ISCA Archive
http://www.isca-speech.org/archive

Afeka-AVIOS
Speech Processing Conference 2012
Tel Aviv, Israel
June 19-20, 2012

Quantifying Vowel Characteristics in Hebrew and Arabic

Noam Amir[1], Osnat Tzenker[1], Ofer Amir[1], Judith Rosenhouse[2]

[1] Dept. of Communication Disorders, Sackler Faculty of Medicine, Tel Aviv University
[2]Swantech Ltd.
Contact email: noama@post.tau.ac.il

Abstract. The dominant characteristics of spoken vowels are the two first formants. Thus the vowel systems of many different languages have been documented and compared through their F1-F2 space. Hebrew and Spoken Arabic, both Semitic languages, have five basic vowels: /i e a o u/, though Spoken Arabic has both short and long versions of each. In this paper we present an overview of the results of several studies on vowel formants in both languages. We first compare formants of isolated phonation and connected speech in Hebrew, showing how the formant space is reduced drastically in connected speech as compared to isolated phonation, and how in some cases a clear assimilation to surrounding context can be observed. We then present results on read speech in two Arabic dialects. We show the subtle differences between the two, and discuss the effect of vowel duration also. Finally we discuss how the results can be influenced by the research methodology, and comment on the relevance of the results to speech processing technologies.

1 Introduction

The first two formants, F1 and F2, are commonly acknowledged as the main carriers of information necessary for vowel identification. As such there have been many studies on this subject, examining different languages, speaker groups, speaking conditions etc. Such studies are labor intensive, requiring the recording and analysis of numerous utterances obtained from a large number of speakers. Many factors that influence the results need to be controlled: speaker age, speaker socioeconomic background, dialectical variations, speech rate, coarticulation effects and others.

Methodological differences between studies can have a large influence on the results in the F1-F2 vowel space. Even controlling for all other factors, articulation has a large influence on the results. For instance, it can be expected that uttering isolated vowels in a hyper-articulate manner, will give different results than having them uttered in carrier words, carrier sentences, or spontaneous speech. While it is not obvious which of the above represents the "true" vowel system of a language, it is actually of interest to compare the results from different such scenarios, in order to determine the factors that influence vowel articulation. Spontaneous speech, in particular, has several interesting characteristics. Due to the high speech rate found in spontaneous speech, the articulators can rarely attain their prototypical position, undergoing a process known as "vowel reduction". The characteristics of vowel reduction vary greatly, though in some languages this reduction is lexical. In English, for example, vowels far from the stressed syllable are pronounced as the centralized vowel "schwa".

In this paper we present a short summary of several studies on vowel characterization in Hebrew and Arabic, underlining the technical aspects of vowel analysis and the interaction between speech technology and the vowel system.

Hebrew has a five-vowel system, /i, e, a, o, u/ with no central "schwa" vowel. The most thorough investigation of the acoustics of the Hebrew vowel system to date was carried out by one of the present authors (Most et al., 2000). In that study, all vowels were uttered in nonsense CVC words with identical initial and final consonants. Our main objectives in the present study were therefore to observe how reduction is manifested in everyday Hebrew speech, and whether it is affected in stressed and unstressed vowels differently. More specifically, we wished to determine whether reduction is manifested as general centralization, or whether it is affected by neighboring consonants.

Arabic presents a much more complex picture than Hebrew. Literary Arabic (LA), on the one hand, is common to most of the Arabic speaking populations of the world. An interesting property of LA is that it is a "quantity language", i.e. vowel duration has a linguistic role. It has therefore three short vowels /i, a, u/, and three long vowels /I, A, U/. Colloquial Arabic (CA), on the other hand, has many different dialects, and traditionally it is not written. Though it is also a quantity language, it is difficult to determine how many distinct vowels it has, a number which can also differ from dialect to dialect. The convention for dialects spoken in Israel is that they have the same five basic vowels as Hebrew, but in short and long versions. There is a severe lack of phonetic study in this area, and therefore in this study we set out to document two of the dialects spoken in Israel. Specifically, we wished to characterize their vowel space, and to determine if duration was the only factor differentiating short and long versions of the same vowel.

