

7/22/2015

אלף האיים – היצירה של קטע קריאה אחיד עבור דוברי עברית
Thousand Islands – A standardized reading passage for Hebrew speakers

עופר אמיר

החוג להפרעות בתקשורת
הפקולטה לרפואה
אוניברסיטת תל אביב

oferamir@tauex.tau.ac.il

אלף האיים – היצירה של קטע קריאה אחיד עבור דוברי עברית

Thousand Islands – A standardized reading passage for Hebrew speakers

תקציר

הקדמה: מחקרים רבים בתחומי הדיבור והשמיעה, כמו גם בתחומים התפתחותיים ותקשורתיים שונים, עושים שימוש תדיר בקטעי קריאה, כמטלה אחידה לבחינת מאפייני הדיבור. עם זאת, עד לאחרונה לא היה קיים קטע קריאה מותאם לדוברי עברית, אשר מיועד למטרה בסיסית זו. לאחרונה פורסם, לראשונה, קטע קריאה בעברית, אשר מיועד למטרה זו, ואולם עד כה לא הוצג מידע על תהליך היצירה שלו, ועל התפלגות הפונמות בתוכו.

מטרות: מחקר זה נועד להציג מידע על תהליך היצירה של קטע הקריאה המאוזן פונמית לשפה העברית, 'אלף האיים'; ולהציג נתונים בסיסיים על התפלגות העיצורים, התנועות, מקומות החיתוך ואופני החיתוך המיוצגים בו.

שיטות מחקר: קטע הקריאה גובש לאחר איסוף קטעי טקסט ממספר מקורות. גרסת הטקסט הראשונית עברה עריכה לשונית, ולאחריה נערכה בחינה של הלימת הטקסט ע"י בלשנית ושלוש קלינאית תקשורת, באופן נפרד ועצמאי. השכיחות של הפונמות השונות בקטע הקריאה נבחנה, כדי להעריך הבדלי שכיחות בין חמש הפסקאות.

תוצאות: השכיחות היחסית של העיצורים, התנועות, מקומות החיתוך ואופני החיתוך בחמש הפסקאות של קטע הקריאה השוותה באמצעות מבחנים סטטיסטיים. העיצור השכיח ביותר בקטע הקריאה היה /m/, ואילו התנועה השכיחה ביותר היתה /a/. הממצאים הדגימו כי השכיחות של הפונמות (עיצורים ותנועות), כמו גם של מקומות ואופני החיתוך היתה דומה בחמש הפסקאות, ללא הבדלים מובהקים סטטיסטיים.

מסקנות ודיון: קטע הקריאה 'אלף האיים' הינו קטע מאוזן פונמית, ושכיחות הפונמות בחמש הפסקאות של קטע הקריאה מייצגת שכיחות דומה ותואמת לשכיחות בשפה העברית.

סיכום: קטע הקריאה מתאים לשימוש למטרות מחקר וקליניקה, עבור דוברי עברית. בשל הייצוג המאוזן של הפונמות בחמש הפסקאות, ניתן להשתמש בקטע הקריאה בשלמותו, או בכל אחת מהפסקאות שבו, באופן מבודד ועצמאי.

Thousand Islands – A standardized reading passage for Hebrew speakers

Abstract

Introduction: Many studies in speech and hearing sciences, as well as in other disciplines, frequently use reading passages, as a standardized speech task. Yet, until recently there was no such reading passage suitable for Hebrew speakers. Such a passage was published recently. Nonetheless, no information on its creation was given, as well as on its phonemic representation.

Aims: This study aimed to describe the preparation process of the phonemically balanced 'Thousand Islands' reading passage, and provide basic data on its phonemic distribution and representation.

Methods: The passage was constructed from online-based sources, and gathered from multiple sites. The primary text was edited and then examined for linguistic coherence and acceptability, by a linguist and by three speech-language pathologists. The prevalence of the different phonemes was examined, to evaluate differences among the five paragraphs.

Results: The prevalence of the consonants, vowels, places of articulation and manners of articulation was examined across the five paragraphs. The most frequently used consonant was /m/, while the most frequent vowel was /a/. No significant differences in phoneme distribution was found among the five paragraphs.

Conclusions: The 'Thousand Islands' reading passage is phonemically balanced; and each of the five paragraphs provides a similar phoneme representation.

Summary: The reading passage is suitable for Hebrew speakers in research and clinical settings. Due to its balanced phonemic structure, it can be used as a whole, or by using each of its paragraphs independently.

