The Journal of the Acoustical Society of America

MARCH 18 2024

The acoustics of Contemporary Standard Bulgarian vowels: A corpus study

Mitko Sabev ; Bistra Andreeva

J. Acoust. Soc. Am. 155, 2128–2138 (2024)
https://doi.org/10.1121/10.0025293

JASA ARTICLE

The acoustics of Contemporary Standard Bulgarian vowels: A corpus study

Mitko Sabev[a)] and Bistra Andreeva
Department of Language Science and Technology, Saarland University, Saarbrücken, Germany

ABSTRACT:
A comprehensive examination of the acoustics of Contemporary Standard Bulgarian vowels is lacking to date, and this article aims to fill that gap. Six acoustic variables—the first three formant frequencies, duration, mean f_0, and mean intensity—of 11 615 vowel tokens from 140 speakers were analysed using linear mixed models, multivariate analysis of variance, and linear discriminant analysis. The vowel system, which comprises six phonemes in stressed position, [ɛ a ɔ i ɤ u], was examined from four angles. First, vowels in pretonic syllables were compared to other unstressed vowels, and no spectral or durational differences were found, contrary to an oft-repeated claim that pretonic vowels reduce less. Second, comparisons of stressed and unstressed vowels revealed significant differences in all six variables for the non-high vowels [ɛ a ɔ]. No spectral or durational differences were found in [i ɤ u], which disproves another received view that high vowels are lowered when unstressed. Third, non-high vowels were compared with their high counterparts; the height contrast was completely neutralized in unstressed [a-ɤ] and [ɔ-u] while [ɛ-i] remained distinct. Last, the acoustic correlates of vowel contrasts were examined, and it was demonstrated that only F_1, F_2 frequencies and duration were systematically employed in differentiating vowel phonemes.
 https://doi.org/10.1121/10.0025293

(Received 15 November 2023; revised 26 February 2024; accepted 27 February 2024; published online 18 March 2024)
[Editor: Jody Kreiman] Pages: 2128–2138

I. INTRODUCTION

This article reports the results of a comprehensive and multifaceted acoustic investigation of the Contemporary Standard Bulgarian (CSB) vowel system. The aim is to identify and assess the acoustic correlates of word stress and vowel contrasts as well as to determine the prosodic contexts and phonetic mechanisms of height neutralization.

Bulgarian is a member of the Southern branch of the Slavonic language family and is the official language in Bulgaria. It is also spoken by a large diaspora abroad. Typologically, Bulgarian belongs to the group of "stress languages"[1] or "intonation languages,"[2] where pitch variation is used for a range of functions such as disambiguating syntactic structures, signalling grammatical distinctions (e.g., statements vs questions), indicating emotional states and attitudes, highlighting important parts of the message, and regulating conversational interaction. In terms of rhythm, Bulgarian occupies an intermediate position on the stress- to syllable-timed continuum and has, thus, been characterized as being of a "mixed" type.[3–5]

The CSB stressed vowel system consists of six contrastive vowels, which phonetically range from high [i u], to mid [ɛ ɤ ɔ], to low [a]. Based on the assumption that—at least in some Bulgarian dialects—the non-high [ɛ a ɔ] are raised in unstressed position and merge with their higher counterparts [i ɤ u], respectively, both Trubetzkoy[6] and Jakobson[7] argued that there were only two contrastive, phonological vowel heights in Bulgarian, as schematized in Fig. 1. This two-height system has generally been adopted in the literature published in Bulgarian and is also retained in this article. However, the received view of vowel reduction in the Bulgarian literature, most thoroughly expounded in the "Academy Grammar,"[8] does not corroborate the assumption that unstressed [ɛ a ɔ] reduce to [i ɤ u]. Instead, the Academy Grammar maintains that only the non-front unstressed pairs, [a-ɤ] and [ɔ-u], may merge in Standard Bulgarian, and that [a-ɤ] are more likely to merge than [ɔ-u]. It is also claimed that the neutralized unstressed qualities are not those of the higher vowel in each pair but rather realizations of intermediate heights, such as [ʌ o]. In other words, not only are non-high vowels raised, but high vowels are also lowered in unstressed position. Another received view of Bulgarian vowel reduction is that there are, in fact, two distinct degrees of reduction, for vowels in first pretonic syllables, i.e., syllables immediately preceding the stressed syllable, are claimed to be more open than other unstressed realizations. Note that this is assumed to apply to high vowels as well, which is an assumption that is left unexplained in the traditional literature. The Academy Grammar has been very influential to the present day, and these received views of Bulgarian vowel reduction have often been repeated or confirmed in various subsequent publications.[9–12]

Eastern Bulgarian accents are characterized by certain phonetic and phonological phenomena that are considered non-standard, two of which are particularly salient in native speakers' sociolinguistic perceptions and attitudes. First, in