2 Methodology

2.1 Data collection – Spontaneous Hebrew speech

Speech samples were taken from everyday speech production of 10 participants, five men and five women. Average ages were 39 (SD=11) for men and 28 (SD=3.7) for women, all of them native Hebrew speakers born in Israel. Seven of these were taken from the CoSIH corpus (Izreel and Rahav, 2004), a corpus of spontaneous everyday spoken Hebrew. Due to a lack of sufficient recordings in the above corpus with a quality good enough for acoustic analysis, three more recordings were carried out within this research, in similar conditions. A total of 5,582 vowel tokens were identified, 2,817 of them from the men's utterances and 2,765 from the women's. An experienced phonetician labeled each of the vowels as lexically stressed or unstressed. Table I shows the distribution over the different vowels, for both stressed and unstressed vowels.

Table I – distribution of the recorded vowels

		/i/	/e/	/a/	/o/	/u/
Men	stressed	288	353	539	248	96
	unstressed	260	244	509	152	76
Women	stressed	325	330	547	408	60
	unstressed	216	217	477	154	83

The distribution of vowel tokens in Table I is not balanced, reflecting their uneven distribution in spoken Israeli Hebrew (Bolozky 1999). Utterances were annotated using Praat software (www.praat.org). Recordings were segmented manually by an experienced phonetician and annotated as stressed or unstressed. Only monophthongs were taken, from content words only, and only from cases where no artifacts were present, such as laughter, chewing, etc. As the most frequent vowel was /a/, its occurrences were further subdivided into three categories for the stressed and unstressed cases separately, according to the place property of the preceding consonant: (1) Frontal; (2) Coronal; (3) Dorsal.

2.2 Data collection – Colloquial Arabic Speech

Two CA dialects were studied here: One is the dialect of Kfar Kassem/Kfar Barra/Jaljulia, in central Israel, also known as the "triangle" region, termed here as the Muthallath Dialect (MD). The second is the dialect of Majd Al-Kurum located in the Galilee, which we term the Galilean Dialect (GD). The participants were chosen according to geographical, social and communal (faith) criteria. They were all natives of one of these towns, all were Muslims, and were students or graduates of an academic institute, with no known hearing or speech problem. Twenty men and twenty women of each region participated in the study. Table II presents the average values of their ages and duration of academic education.

The five short and five long vowels /i, e, a, o, u, I, E, A, O, U/ of the two dialects were studied. For each of the two dialects the test material comprised three real-word lists, i.e., three different words per vowel, altogether 30 words per dialect. The lists included 24 CVC monosyllabic words, and six disyllabic words in the CVCVC pattern. These six words were necessary to provide examples of short /i /or /u/ vowels, which do not occur in the monosyllabic CVC pattern in these Arabic dialects.

The test words were inserted in a carrier sentence "I am reading the word .... which is written on the page." The central position of the test word was intended to prevent phrase final effects of lengthened vowels. This procedure yielded 1200 vowels (30 test words x 40 participants) per dialect, and 2400 vowels altogether.

2.3 Formant extraction

Automatic formant extraction based on source/filter modeling is relatively easy to perform using Praat or custom written software. However this process is notoriously error prone. Several attempts at automatic analysis of the data using Praat with different parameters revealed numerous errors. This is an interesting challenge for speech technologists, especially when the purpose is phonetic analysis, as opposed to speech coding applications which are less sensitive to such errors. It was finally decided to process the entire body of data

Table II - Average age and academic education (in years) of the two groups (SD values are in parentheses)

Category Group		Mean age (SD)	Mean academic education (years)
MD	Men	22.95 (2.48)	2.95
	Women	23.35 (2.57)	3.45
	Total	23.1 (2.53)	3.2
GD	Men	24.75 (2.8)	2.8
	Women	24.28 (2.57)	2.57
	Total	24.63 (2.68)	2.68

manually. A Matlab GUI-based program was written, which presented the experimenter with a comprehensive picture: the poles of the LPC model, the FFT spectrum of the vowel and the estimated spectrum based on the LPC model. Two degrees of freedom were available to the experimenter: LPC model order, and the sampling frequency to which the signal was downsampled. The experimenter attempted to select the highest LPC model before "formant splitting" occurred. Generally, a sampling frequency of 8 kHz was selected for men, and 11 kHz for women. After training, an experienced phonetic researcher could analyze approximately 250 tokens per hour. A screenshot of this GUI is shown in Figure 1.