מבוא

במחקרים רבים בתחום הדיבור והשמיעה נדרשים הנבדקים לבצע מטלת קריאה, כאמצעי להפקת דיבור במצב מבוקר ואחיד. מטלות הקריאה נמצאות בשימוש תדיר גם בטיפול דיבור שונים, כגון אלה המתמקדים בתרגול בקרת קצב-דיבור, פיתוח ותרגול מאפייני קול ספציפיים ובמחקרים בתחום מדעי ההתנהגות. עד לאחרונה לא היה נמצא קטע קריאה מקובל ואחיד בעברית, אשר יהיה מתאים לאנשים בוגרים, ואשר יהיה ניתן להשתמש בו למטרות מחקר וטיפול. בשל כך, חוקרים ומטפלים דוברי עברית, אשר מתמקדים בתחומי השפה, הדיבור והשמיעה נוהגים להשתמש בקטעי קריאה שונים שאינם אחידים.

קטע הקריאה המוכר ה-"Rainbow Passage" [1] למשל, משמש כקטע הקריאה הסטנדרטי המקובל במרבית המחקרים הנערכים בשפה האנגלית, בשל העובדה שהוא מספק ייצוג רחב של שכיחות הפונמות בשפה האנגלית. העובדה שקטע קריאה זה מוכר ונמצא בשימוש נרחב, מגביר את האחידות בין מחקרים שונים המבוצעים בשפה האנגלית, ומשפר את יכולת ההשוואה בין המחקרים השונים. קטעי קריאה דומים פורסמו עבור שפות שונות בעולם, מתוך הכרה בחשיבות השימוש בקטעי קריאה מאוזנים פונמית ובחשיבות של השימוש בקטעי קריאה אחידים למטרות מחקר וקליניקה. מכיוון שלא פורסם קטע קריאה דומה בשפה העברית, עלה הצורך לגבש קטע קריאה מתאים, אשר יוכל לשמש חוקרים מדיסציפלינות שונות שנדרשים להעריך דוברים במטלות דיבור וקריאה.

לאחרונה פורסם קטע הקריאה 'אלף האיים', שהינו מאוזן פונמית ומיועד לדוברי עברית בוגרים [2]. קטע הקריאה הוצג במסגרת מחקר על עמדות מאזינים כלפי גברים ונשים הסובלים מהפרעות קול. הקטע פורסם, בעברית, כנספח של המאמר, וניתן להוריד באופן ישיר מאתר האינטרנט של המחבר:

www.oferamir.co.il

לאור הצורך שעלה לאחר הצגת קטע הקריאה, וכהשלמה לפרסום קטע הקריאה בספרות המקצועית, מיועד המאמר הנוכחי לתאר את תהליך גיבוש ויצירת קטע הקריאה, ואת בחינת התפלגות הפונמות בחלקיו השונים, כדי לאפשר ולהקל על חוקרים וקלינאי תקשורת דוברי עברית את השימוש העתידי בו.

שיטות מחקר

קטע הקריאה 'אלף האיים' חובר ופותח ע"י צוות מעבדת הקול של החוג להפרעות בתקשורת באוניברסיטת תל-אביב, כחלק מפרויקט מחקר ארוך שנים המתבצע במעבדה זו. קטע הקריאה הינו טקסט מידעי, אשר גובש על בסיס מקטעי טקסט קצרים שקובצו ממגוון של אתרי אינטרנט. לשם כך, נערך חיפוש ראשוני באתרי אינטרנט אשר סיפקו מידע על תוכן הקטע, ובשלב הראשון נבחרו מספר משפטים מכל מקור מידע. קטע הקריאה הראשוני שגובש ממשפטים אלו עבר עריכה לשונית, ולאחריה עיבוד וארגון מחדש לחמש

פסקאות קצרות, כאשר כל פסקה מכילה חמישה או שישה משפטים. הבחירה לארגן את קטע הקריאה לחמש פסקאות קצרות נעשתה במטרה לגבש קטע קריאה קצר, אשר יאפשר שימוש גם בחלקיו השונים, באופן עצמאי.

לאחר הגיבוש הראשוני של הטקסט, עבר קטע הקריאה עריכה לשונית חוזרת על ידי צוות המעבדה. בשלב הבא הועבר קטע הקריאה להערכה נפרדת ע"י בלשנית וע"י שלוש קלינאיות תקשורת העוסקות בשפה אצל דוברים בוגרים. הליך זה נועד להבטיח שקטע הקריאה הסופי יהיה תקין לשונית, יהיה נוח לקריאה ויתאים כמטלת קריאה עבור דוברי עברית בוגרים.

