NOVEL MEASURES FOR VOWEL REDUCTION

Noam Amir, Ofer Amir

Department of Communication Disorders, Sackler Faculty of Medicine, Tel-Aviv University
noama@post.tau.ac.il, oferamir@post.tau.ac.il

ABSTRACT

Novel measures for vowel reduction are presented here, for examining vowel space as a whole, and for quantifying reduction of individual vowels. These measures were used to evaluate the degree of vowel reduction in continuous speech, as manifested in the F1-F2 plane. The new measures were applied to a set of 1500 tokens, extracted from a database of spontaneous Hebrew speech (30 tokens of each vowel, recorded from five men and five women). Using a similarity measure, we found that vowels were reduced by a factor of 2.09 for men and by 2.93 for women. The reduced vowel space for men was more distorted than for women. Error measure estimations were larger for men in comparison to women (0.0714 versus 0.0525, respectively). While vowel reduction in women exhibited a relatively symmetric pattern across vowels, it showed a skewed pattern in men. This was attributed to a more pronounced reduction in the back vowels /o/ and /u/.

Keywords: vowel space, vowel reduction, formants, formant space

1. INTRODUCTION

Vowel reduction is the relationship between the idealized vowels of a given language and their manifestation in spontaneous (or "continuous" / "natural") speech (SS). Description and quantification of vowel reduction is of interest to various disciplines, such as articulatory and acoustic phonetics and others. Vowel reduction is manifested mainly in two ways:

1. Vowel "Centralization": vowels in SS tend to be similar to each other. They fall between their "idealized" values and a central point on the F1-F2 space (i.e., /↔/).
2. Increased Variability: SS is typically characterized by increased variability, in comparison to phonetically controlled speech, due to changes in speech rate, stress and other linguistic factors in SS.

Many studies have been performed to determine the influence of various factors on the degree of vowel reduction. These studies have confronted two typical problems: (a) finding suitable metrics for quantifying the degree of reduction, and (b) obtaining reliable data to represent the "typical production" of the target vowels.

Blomgren et al. [1], for example, reviewed several metrics for measuring individual vowel reduction and overall vowel space reduction. Overall vowel space was characterized by the extremities of the vowel polygon in the F1-F2 space (the triangle spanned by /i/, /a/ and /u/). The measures for quantifying vowel-space reduction were:

1. **Formant Spacing**: the compact-diffuse (C-D) feature, which is F2-F1, and the grave-acute (G-A) feature, which is (F1+F2)/2.
2. **Vowel Area**: the area of the triangle spanned by the above vowels.
3. **Euclidean Distance**: the distance between the F1-F2 point of a vowel and the centroid of the triangle spanned by the above three vowels.

Different measurement methods can be applied for quantifying the entire vowel space and its reduction, as a whole, in comparison to quantifying the reduction of specific vowels. The Euclidean distance between the reduced and "standard" vowel can provide an estimate their distance from each other. Yet, measuring the two vowels in relationship to a theoretic or recognized central point (i.e., /↔/) could be more instructive. This central point could be identified as the geometric centroid of the vowel space [1], or as the point obtained from a theoretical vocal tract in a neutral configuration [2]. Despite the existence of measures for individual vowel reduction, valid measures for describing the different ways in which the entire vowel space is transformed are lacking. Furthermore, existing measures are often applied to a limited subset of vowels, which could lead to overlooking subtle differences in vowel space configuration.

The Hebrew vowel system is composed of the five vowels, /i, ε, a, o, u/. This vowel system does not include a phonological vowel reduction, nor does it include the neutral Schwa (/↔/) as a phoneme. The acoustic properties of Hebrew vowels, as manifested in isolated utterances (IU), were described in detail by Most et al. [3].

In a sense, the study of a five-vowel system lends itself readily to examination of vowel reduction. Because vowel-duration and lax/tense differences are not phonemic in Hebrew, the points in the F1/F2 space, lying nearly on a geometric triangle, are virtually a complete description of the vowel system. Therefore, geometric calculations are easily applied to this type of vowel distribution.