3 Results

3.1 Vowels of Hebrew Spontaneous Speech

The full results of the analysis are beyond the scope of this paper. Instead, we present several qualitative results with possible implications for speech technology such as speech recognition and synthesis. Figures 2 (Men) and 3 (Women) show three cases of the F1-F2 formant space: for isolated utterances (Most et al. 2000), as well as for the stressed and unstressed vowels in the present study.

Centralization is apparent in spontaneous speech, though for men this is not as symmetrical as for women. For men there was mainly a heightening of F1 in the stressed vowels /i, u/ and a lowering in F1 only for unstressed /a/. For women the overall centralization is more apparent, with a similar lowering in F1 for unstressed /a/ as compared to stressed /a/.

Regarding women's vowel /a/ with different preceding consonants in Figure 4, it is apparent that the stressed syllables are less affected than the unstressed ones. A similar result was obtained for men. However, both the stressed and unstressed productions of /a/ were strongly influenced by the preceding consonant. Unstressed occurrences were more influenced than the stressed ones.

3.2 Vowels of Colloquial Arabic

Here we also present only our main results. A comprehensive picture of the F1-F2 vowel spaces for both genders and both dialects is shown in Figure 5. In each subplot, the external polygon is for the long vowels, while the internal polygon is for the short vowels. Several trends are apparent:

1. Short vowels are more centralized than long vowels, in all cases.
2. As expected, vowels spaces of women are similar in shape than those of men, but displaced in frequency, due to their shorter vowel tract.
3. Vowel spaces for long vowels are very similar across dialects.
4. Vowel spaces for short vowels are different across dialects, mainly in F1.

Durations of the vowels were also examined, as presented in Figure 6. Clearly, the short vowels are about half the duration of the long vowels, regardless of gender and dialect. However, durations of all vowels were not equal, with /a/ and /A/ tending to be longer than other vowels.