תוצאות

בחינה ראשונית של התפלגות הפונמות בקטע הקריאה השלם ממחישה כי העיצור השכיח ביותר בקטע הקריאה הוא הפונמה /m/, אשר הופיע 164 פעמים (18.9%). העיצורים אשר דורגו אחריו בתדירות ההופעה היו העיצורים /l/ ו-/t/, אשר הופיעו 84 ו-82 פעמים בקטע הקריאה (9.4%-ו 9.7%) בהתאמה. לעומתם, העיצור הנדיר ביותר בקטע הקריאה היה /ʔ/, אשר הופיע שלוש פעמים בלבד (0.3%).

טבלה מס' 1: התפלגות הפונמות (עיצורים ותנועות) בחמש הפסקאות של קטע הקריאה

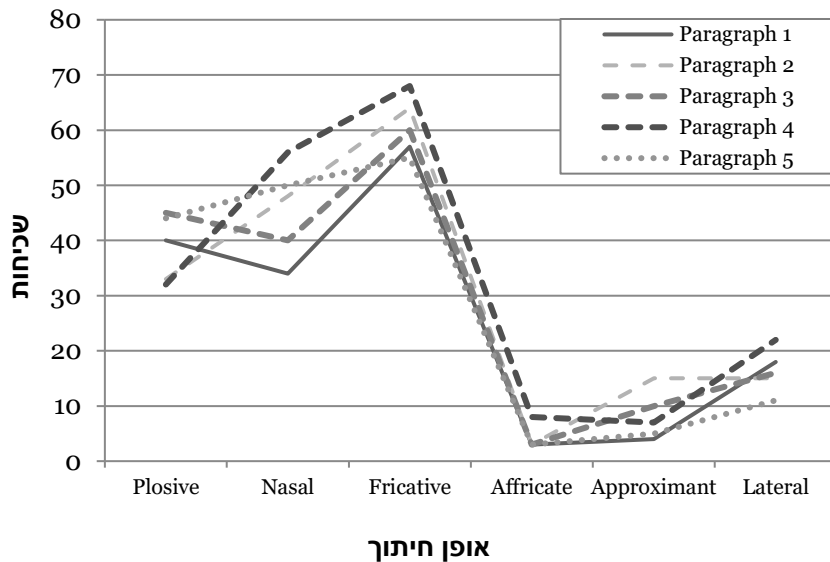
Paragraph	Phoneme													
	/p/	/b/	/m/	/f/	/v/	/l/	/t/	/d/	/n/	/s/	/z/	/ts/	/ʃ/	
1	3	6	25	5	9	18	21	1	9	2	3	3	7	
2	1	7	40	3	9	15	13	2	8	2	2	3	17	
3	2	7	29	9	8	16	18	5	11	3	0	3	9	
4	0	5	37	7	13	22	14	3	19	3	1	8	9	
5	4	6	33	2	8	11	18	5	17	10	1	3	5	
Total	10	31	164	26	47	82	84	16	64	20	7	20	47	

	Phoneme													
	/j/	/k/	/g/	/ɸ/	/x/	/h/	/ʔ/	/i/	/E/	/a/	/o/	/u/		
1	4	8	1	13	12	6	0	21	35	39	20	4		
2	15	5	2	9	9	13	3	28	38	56	10	6		
3	10	11	2	14	12	5	0	43	29	49	17	10		
4	7	5	5	12	7	16	0	29	41	47	20	7		
5	5	7	4	15	5	9	0	28	24	48	18	13		
Total	41	36	14	63	45	49	3	149	167	239	85	40		

כידוע, מקובל לתאר את מערכת התנועות של השפה העברית כמכילה חמש תנועות [3]. מבין חמש התנועות, התנועה השכיחה ביותר בקטע הקריאה היתה התנועה /a/, אשר הופיעה 239 פעמים (35.1%). לעומתה, התנועה הנדירה ביותר היתה /u/, אשר הופיעה 40 פעמים (5.9%). על מנת להעריך את ההבדלים בשכיחות הפונמות בקטע הקריאה, נערכו ניתוחים סטטיסטיים להשוואה בין הפונמות ובין הפסקאות. הניתוחים הסטטיסטיים הדגימו כי לא היה הבדל מובהק בין הפסקאות בהתפלגות כלל העיצורים

($\chi^2_{(19)}=107.54, p=0.18$). בנוסף, על מנת להעמיק בבחינת התפלגות הפונמות, אורגנו העיצורים על פי אופן החיתוך שלהם, ולאחר מכן – על פי מקום החיתוך שלהם. טבלה מס' 1 מציגה את התפלגות השכיחות של עיצורים ותנועות בכל אחת מחמש הפסקאות של קטע הקריאה.

