Acoustic Variability of Vowels in Greek Spontaneous Speech

Katerina Nicolaidis
Aristotle University of Thessaloniki, Greece
E-mail: knicol@enl.auth.gr

ABSTRACT

To date numerous studies have investigated the acoustic characteristics of vowels produced in isolated words or other well-controlled types of speech material in several languages. However, relatively limited research has been carried out for vowels produced in spontaneous speech. In addition, relatively limited data are available for vowels in some languages, such as Greek. The aim of this study is to investigate acoustic variability during the production of the 5 Greek vowels /i, e, a, o, u/ in spontaneous speech. Data were recorded from two Greek male speakers producing a monologue. Formant frequency and duration measurements were made for all vowels produced in consonantal contexts. The study examines the extent of acoustic variability present in the data and some of the factors that contribute to such variability, including duration and context. Results show extensive variability for all vowels and the presence of overlapping formant distributions especially in the central part of the vowel space.

1. INTRODUCTION

One of the fundamental challenges of research in phonetics concerns the issue of variability in speech production. To date, most experimental work on articulatory and acoustic variability has been conducted on the basis of speech material that is relatively well-controlled. Limited research has been carried out on spontaneous speech despite its importance for the development of a comprehensive theory of speech production and for applications in speech technology.

Segment production in spontaneous speech is commonly characterized by large spatio-temporal variability. This is due to several factors including context, stress, location of a segment in the word/phrase, changes in rate of production, speaking style, speaker, discourse constraints that affect prosodic factors such as focus position, and so on. Segments may show different degrees of reduction and vary along a hyper-hypo speech continuum, i.e. from very reduced (hypo-forms) to non-reduced (hyper-forms). Such variability results from a continuous interplay between output constraints, ensuring successful communication, and system constraints, ensuring articulatory economy [1].

Previous research on variability in vowel production has looked into the effects of, among others, stress, context, and speaking rate, usually on the basis of well-controlled speech material. Variability in formant patterns was found to depend on (a) vowel duration, (b) contextual environment, and (c) articulatory effort [2]. Findings indicated that a decrease in duration may result in vowel reduction, i.e. in formant undershoot, due to an increase in contextual influences. Vowel reduction may or may not result in centralization depending on the identity of the vowel and consonantal environment involved, in particular the degree of the locus-target distance. In addition, undershoot may be compensated for by an increase articulatory movement velocity. Such a change in the style of movement may relate to a need for more articulatory precision in specific communicative situations.

Research focusing on the acoustic variability of vowels in spontaneous speech has been relatively limited. Koopmans-van Beinum [3] examined 12 Dutch vowels in several types of speech material ranging from the production of isolated vowels to free conversation. The author reported centralisation and a reduction of the vowel space when going from the more well-controlled types of speech material to free conversation. Moreover, Harmegnies & Poch-Olivé [4] report shifts in the formant frequencies of vowels produced in spontaneous speech relative to isolated words in Spanish. In particular, a strong tendency for centralization and larger intra-cluster variability was reported for vowels in spontaneous speech.

Limited research is also available on the acoustic characteristics of the Greek vowels (e.g. [5, 6]). Greek has a simple 5 vowel system consisting of the /i, e, a, o, u/ vowel categories. In a recent study, Fourakis et al. [6] provide a comprehensive analysis of the influence of factors such as stress, tempo and focus on several acoustic parameters. Shrinkage of the acoustic space is reported for the unstressed condition, together with a shift of the vowel space in a higher position in the F1x F2 acoustic space. A reduction of the vowel space is also reported in the fast tempo condition but is large only for the stressed vowels.

The aim of the current study is to examine variability during vowel production in Greek spontaneous speech. The study will report variability in vowel duration and F1, F2, F3 frequencies. It will then attempt to explain F1 and F2 variation as a function of several parameters including duration and context. The study is expected to add to the current knowledge available on the variability of the acoustic parameters of vowels as well as on the acoustic characteristics of Greek vowels. To our knowledge, this is the first study that examines variability of Greek vowels produced in spontaneous speech.