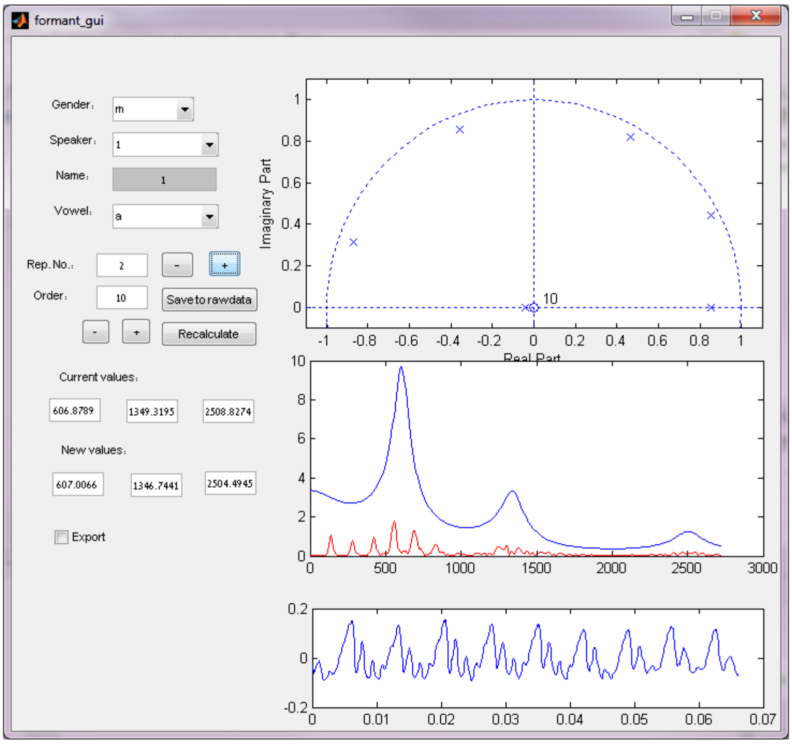

Figure 1 – Screenshot of the formant analysis GUI

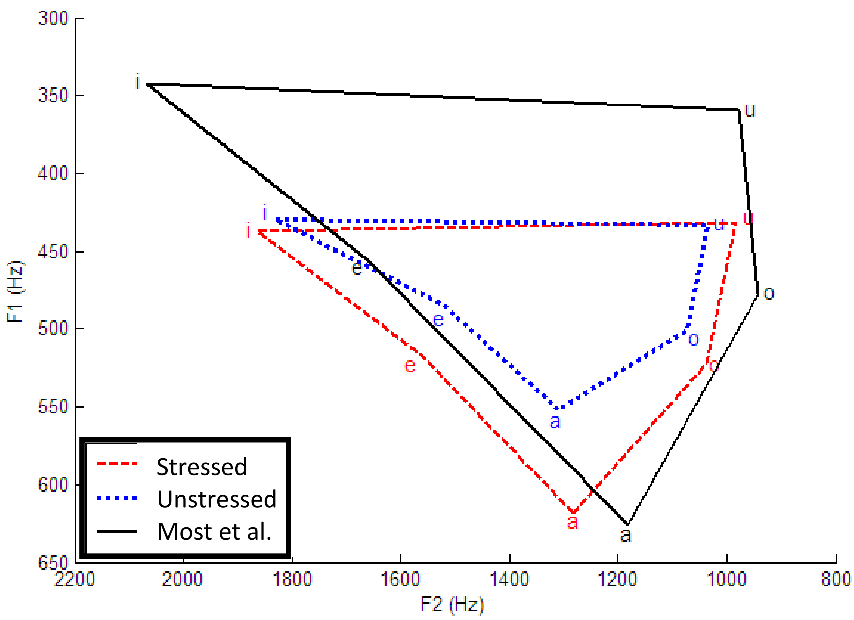

Figure 2 – F1-F2 vowel spaces for men in Hebrew

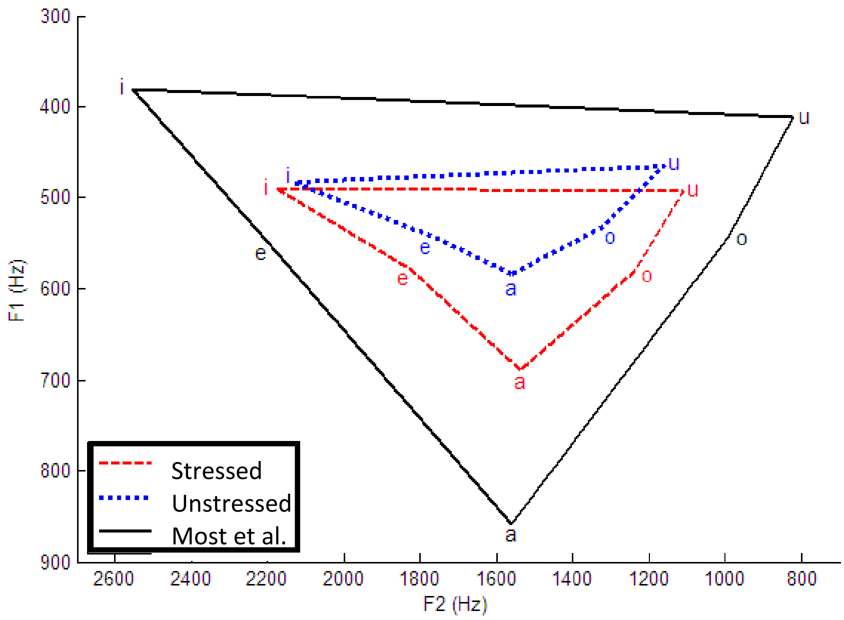

Figure 3 – F1-F2 vowel spaces for women in Hebrew