[a)]Email: msabev@lst.uni-saarland.de

high(er)	i	ɤ	u
non-high	ɛ	a	ɔ

FIG. 1. The CSB stressed vowel system.

addition to the reduction of [a] and [ɔ], Eastern Bulgarian may also reduce unstressed [ɛ] to [i]. Second, consonants are regularly allophonically palatalized before front vowels. Both of these phenomena are strongly stigmatized and neither of them occurs in Standard, and generally Western, Bulgarian.[13,14] However, although Standard Bulgarian historically sides with Western Bulgarian accents in terms of not allowing neutralizing [ɛ]-reduction or allophonic palatalization, like any standardized variety, CSB, to a large extent, has a non-regional status, and its speakers come both from the West and the East of Bulgaria. Geographical origin alone, therefore, is not a sufficient predictor of accent.

The Academy Grammar incorporates an acoustic examination of Standard Bulgarian vowels and word stress. Bulgarian word stress is described as "free dynamic accent." It is free because its location can be contrastive, as in [ˈvɤlna] 'wool' vs [vɤlˈna] 'wave,' or [ˈtʃɛtɛ] 'read-3sg.AOR' vs [tʃɛˈtɛ] 'read-3sg.PRES.' It is dynamic (rather than melodic) accent as it is phonetically realized through relative intensity, duration, pitch, and vowel quality. Intensity is identified as the main factor for the perception of word stress, although it is also cautioned that higher intensity alone is not necessarily sufficient to distinguish stressed from unstressed syllables, and increased duration and pitch are often employed as well, since these phonetic resources, not being lexically contrastive, are readily available to serve as cues to prominence. It is further stated that "unstressed vowels have unclear articulation and undergo quantitative and qualitative reduction" (see Ref. 8, p. 162). Along with higher intensity, stressed vowels are also described as having generally higher fundamental frequencies and longer durations. At the same time, it is maintained that prosodic factors, as well as the fact that vowels have different intrinsic intensities and fundamental frequencies, may result in an unstressed vowel having a higher value for those variables than the stressed vowel in the same word. The non-high vowels [ɛ a ɔ] are described as having higher intrinsic intensity than the high [i ɤ u], where [a] is loudest and [i] is least loud. With regard to intrinsic f_0, the six vowels are ranked from lowest to highest frequency as follows: [a ɔ ɛ ɤ i u] (see Ref. 8, pp. 160–162).

The Grammar reports F_1, F_2, and F_3 frequencies for each vowel in various positions. For [a] and [ɔ], there appears to be no F_1 difference between stressed and first pretonic realizations, which is at variance with the claim of two degrees of reduction: pretonic vowels appear not to reduce at all. The F_1 frequency given for unstressed non-pretonic [ɛ] is, in fact, higher than that for stressed [ɛ], which is probably an error. No formant frequencies are reported for [ɤ] and [i] in unstressed positions (see Ref. 8, pp. 32–59).

Although the Academy Grammar describes acoustic variables that are highly relevant to word stress and vowel contrasts, from a modern perspective, the volume suffers from a number of methodological and presentational flaws. It is often unclear whether mean values or single measurements are provided. The reported formant frequencies are, in all likelihood, based on data collected from only three subjects in the late 1960s for the doctoral thesis of the first author, D. Tilkov.[15] The subjects were the author himself, another male, and a female speaker. The formant frequencies published in the Academy Grammar are incomplete and sometimes implausible. All acoustic variables are presented as raw values without any normalization, outlier treatment, or other statistical analysis. This means that it is impossible to establish the significance of any apparent differences or estimate the relative importance of each acoustic variable in implementing a particular distinction between vowel groups (stressed vs unstressed, high vs non-high, etc.).

A series of more recent publications have been devoted to Bulgarian vowel reduction and have challenged or refuted many of the received views upheld in the Academy Grammar. One corpus study[16] of speech read by 20 CSB speakers found no evidence of unstressed high vowel lowering, nor of [a-ɤ] being more likely to merge than [ɔ-u], while confirming that unstressed [ɛ-i] do not merge in CSB. These findings were corroborated in an ultrasound and acoustic investigation[17] of the speech of three male informants. A more extensive acoustic study[14,18] of careful speech (highly controlled nonsense words in carrier sentences) read by 12 Western (CSB) and 8 Eastern Bulgarian speakers also found no lowering in unstressed high vowels and confirmed that unstressed [ɛ-i] do merge in Eastern but not in Western Bulgarian. The study also found that, contrary to the received view, Western Bulgarian unstressed [ɔ-u] underwent greater contrast reduction than unstressed [a-ɤ], and vowels had significantly *higher* realizations in first pretonic than in other unstressed syllables in the CSB, whereas in Eastern Bulgarian, the various unstressed positions were undifferentiated in vowel height. A short paper in recent conference proceedings,[19] which is based on the same speech corpus as the present article, found no evidence of two degrees of reduction or unstressed high vowel lowering and also demonstrated that [a-ɤ] and [ɔ-u] are completely merged in unstressed position.