 ISBN 1-876346-nn-n

2. METHODOLOGY

Data were recorded from two Greek male speakers (CN and JM) producing a monologue on a subject of their interest. In total, 2.5 min for speaker JM and 1.2 min for CN were recorded. If pauses and filled pauses are excluded, approximately 3 min of speech were analysed and data on approximately half of the vowels produced are reported in this study. The subjects were 29 and 30 years of age and spoke standard Modern Greek. The recordings were made in an acoustically damped studio at the Speech Laboratory at Reading University, UK, using a high quality cardioid studio condenser microphone, Sennheiser MKH 40 P48. Data were acquired with the Reading University-IBM multichannel acquisition system which was used to record simultaneously acoustic, electropalatographic and electrolaryngographic data. Acoustic data were sampled at 20,000 Hz.

Data were analysed using the PRAAT speech analysis system. In particular, duration and F1, F2 and F3 measurements were made for all vowels produced in consonantal context; diphthongs and vowels preceded or followed by other vowels were excluded from the analysis. Vocalic sounds produced in dysfluent parts of the data, e.g. in filled pauses (usually [e:]) and hesitations, were also excluded from the analysis. Duration measurements were taken from the onset till the end of the vowel formant structure (and the end of voicing when there was no carryover voicing present during the next consonant). Formant measurements were made from wideband spectrograms at the temporal midpoint of the vowel. Measurements were based on the formant tracking procedure available in the system, manually checked for accuracy for all vowels, as well as on FFT derived spectra at the mid of each vowel. For the present study 302 vowels were analysed. The frequency of the occurrence of the 5 vowels in the data was as follows: /i/=73, /e/=62, /a/=103, /o/=52, /u/=12. The highest frequency was thus evident for the /a/ vowel and the lowest for the /u/. A phonemic and narrow phonetic transcription was made of all the data and vowels were categorized into stressed/unstressed, in focus/nonfocus position, and in final/non final position in the tone unit. Table 1 shows the number of vowels produced in the different categories identified.

The statistical analysis involved the use of regression. The analysis was carried out for the vowels [i, e, a, o] because there was a sufficient number of tokens available. Formant frequencies (F2 and F1) were regressed on the duration of the vowel, the subject, and a set of dummy variables describing the consonantal context. In particular, a set of a total of eight variables for labial, alveolar, palatal and velar contexts, preceding and following the vowel, were included in the model. Initially the general model was estimated; subsequently using t-tests the insignificant variables were gradually excluded. The final reduced-form model, which includes only the significant variables, is reported in this study. Other important variables, e.g. stress, focus, boundary, were not included in the model as they could make estimation problematic given (a) the number of observations available, (b) the correlation of these variables with duration. Two further regressions were carried out in order to investigate the (a) relationship between duration and stress, focus, boundary, (b) duration differences between the five vowels.

	i	e	a	o	u
Unstressed	32	40	74	33	5
Unstressed-boundary	5	3	12	2	0
stressed-nonfocus	19	13	13	10	3
stressed-focus	9	3	2	5	3
stressed -nonfocus-boundary	1	1	0	0	0
stressed-focus-boundary	7	2	2	2	1
Total	**73**	**62**	**103**	**52**	**12**

Table 1: Number of tokens analysed per vowel category. The vowels are categorized (a) into stressed/unstressed, (b) in focus/nonfocus position and (c) in final/nonfinal position in the tone unit (final position is marked with the indication 'boundary').

3. RESULTS

Duration measurements: The regression analysis showed significant differences in the durations of the vowels [F(4,297)=4.47, p<0,0015]. The means show that the open /a/ vowel was produced with the longest duration while the two high vowels /i/ and /u/ with the shortest (see Table 2). These findings are in agreement with the results reported in Fourakis et al. [6], Dauer [7] and Fourakis [8] for the Greek vowels. One difference however between the current and previous studies concerns the duration of the high vowels. In Dauer [7] and Fourakis [8], /i/ and /u/ were reported to be of similar durations while in Fourakis et al. [6] /u/ was reported to be longer than /i/. Our data show shorter duration for /u/ than /i/; this finding needs however to be considered with caution since it is based on very limited data available for the back vowel.