תרשים מס' 1 מציג את השכיחות של אופני החיתוך השונים בחמש הפסקאות של קטע הקריאה. כפי שניתן לראות, השכיחות של אופני החיתוך השונים דומה בחמש הפסקאות של קטע הקריאה. בנוסף, ניתן לראות

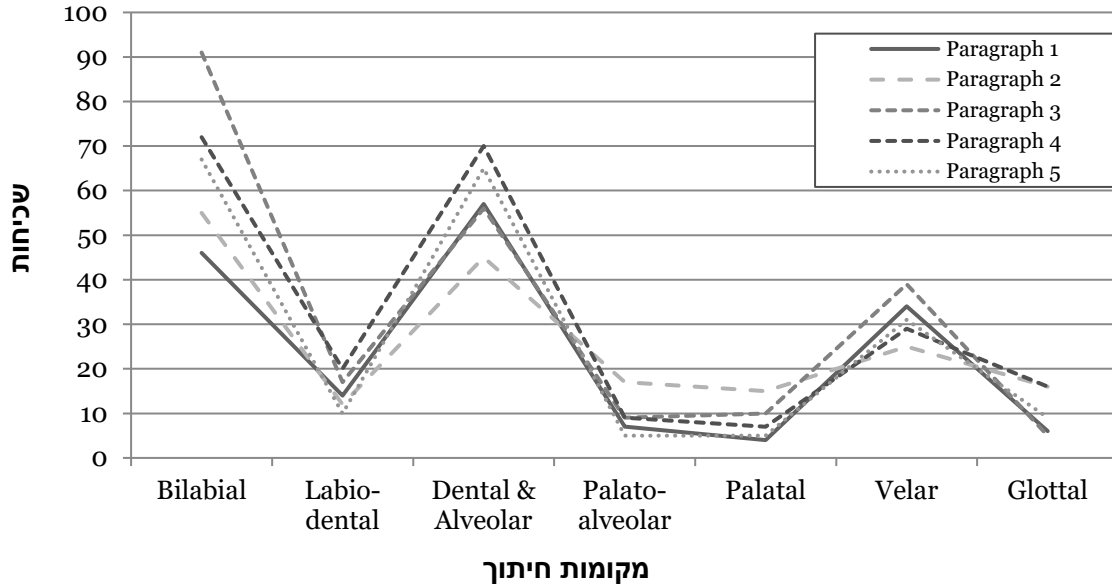


כי אופן החיתוך השכיח ביותר הוא החוכך (fricative), אשר כלל עיצורים כגון /f, s, v/, בעוד שאופן החיתוך הנדיר ביותר היה המחוכך (affricate), אשר כלל רק את העיצור /ts/. ניתוח סטטיסטי אישש כי השכיחות של אופני החיתוך השונים היתה דומה בחמש הפסקאות השונות של קטע הקריאה, וכי לא היו הבדלים מובהקים סטטיסטית בשכיחות אופני החיתוך בין הפסקאות ($\chi^2_{(4)}=23.134, p=0.16$).