2. METHOD

2.1. The Speech Corpus

Speech samples were taken from CoSIH (Corpus of Spoken Israeli Hebrew) [4], which is a speech corpus intended to serve as a "snapshot" of spoken Hebrew. It is based on recordings of 45 subjects who wore two unobtrusive microphones on their lapels, during an entire waking day.

Recordings of five men and five women from this corpus were taken for this study. All speakers were native speakers of Hebrew, residing in Israel since birth. Mean age was 30.5 years (SD=14.6) for the men group, and 26.3 years (SD = 5.5) for the women. The first 30 productions of each vowel, from each speaker were identified and saved separately as a sound file, with a sampling rate of 48kHz. A total of 1500 vowel productions were obtained (30 productions X 5 vowels X 10 speakers). A segment of 30-50 ms, from the center of each token was extracted, which was judged visually to be stationary.

2.2. Formant Extraction

Linear Predictive Coding (LPC) analysis was applied to the signal. The first two formant frequencies were identified as the frequencies of the first two conjugate pole pairs obtained from this analysis. The results of this analysis may vary considerably, when applied automatically, depending on sampling rate and order of the LPC analysis. Therefore, we performed a manually supervised analysis of each token separately.

2.3. Measures of Vowel Reduction

Several measures for vowel reduction were examined. Our objective was to identify quantitative measures that could be used for vowels in IU and SS, as well as identifying measures that would strictly applied for IU or SS. Initially, we applied the measures summarized by Blomgren et al. [1].In addition, we examined additional novel measures. These are listed below.

2.3.1. Previously Used Measures:

I) Triangle Area: An overall measure of vowel space reduction, which is the area of the /i-a-u/ triangle in IU divided by its equivalent in SS. The square root of this area provides a measure comparable to the similarity ratio below.

II) C-D, G-A Measures: As described above, the C-D feature is F2-F1, and the G-A feature is (F1+F2)/2.

III) Euclidean Distances: The distances between the vowel points in the F1-F2 plane and the centroid of the vowel polygon. Taken alone, these distances provide limited information. However, they can serve as a baseline for measuring reduction.

2.3.2. Novel Measures:

I) Vowel Space Similarity: visual inspection of the IU and SS vowel polygons revealed clear similarity. Two measures quantiy this: the axis scaling that yields the best fit, and the mean square error between the best fit and target polygon.

To calculate these two measures, we assume that the nodes of the two polygons (IU, SS) are the two sets of N pairs (Xi,Yi) and (X'i,Y'i), where N is the number of nodes (in this case, N=5). First, we center both polygons on the origin. The centroid of each polygon is then obtained by taking the mean of all x values and y values of its nodes. The nodes of the centered polygons are thus defined as: xi=Xi-mean(Xi), yi=Yi-mean(Yi), x'i=X'i-mean(X'i), y'i=Y'i-mean(Y'i). For a given scale factor “*a*”, the mean square error E between the two polygon's nodes is:

$$E = \sum_i (x_i - a x_i^{'})^2 + \sum_i (y_i - a y_i^{'})^2 \quad (1)$$

Deriving the expression for E with respect to “*a*”, and equating to zero, gives the optimum value for “*a*”:

$$a = \frac{\sum_i x_i x_i^{'} + \sum_i y_i y_i^{'}}{\sum_i x_i^2 + \sum_i y_i^2} \quad (2)$$

In the sequel, we term “*a*” the *similarity ratio*. The mean square error E is an absolute measure, thus it must be normalized to compare results for men and women, because the IU polygons of these two are of different area. Thus, the measure we propose is the *similarity error*, denoted by “*e*”, which we define as:

(3) $$e = \frac{\sqrt{E}}{\overline{R}}$$

Where $\overline{R}$ is the average Euclidean distance of all IU vowels from the polygon centroid.

II) Vowel Space "Skewedness": This is measured by the distance to which the IU centroid is shifted from the SS centroid. This raw measure overlooks two factors: (a) women's vowel space is larger than then that of men, and (b) reduction for men and women may not be identical. Therefore, an improved measure was obtained by dividing the distance by the specific similarity ratio for each gender, then normalizing by the ratios of the men-women vowel triangle areas.