	DUR	*F1*	*F2*	*F3*
i	69.1 (32.1)	360.3 (47)	1892.3 (184.4)	2492.8 (145.4)
e	80.9 (31.7)	475.3 (65.2)	1671.6 (137.2)	2377.1 (178.7)
a	85.4 (25.9)	575 (73.9)	1452.7 (134)	2201.9 (193.4)
o	78.4 (33.5)	462 (58.5)	1202 (138.9)	2207.3 (214.1)
u	59.8 (14.9)	376.7 (54.6)	1163.1 (157.6)	2290.3 (196.9)

Table 2. Means and standard deviations (in parentheses) for the duration, F1, F2, and F3 frequencies of the vowels /i, e, a, o, u/.

In addition, the regression analysis showed a significant relationship between duration and stress, focus, boundary [F (3, 298)=50.6, p<0.000], i.e. presence of these three variables increased duration significantly. Figure 2 plots mean durations for each vowel at the different stress, focus and boundary conditions (the graph excludes data for which there was only one token available, see table 1). The graph shows a general increase in duration from the unstressed to the stressed (nonfocus and focus), to the boundary conditions (unstressed and stressed). A very large increase in duration is evident in the boundary condition.

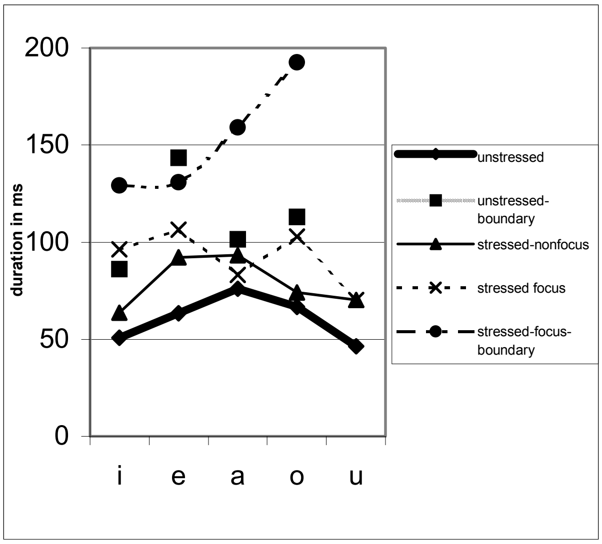

Figure 2. Mean durations of the five vowels in the different, stress, focus and boundary conditions.

Formant frequency measurements: The F1 and F2 frequencies of all vowel tokens are plotted in Figure 3. The graph shows extensive variability for all vowels and the presence of overlapping formant distributions. While there is good differentiation among the three point vowels /i, a, u/, overlap is evident between these and the mid vowels /e/ and /o/. There is extensive overlap in the center of the vowel space and a general merging of vowel formant distributions which results in less differentiated vowels.

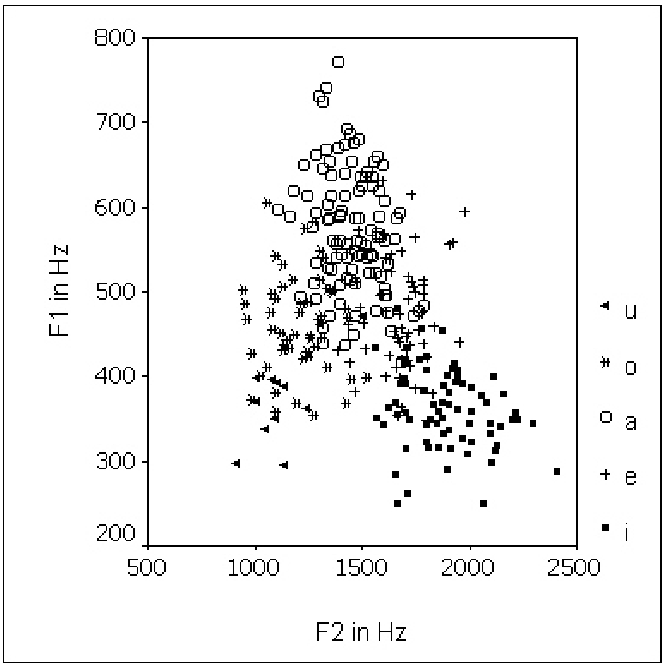

Figure 3. Greek vowels produced in spontaneous speech plotted in an F1 X F2 space.

Variability is evident both in F1 and F2 resulting in a considerable spread of the frequency values for all vowel categories and considerable overlap especially between the open vowel /a/ and the mid vowels /e/, /o/ at the center of the vowel space. The means and standard deviations of the first three formants for all vowels are presented in Table 2.