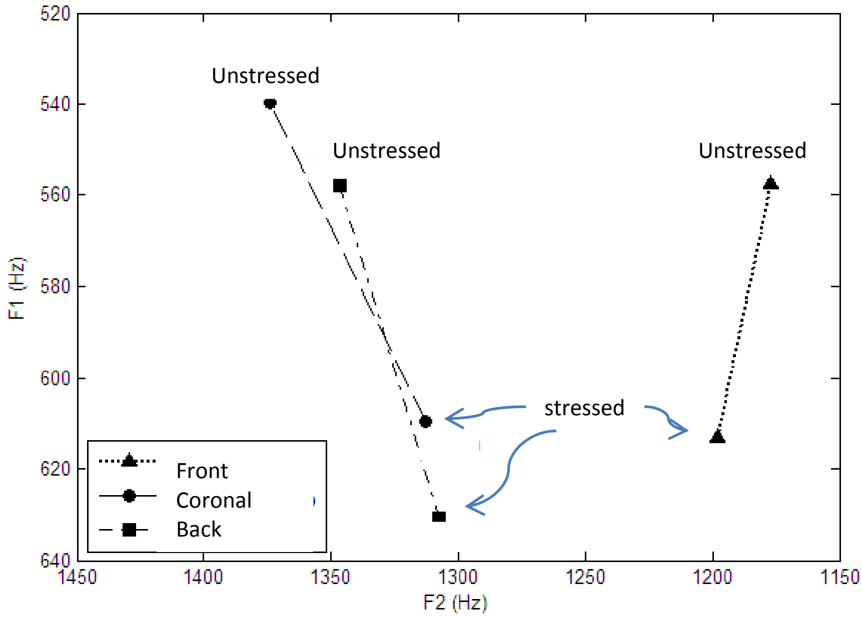

Figure 4 – F1-F2 values for women's Hebrew /a/ in stressed and unstressed positions, for three classes of preceding consonants.

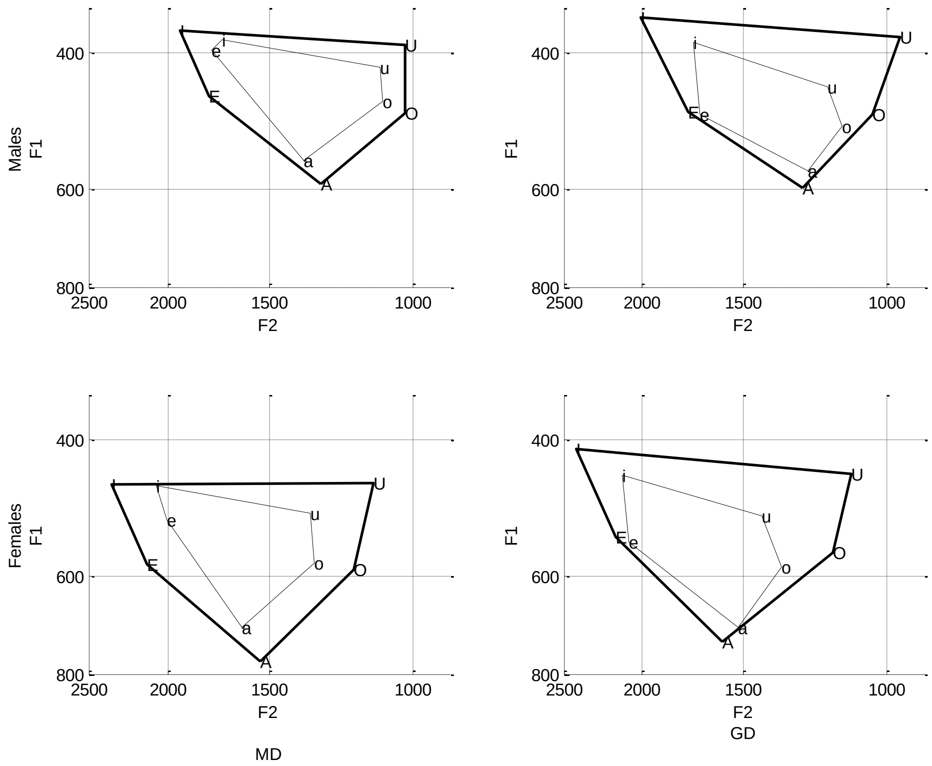

Figure 5- Vowel spaces of Colloquial Arabic for both genders and both dialects.