III) Individual Vowel Reduction: While vowel space similarity, outlined above, indicates the degree of reduction of the entire vowel space, the reduction of specific vowels can be measured separately. This was calculated by the ratio of the Euclidean distances in IU divided by the Euclidean distance for SS, for each vowel independently.

IV) Increase in Spread: The error ellipses for vowels in SS have been found to be much larger than those obtained for IU. This was measured for each vowel, by finding the ratios between the areas of the error ellipses in IU versus SS.

3. RESULTS

Results are presented with reference to the values reported by Most et al. [3], as the prototypical values for citation-form Hebrew speech. Table 1 presents formant frequency averages and standard deviations of the present study. Figure 1 presents the vowel polygons for SS, with those of IU [3] included for reference.

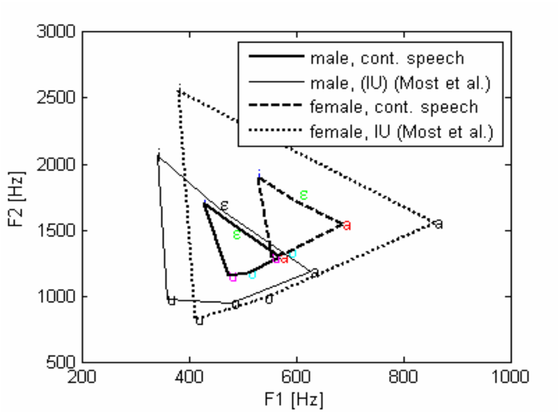
Table 1: Formant frequency averages (Hz) and SD for men (M) and women (W) in SS.

Formants	Gender	Vowels				
		/a/	/ɛ/	/i/	/o/	/u/
F1	M	569	491	427	509	473
		(93)	(91)	(91)	(101)	(99)
	W	688	597	529	586	554
		(132)	(126)	(114)	(132)	(121)
F2	M	1291	1514	1701	1175	1160
		(171)	(175)	(302)	(190)	(198)
	W	1548	1729	1903	1340	1297
		(194)	(257)	(419)	(278)	(256)

3.1. General Measures:

Triangle areas: The areas of the /i, a, u/ triangles were calculated for men and women, in IU and SS. Areas for men were 99,360Hz^2 and 20,463Hz^2, for IU and SS respectively. Areas for women were 294,420Hz^2 and 34,531Hz^2 for IU and SS, respectively.

Figure 1: The vowel polygons for SS and IU [4]



Square roots of the ratios of areas were calculated for IU and SS, yielding: 2.20 for men and 2.92 for women.

Similarity ratios: The scaling factors for achieving optimal similarity between the IU and SS polygons were 2.09 for men and 2.93 for women. A graphic depiction of the results of the similarity analysis is presented in Figure 1.

Similarity Errors: Square roots of the mean square error in similarity between IU and SS polygons, normalized by mean vowel Euclidian distance were 0.0717 and 0.0525 for men and women, respectively.

Centroid Offsets: The centroid offset between the IU and SS polygons were 41.97Hz (F1 axis) and 1.12Hz (F2 axis) for men, and 42.71Hz (F1 axis) and -62.13Hz (F2 axis) for women. This yields Euclidean distances of 41.99 and 75.39 Hz, respectively. The men SS vowel polygon shifted primarily in the positive F1 direction, aligning the /i- ɛ -a/ planes in IU and SS close to each other, as shown in Figure 5. In contrast, the women SS polygon shifted in the positive F1 and negative F2 directions, resulting in a seemingly "less skewed" shift for women, although the SS polygon centroid is further from the IU centroid for women than for men. Normalizing the Euclidean distance of these two vectors by their respective similarity ratios, and dividing the result for women by the square root of the ratio of the men/women vowel triangles yields a normalized offset distance of 20.11Hz for men and 14.54 Hz for women.

3.2. Individual Vowel Measures:

Euclidean Distance Ratios: The ratios of the Euclidean distances from the centroids in the IU and SS were calculated for each vowel independently. For lack of space we note only that the ratios between maximal and minimal reduction ratios were similar for both genders: 1.24 for men, 1.27 for women.