The results of the regression analysis examining the influence of duration, subject, and context on the F1 and F2 frequencies of the vowels /i, e, a, o/ showed a significant effect of duration on the F1 and F2 frequencies of the four vowels (see table 3) (a regression analysis was not carried out for /u/ due to the lack of sufficient observations). The only exception was for the F2 frequency of the vowel /a/. The shifts in frequency values as a result of a change in duration are shown in Figure 4. The graph plots the five vowels, using mean values, and the horizontal and vertical arrows indicate the direction of F1 and F2 change when the duration of the vowels decreases (based on the regression coefficients).

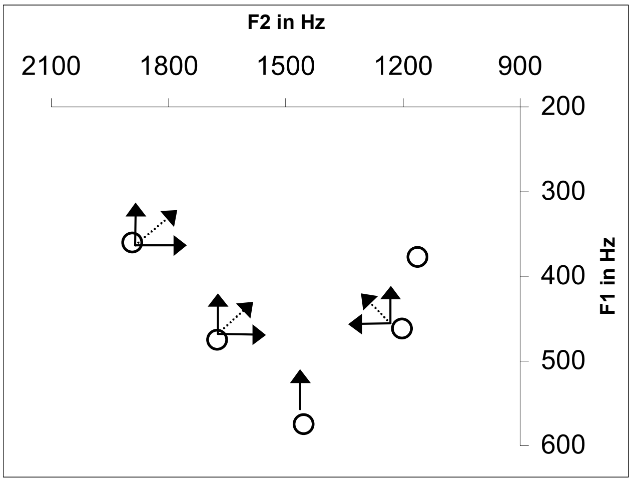

Figure 4: The Greek vowels plotted in an F1 X F2 space (mean formant frequencies used). The arrows show the influence of a decrease in duration on F1 and F2 frequencies.

Mid arrows were included to indicate the approximate composite direction of movement of the vowel in the F1 X F2 space. For all four vowels, a decrease in F1 is evident as a result of a decrease in duration suggesting an upward shift of the vowel space. In addition, a decrease in F2 frequency occurs for the front vowels and an increase for the back /o/. This indicates movement towards a more central part of the vowel space for /e/ and /o/, which is addition to the F1 shift for /a/, results mainly in the open and mid vowels coming closer together in the center of the vowel space.

With reference to the subject variable, significant speaker differences in formant frequency values were evident only for the point vowels /i/ and /a/. Finally, the influence of the context revealed significant effects on the vowel F1 and F2 frequencies. As is evident in Table 3, significant effects were more commonly found from the preceding consonant at the vowel midpoint. For the front vowels /i, e/, preceding consonants produced at the bilabial (b1), alveolar (a1), and

 ISBN 1-876346-nn-n © 2003 UAB

palatal (p1) places of articulation caused a significant reduction of the second formant frequency at the vowel midpoint (negative regression coefficient). A similar effect is shown for the /a/ vowel in a bilabial context, while a significant increase in F2 frequency at the /a/ vowel midpoint was evident in the palatal and velar (v1) environments. For the open vowel there was also a significant effect from the following consonantal context with a significant decrease in F2 frequency in the environment of a following bilabial (b2) consonant. For the /o/ vowel a significant increase in F2 frequency was evident in the environment of a preceding alveolar consonant. Significant context effects on F1 were more limited but also evident mainly from the preceding than the following consonant with the exception of /a/. No significant results were found from the following alveolar, palatal and velar environments on any of the vowels.

	CONS	DUR	SUB	B1	A1	P1	V1	B2
/i/ F2 r^2=0.53	2084 27.7	1.81 3.54	-203.2 -6.3	-172 -2.2	-246 -3.6	-138.6 -1.8	*	
/i/ F1 r^2=0.33	294 17.7	0.4 2.6	44.4 4.6		21.4 1.7		*	
/e/ F2 r^2=0.28	1791 25.2	0.93 1.9		-265 -4.04	-180 -2.8	-161.7 -2.32	*	
/e/ F1 r^2=0.19	519.6 14.5	0.48 1.95		-81.9 -2.5	-82.4 -2.5	-113.6 -3.2	*	
/a/ F2 r^2=0.41	1463 36.7		-70.8 -3.2	-74.3 -1.7		239.2 5.4	73 1.6	-97.5 -3.7
/a/ F1 r^2=0.32	413.3 15.1	1.5 6.0	19.5 1.5					67.9 3.4
/o/ F2 r^2=0.41	1192 25.9	-1.2 -2.7			158.6 4.9			
/o/ F1 r^2=0.10	418.7 21.1	0.6 2.4						