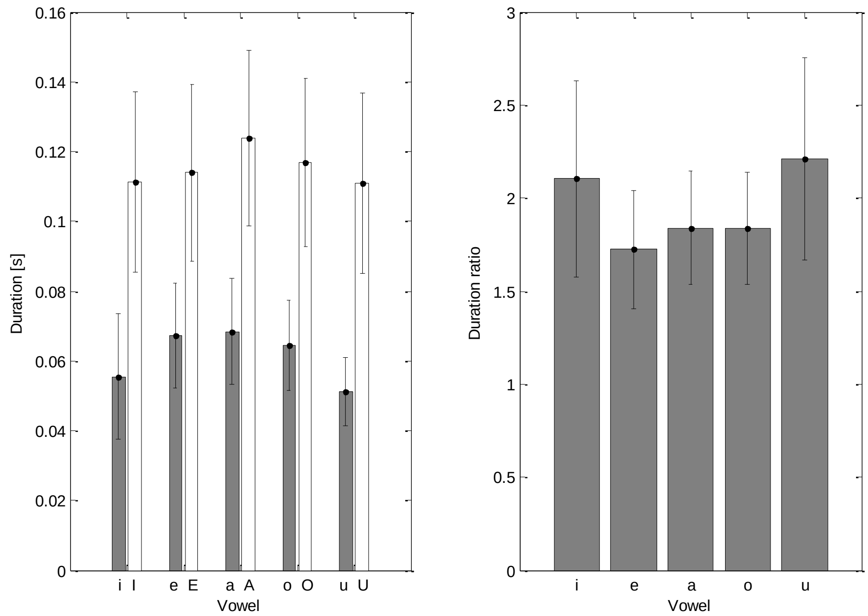

Figure 6- Duration of long and short vowel pairs (left) and long/short duration ratios (right)

4 Discussion

Several interesting observations have been made above concerning both experiments, with possible implications for speech technologies.

On the analysis side, it appears that formant extraction is still a labor intensive process that is not yet fully automated. It would be extremely helpful to phonetic study to develop methods that could automate this process further: both in automatic annotation of vowels, and ensuing extraction of the formant values. Analysis of large speech corpora is becoming more and more widespread, however without such tools it is a formidable task.

The analysis of spontaneous Hebrew speech shows that studies of isolated vowels are a very rough template for the actual occurrence of vowels in everyday speech. In such speech, the realization of vowels depends both on lexical stress and phonetic context. In speech synthesis, these effects must be taken into account if we wish to produce natural sounding speech. Presumably it must also be taken into account in speech recognition applications.

Analysis of two Arabic dialects revealed interesting contrasts between them. Though long vowels of both dialects were very similar, the short vowels were considerably different, mainly in F1. Thus, even in a small country like Israel we observe dialectical differences. It could be expected that larger differences will be found for dialects from more distant regions, leading us to the conjecture that there is possibly a large number of phonetically different Arabic dialects. Interestingly, the distinction between long and short vowels was found not only in duration, as the vowel space for the short vowels was found to be more centralized than that of the long vowels. Duration, however, remained a distinctive difference, with long vowels being approximately twice as long as short vowels. All of these findings are of major interest to any application of speech technology to the Arabic language.

Further research in both directions described here is certainly due. Such experiments have their own value as basic phonetic research: Advances in speech technology could certainly help carry out such research more efficiently. In turn, results from this type of research could potentially aid in improving the results of speech synthesis and recognition.

References

1. Most, T, Amir, O. & Tobin, Y.(2000). The Hebrew vowel system: Raw and normalized acoustic data. *Language and Speech*, 43(3), 295-308.
2. Izre'el, S. & Rahav, G. (2004). The corpus of Spoken Israeli Hebrew (CoSIH). Phase I: The Pilot Study, LREC

3. Bolozky, S., (1999). "On the Special Status of the Vowels a and e in Israeli Hebrew." *Hebrew Studies* 40. pp. 233-250.