תרשים מס' 1: שכיחות אופני החיתוך בחמש הפסקאות של קטע הקריאה

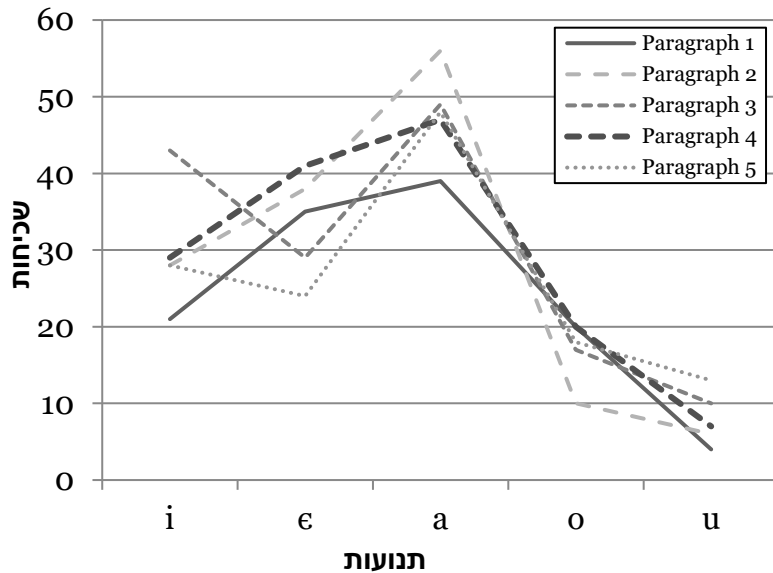
תרשים מס' 2 מדגים את התפלגות העיצורים בקטע הקריאה, על פי מקומות החיתוך שלהם. כפי שניתן לראות, השכיחות של מקומות החיתוך דומה בחמש הפסקאות של קטע הקריאה. בנוסף, ניתן לראות כי בכל הפסקאות של קטע הקריאה, מקומות החיתוך השכיחים ביותר היו: Bilabial, Dental & Alveolar ו-Velar.

בחינה ראשונית של ההבדלים בין הפסקאות הצביעה על הבדלים מובהקים סטטיסטית בתפוצה של מקומות החיתוך בחמש הפסקאות של קטע הקריאה ($\chi^2_{(4)}=49.41, p=0.017$). ואולם, לאחר ביצוע תיקון למדידות מרובות (Bonferroni correction for multiple comparison) נמצא כי ההבדלים המספריים הקטנים בשכיחות מקומות החיתוך בין הפסקאות השונות לא היו מובהקים סטטיסטית ($adjusted\ p>0.05$). בהתאם לכך ניתן לקבוע כי חמש הפסקאות של קטע הקריאה מספקות ייצוג דומה של מקומות החיתוך השונים.



תרשים מס' 2: שכיחות מקומות החיתוך בחמש הפסקאות של קטע הקריאה

לבסוף, נערכה בחינה של שכיחות חמש התנועות בפסקאות השונות של קטע הקריאה. תרשים מס' 3 מציג את השכיחות של חמש התנועות בחמש הפסקאות. בדומה לממצאים ולמסקנות שהתקבלו ביחס לשכיחות



העיצורים, ניתן לראות כי גם השכיחות היחסית של התנועות השונות בחמש הפסקאות היתה דומה. ניתוח סטטיסטי של ההבדלים בשכיחות התנועות בחמש הפסקאות הדגים כי לא היה הבדל מובהק סטטיסטית בשכיחות של התנועות השונות בחמש הפסקאות של קטע הקריאה ($X^2_{(4)}=24.03, p=0.08$).

תרשים מס' 3: שכיחות התנועות השונות בחמש הפסקאות של קטע הקריאה

ממצאי הבחינה האמפירית של התפלגות העיצורים, התנועות, מקומות החיתוך ואופני החיתוך בקטע הקריאה מצביעים על כך שלא קיימים הבדלים מובהקים בין חמש הפסקאות. תוצאה זו היתה מובהקת סטטיסטית ועקבית עבור העיצורים, כמו גם עבור התנועות. בהתאם לכך, ניתן לקבוע כי חמש הפסקאות של קטע הקריאה מספקות ייצוג דומה ועקבי של התפלגות הפונמות. ממצא זה מהותי, כיוון שהוא מאפשר להשתמש בכל אחת מהפסקאות של קטע הקריאה כייצוג תקף ומקביל של התפלגות הפונמות בשפה העברית.

במהלך היצירה, ההערכה והתיקוף של קטע הקריאה, נעשה מאמץ מיוחד לאתר מידע אמפירי ותקף על השכיחות היחסית של הפונמות בשפה העברית. ואולם, סקירה של מאגרי המידע הממוחשבים לא העלתה מידע מסוג זה אשר תוקף ופורסם עד כה. בשל כך, לא היה ניתן להשוות את התפלגות הפונמות שנצפתה במחקר הנוכחי עם מחקרים קודמים.