*Table 3. Results of the regression analysis of F1 and F2 on duration, subject and context. (CONS: intercept; b=bilabial, a=alveolar, p=palatal, v=velar, 1=preceding environment, 2=following environment; * indicates absence of tokens in particular environment). Left column shows the dependent variable and the r^2 value. The top number in each cell is the value of the estimated coefficient and the bottom one is the t-ratio.*

4. DISCUSSION

Our formant frequency data for the Greek vowels produced in spontaneous speech have shown lack of a clear differentiation between vowel categories in the F1 X F2 acoustic space. Instead extensive variability and overlapping formant distributions were evident. Similar results have been reported for the five Spanish vowels produced in spontaneous speech in [4]. In our data, extensive overlap was evident in the central part of the vowel space between the open vowel /a/ and the mid vowels /e/ and /o/. Since it is known that formant displacement is duration and context dependent, an attempt was made to examine the effect of duration and context on formant variability. A decrease in duration was found to cause a significant change in formant frequencies, which resulted in a shift of the vowel space towards a higher location in the F1 X F2 space (for the four vowel categories examined). Fourakis et al [6] also report a similar shift in the vowel space of the unstressed vowels in comparison to stressed nonfocus vowels both in slow and fast speech. A decrease of F1 for all vowels thus suggests a higher tongue position when duration decreases. This may result in centralisation for vowels like /a/ but not for /i/. In addition, the decrease in F2 for /e/ and increase for /o/ indicates approximation of these vowel categories, which together with the decrease of F1 for /a/, results in extensive overlap between the three vowel categories at the center of the vowel space.

Significant context effects were also found in spontaneous speech. It is interesting that these were mainly from the preceding consonantal environment suggesting the presence of carryover influence rather than anticipatory on the vowel midpoint. More significant effects were also found for the F2 than F1 frequency from a variety of consonantal contexts suggesting more systematic contextual effects along the front-back axis. This may to some degree relate to constraints (articulatory and perceptual) that are imposed on tongue height variability due to the overall shift of the vowel space towards a higher position when there is a decrease in duration.

In conclusion, since the data sample in this study was relatively small and some vowels were under-represented in the sample, an analysis of more vowel data from more speakers and a comparison of formant displacements in different speaking styles is necessary for a comprehensive study of variability in vowel production.

REFERENCES

[1] B. Lindblom, "Explaining phonetic variation: a sketch of the H&H theory," in *Speech Production and Speech Modelling*, W. Hardcastle and A. Marchal, Eds, pp. 403–439. Dordrecht: Kluwer, 1990.

[2] S-J. Moon and B. Lindblom, "Interaction between duration, context, and speaking style in English stressed vowels," *Journal of the Acoustical Society of America*, vol. 96(1), pp. 40-55, 1994.

[3] F.J. Koopmans,-van Beinum, *Vowel Contrast Reduction, An acoustic and Perceptual study of Dutch Vowels in Various Speech Conditions*, Ph.D. thesis, University of Amsterdam, 1980.

[4] B. Harmegnies and D. Poch-Olivé, "A study of style-induced vowel variability: laboratory versus spontaneous speech in Spanish," *Speech Communication*, vol. 11, pp. 429-437, 1992.

[5] A. Jongman, M. Fourakis and J.A. Sereno, "The acoustic vowel space of Modern Greek and German," *Language and Speech,* vol. 32(3), pp. 221-248, 1989.

[6] M. Fourakis, A. Botinis and M. Katsaiti, "Acoustic characteristics of Greek vowels," *Phonetica*, vol. 56, pp. 28-43, 1999.

[7] R. Dauer, *Stress and Rhythm in Modern Greek*. Ph.D. dissertation, University of Edinburgh, 1980.

[8] M. Fourakis, "An acoustic study of the effects of tempo and stress on segmental intervals in Modern Greek," *Phonetica*, vol. 43, pp. 172-188, 1986.