביצועים האופטימליים שניתן להשיג.

ההבדלים בין תוצאות התרחישים היו בעיקר כאשר השוונו את השיטות לשיטת הייחוס *radius-based* . בתרחיש הראשון האלגוריתם שהצגנו הפיק תוצאות טובות בהרבה מה-*radius-based*, ואילו בתרחיש השני נראה שהביצועים של שתי השיטות קרובים זה לזה.

ראינו שסך הכול השיטה שהצגנו מפיקה ביצועים טובים עבור שני התרחישים, גם בתרחיש הראשון שהינו פשוט יותר וגם בתרחיש השני, שהיה מורכב יחסית.

ישנם שני יתרונות בולטים לשיטה:

1. השיטה יעילה ומהירה הרבה יותר מבדיקת כל הקומבינציות האפשריות של תת קבוצת מיקרופונים ובחירה מתוך הקומבינציות האלו את תת הקבוצה הטובה ביותר. בעוד שסיבוכיות בדיקת כל הקומבינציות תלויה אקספוננציאלית במספר המיקרופונים במערכת, סיבוכיות השיטה שלנו נמוכה הרבה יותר.
2. בניגוד לשיטות אחרות, כמו לדוגמא שיטת ה-*radius-based*, בשיטה זו ניתן לבחור אילוץ על ערך ה-*SNR* במוצא ה-*MVDR* של המיקרופונים

שנבחרו. השיטה מוודאת שאכן עבור המיקרופונים שנבחרו יתקבל ה-*SNR* הרצוי או אפילו ערך גבוה ממנו.

החיסרון העיקרי של השיטה הוא שלא בהכרח תיבחר הקבוצה שמחירה הוא המינימלי, אלא תיבחר קבוצה שמחירה קרוב למחיר המינימלי. הסיבה לכך היא שהשיטה מוצאת ערכי $\boldsymbol{p}$ רציפים שעבורם המחיר מתקבל מחיר מינימלי, וכאשר מתרגמים את $\boldsymbol{p}$ לערכים בוליאניים הדיוק נפגע.
זוהי בעיה עקרונית של השיטה שלא ניתן לפותרה כיוון שאנחנו מנסים לפתור בעיה בוליאנית בעזרת כלי שנועד לפתור בעיות רציפות וזוהי נקודת החולשה המשמעותית ביותר של השיטה.

רעיונות להמשך

בפרויקט זה עסקנו בבעיה של סינון אות דיבור מתוך רעש. ניתן להשתמש בעקרונות ובכלים שהשתמשנו בהם כדי להתמודד עם בעיה זו, על מנת להתמודד עם בעיה של הפרדת דוברים.

בשונה ממה שהנחנו מקודם, כעת נניח שמבנה האות הרועש הוא:

$$\boldsymbol{y} = \boldsymbol{a}s + \boldsymbol{b}r + \boldsymbol{n}$$

כאשר r הוא אות נוסף שאינו רצוי ו-$\boldsymbol{b}$ הוא וקטור ה-*ATF* של האות r.

<u>ננסח את הבעיה החדשה בצורה מתמטית:</u>

$$\min_{\boldsymbol{w}} \boldsymbol{w}^{\boldsymbol{H}} R_{nn} \boldsymbol{w}$$

$$s.t. \quad \boldsymbol{w}^{\boldsymbol{H}} \boldsymbol{a} = 1\,,$$

$$\boldsymbol{w}^{\boldsymbol{H}} \boldsymbol{b} = 0$$

כלומר, אנחנו רוצים למצוא מסנן $\boldsymbol{w}$ כך שעוצמת הרעש במוצא תהיה מינימלית, כיווניות האות הרצוי תהיה 1 וכיווניות האות הלא-רצוי (ההפרעה) תהיה 0.

באופן דומה לפיתוחים שהצגנו עבור בעיית סינון אות דיבור מתוך רעש, ניתן לנסח מתמטית את בחירת המיקרופונים האופטימלית במקרה של ריבוי דוברים באופן הבא:

$$\min_{\boldsymbol{w_p}, \boldsymbol{p} \in \{0,1\}^M} \|diag(\boldsymbol{p})\boldsymbol{c}\|_{\boldsymbol{1}}$$

$$s.t. \quad \boldsymbol{w_p^H} R_{nn,p} \boldsymbol{w_p} \leq \frac{\beta}{\alpha}\,,$$

$$\boldsymbol{w_p^H} \boldsymbol{a_p} = 1$$

$$\boldsymbol{w_p^H} \boldsymbol{b_p} = 0$$

ניתן לפתח את הבעיה הזו תחת *relaxations* כלשהן כך שתהיה *SDP* ואז יהיה ניתן לפתור את הבעיה בעזרת תוכנות מחשב במהירות ויעילות.

תודות

תודה לפרופ' שרון גנות ולד"ר עופר שוורץ על ההנחיה והתמיכה לכל אורך הדרך: יעוץ בבחירת מאמר, נתינת כלים מתאימים כאשר לא הצלחנו להתקדם הלאה, ועזרה בתכנון הפרויקט. תודה על הסבלנות למרות האתגר בלהנחות זוג שבמשך כמעט חצי מתקופת העבודה על הפרויקט היה זמין באופן חלקי בגלל שירות המילואים. תודה שהפניתם אוזן קשבת ודחפתם אותנו לא לוותר, לחקור ולהבין את הנושאים הנלמדים לעומק.

תודה למר פיני טנדייטניק וליואב אליסון מהנדסי המעבדה על העבודה המסורה והמקצועית. למרות הדרישות המורכבות שלנו ולמרות אילוצי הזמן, בזכותם הצלחנו ליצור הקלטות דאטה איכותיות ומדויקות. הצלחת הניסויי היא פרי של השקעתכם ותשומת הלב לפרטים הקטנים ביותר.

תודה לד"ר טום טירר על הסברת נושא האופטימיזציה הקמורה. הבנה בסיסית בתחום הייתה נדרשת עבורנו בשביל להבין בצורה טובה את השיטות שהוצעו במאמר ואף עזרה לנו להסתכל באופן ביקורתי על *relaxations* שנעשו במאמר.

תודה לפרופ' אמיר לשם, המרצה שלנו בקורס "ניתוח סטטיסטי של מידע". למידת הקורס הקנתה לנו רקע משמעותי בשערוך סטטיסטי שהיה מאוד רלוונטי בעבודה על הפרויקט.

תודה שנתתם לנו את ההזדמנות ללמוד, להעמיק ולהתמקצע בתחומים המעניינים והחשובים הללו.