Formant Spacing: The C-D, G-A measures were calculated. Due to lack of space we note only that The C-D measures in SS tend to centralize towards the C-D value for the centroid. The C-D values for the IU and SS centroids are similar, because the centroids are close to each other. Yet, such finding was not observed for the G-A measure.

4. DISCUSSION

Results of the present study demonstrate vowel reduction in Hebrew, involving vowel centralization and an increase in variability. Interestingly, vowel reduction was manifested differently in men and women. Specifically, vowel triangle area ratios revealed a reduction in vowel space by a factor of 2.2 for men and by 2.92 for women. In addition, values for the similarity ratio were 2.09 for men and 2.93 for women. While these two measures are related, the similarity ratio measure might be considered advantageous, as it accounts for *all* vowels. Furthermore, it can be applied to vowel systems with any number of vowels, because it is based on a mean-squared-error criterion, applied to all vowels, rather merely on the three cardinal vowels (/i-a-u/).

The similarity error measure, introduced here, demonstrated the irregularity in the reduction in the men's vowels with an error 1.5 times larger than in women. Therefore, it is suggested that combining the similarity ratio measure with the similarity error measure provides valuable and complementary information.

In the present study, the area of the IU vowel triangle for men was 99360, whereas Blomgren et al [1] obtained a value of 200441, for five English-speaking men. While the present study was not aimed at comparing vowel systems of different languages, this preliminary observation suggests that the basic vowel triangle for American English is considerably larger than that of Modern Hebrew. Such comparison could provide a valuable measure for quantifying differences among vowel systems obtained in various languages.

The Euclidean distance ratios reflect the reduction of each vowel independently. Irregularity in vowel reduction was observed here too. Maximum reduction for men occurred in the low vowel /a/, and minimum reduction occurred in the back vowel /o/. In women, maximum reduction occurred in the front vowel /ɛ/, and minimal reduction occurred in the front-high vowel /i/. Nevertheless, the ratio between maximum and minimum reduction was similar for men and women. Visual inspection of the IU versus SS vowel polygons in Figure 1 suggests that reduction in women's speech is more symmetric than in men. Although the men's centroid for SS is relatively close to that of IU by means of absolute distance, appropriate normalization provides a measure of "skewness" that is 4/3 larger for men then for women.

Comparison of the data obtained in the IU and SS revealed that the centroids of both vowel polygons did not overlap. This result can be attributed to the fact that Hebrew does not have the central vowel, Schwa, as a phonemic vowel. Therefore, speakers are not aiming at a specific phonemic target. However, due to the novelty of this result, this should be further investigated in Hebrew, in comparison to other languages that include the Schwa phoneme.

Of the two formant-spacing measures, the C-D measure was more indicative. As it denotes F2-F1, vowel centralization in the F1 versus F2 plane causes this value to approach that of the centroid. Thus, all values for SS fall between the IU values and those of the centroid. Notably, no such marked reduction was found in the G-A measure.

5. SUMMARY

Past research presented limited measures of vowel reduction. Therefore, our aim was to define specific measures for reduction, which could be applied to vowel systems with different numbers of vowels. We trust that the measures described here could be used by others, to describe and quantify vowel reduction in different languages and in various reduction conditions; and to compare results among different languages.

6. REFERENCES

[1] Blomgren, M., Robb, M., Chen, Y. 1998. A note on vowel centralization in stuttering and nonstuttering individuals. *J. Speech Lang. Hearing Res.* 41, 1042-1051.

[2] Fourakis, M. 1991. Tempo, stress and vowel reduction in American English. . *J. Acoust. Soc. Am.* 90, 1816-1827.

[3] Most, T., Amir, O., Tobin, Y. 2000. The Hebrew vowel system: raw and normalized acoustic data. *Language and Speech* 43, 295-308.

[4] Izre'el, S., Rahav, G. 2004. The corpus of spoken Israeli Hebrew (CoSIH); Phase I: the pilot study. Proc LREC 2004.