מאידך, מקור המידע בר-התוקף והמהימן היחיד, אשר סיפק נתונים אמפיריים על השכיחות היחסית של הפונמות בעברית התקבל מעבודת דוקטורט בודדה (Schocken, 2008). בעבודת דוקטורט רחבת היקף זו נבחנה השכיחות היחסית של הפונמות בשפה העברית, מתוך מדגם אקראי של 13,720,000 מילים בעברית, אשר נאספו ממאגר גדול של דפי אינטרנט. בשל ההבדלים המתודולוגיים בין מחקרה של Schocken לבין המחקר הנוכחי, ומכיוון שלא התאפשרה לנו הנגישות אל הנתונים הגולמיים של מחקר זה; לא ניתן היה לבצע בחינה סטטיסטית של ההבדלים בהתפלגות הפונמות בין שני המחקרים. מאידך, בחינה איכותנית של הנתונים מדגימה ממצאים בעלי דמיון רב בין שני המחקרים. לדוגמא, בדומה לממצאים שתוארו לעיל, גם במחקרה של Schocken היה העיצור /m/ השכיח ביותר, והופיע בתדירות יחסית של 11.28% (ערכי השכיחות אשר חושבו במחקרה של Schocken, חושבו באופן שונה מאשר במחקר הנוכחי, וכפי הנראה, בשל כך, קיימים הבדלים בערכים המוחלטים בשני המחקרים). בנוסף, במחקרה של Schocken דווח כי שני העיצורים העוקבים ברמות השכיחות שלהם היו /t/ ו-/l/, אשר הופיעו בשכיחות יחסית של 10.41% ו-8.23%, בהתאמה. זאת ועוד, התפלגות השכיחות של התנועות היתה דומה בשני המחקרים והצביעה על אותן מגמות. לדוגמא, התנועה /a/ היתה השכיחה ביותר במחקרה של Schocken, והופיעה בתדירות של 59.2%, בעוד התנועה /u/ היתה הנדירה ביותר, והופיעה בשכיחות יחסית של 7.1%.

לסיכום, למרות שלא קיימים נתונים נורמטיביים אשר פורסמו ואשר מספקים מידע מהימן על השכיחות של הפונמות השונות בשפה העברית, אשר אליהם יהיה ניתן להשוות את ממצאי המחקר הנוכחי, אנו מסיקים כי קטע הקריאה 'אלף האיים' מספק ייצוג הולם ומאוזן פונמית של השפה העברית. בנוסף, ממצאי המחקר הדגימו כי לא קיימים הבדלים מובהקים בין חמש הפסקאות של קטע הקריאה בשכיחות היחסית של העיצורים, של התנועות, של מקומות החיתוך ושל אופני החיתוך. ממצא זה מכווון לכך שבעת ביצוע מטלות קריאה, ניתן להשתמש בכל אחת מהפסקאות של קטע הקריאה בנפרד, או בכל שילוב של הפסקאות, כמייצג את השכיחות היחסית של הפונמות בעברית. עם זאת, מכיוון שמטרתו הראשונית של המחקר הנוכחי היתה להציג את קטע הקריאה; לש מקום לבחינה עתידית של ערכי השכיחות של הפונמות בעברית, ולא ראוי להתייחס לערכי

השכיחות של הפונמות השונות כמייצגים, באופן גורף, את השכיחות של הפונמות בשפה העברית. לאור כך, מומלץ לבצע מחקרי המשך עתידיים, בהם תמשיך הבחינה של שכיחות הפונמות, תוך מתן התייחסות – בין היתר - להבדלים בהגייה התקנית לעומת ההגייה השכיחה של מלים מסוימות, וכן לאופני ההגייה השונים הנהוגים בקרב דוברי עברית ילידים.

סיכום

קטע הקריאה, 'אלף האיים', אשר פורסם בשנת 2013, מתאים לשימוש כקטע קריאה אחיד עבור דוברי עברית. בשל הייצוג המאוזן של הפונמות בחמש הפסקאות, ניתן להשתמש בקטע הקריאה השלם או לחלופין, בכל אחת מהפסקאות באופן עצמאי.

ביבליוגרפיה

1. Fairbanks G. Voice and Articulation Drillbook. 2nd ed. New York: Harper & Bros. 1960.
2. Amir O, Levine-Yundof R. Listeners' attitude toward people with dysphonia. *J Voice*, 2013;27:524.e1-10.
3. Most T, Amir O, Tobin Y. The Hebrew vowel system: Raw and normalized acoustic data. *Lang Speech*, 2000;43:295-308.
4. Schocken R. Characterization and evaluation of speech-reading support systems for hard-of hearing people. *Unpublished doctoral dissertation*. Tel-Aviv University, Tel-Aviv, Israel (Hebrew). 2008.