All recent experimental work on Bulgarian vowel reduction explores only the first two formant frequencies and duration. Although these variables are central to the study of vowel reduction, there are other acoustic parameters that may play an important role in the implementation of word stress and vowel contrasts, namely f_0, intensity, and F_3 frequency. In terms of vowel contrast, work on reduction has naturally focused on the neutralization or preservation of phonological height. No previous study has looked into the acoustic correlates of the other vowel contrasts: backness and roundness. This article reports an extensive, up-to-date examination of the first three formant frequencies, duration, mean f_0, and mean intensity of vowels in a large corpus of speech read by 140 CSB speakers and specifically addresses the following topics.

A. First pretonic vs other unstressed vowels

Based on recent findings, we expect to discover no spectral or durational differences between these two types of unstressed vowels. We cannot rule out, however, differences in mean f_0 or mean intensity, which may be related to higher-level prosodic effects.

B. Stressed vs unstressed vowels

Consistent with previous work, our prediction is that in unstressed syllables, non-high vowels undergo raising (F_1 frequency reduction) and changes in F_2 frequency will reveal some centralization. Whether stress affects F_3 frequency in any of the vowels is yet to be established. Stress-related differences in duration, mean f_0, and mean intensity may also be expected as per earlier findings and claims.

C. Height neutralization

It has been demonstrated that the vowels in each of the unstressed pairs [a-ɤ] and [ɔ-u] are merged completely, both acoustically[19] and perceptually,[20] and we have no reason to expect to find significant differences in the variables that have not been previously examined: F_3 frequency, mean f_0, and mean intensity.

D. Acoustic correlates of vowel contrasts

Although it has been demonstrated that CSB non-high and corresponding high vowels differ primarily in F_1 frequency and sometimes also in F_2 frequency and duration,[14,19] further significant differences may be found in F_3 frequency, as well as mean f_0 and mean intensity, in accordance with the intrinsic fundamental frequency and intensity of vowels that have been reported for Bulgarian[8] and other languages.[21,22] The "timbre" or "colour" contrasts, on the other hand, have not been closely studied, so far, with modern experimental methods. Based on research on other languages and general acoustic phonetics, one may expect F_2 frequency to be affected by backness and roundness and, possibly, F_3 frequency to vary with roundness. Once again, significant differences in mean f_0 and mean intensity may emerge.

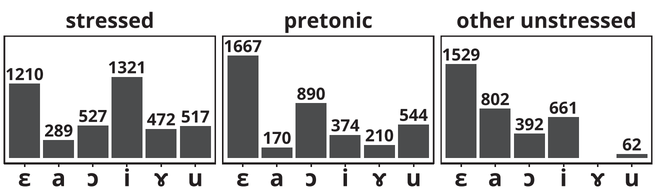

FIG. 2. Number of vowel tokens by phoneme and position.

II. METHODOLOGY

The material analysed is continuous read speech from the Bulgarian phonetic corpus *BulPhonC*, version 3, consisting of 319 phonetically rich sentences.[23] The corpus was designed for the development of automatic speech recognition technology. The recordings were made in an echo-cancelling studio with a Sennheiser MK 4 omnidirectional microphone (Wedemark, Germany) on a TASCAM DP32 digital recorder (Santa Fe Springs, CA) at a sampling rate of 48 kHz and 24 bits, filtered and down-sampled to 16 kHz. Canonical transcriptions and automatic phoneme segmentation are available.[24] We used a subset of the data containing the vowels in open syllables, read by 140 speakers (81 female and 59 male). The mean speaker age was 37 years [standard deviation (SD), 16; median, 30]. The vast majority of the speakers were representative of CSB according to the judgments of M.S. and B.A., both of whom are native CSB speakers. Approximately 10% of the speakers had subtle eastern traits, which the authors deemed not to be relevant to the investigation. Regional variation was, therefore, largely controlled for, and that is borne out by the results reported in Sec. III: as will be seen, the findings are in keeping with what is to be expected for Western and Standard—but not Eastern—Bulgarian.[8,13,14]

Praat[25] scripts were used to measure vowel duration, F_1, F_2, and F_3 frequencies at vowel midpoint, as well as mean f_0 and mean-energy intensity over full vocalic intervals. Formant measurement was conducted with Praat's Burg algorithm with a maximum of five formants, window size of 0.025 s, pre-emphasis from 50 Hz, and maximum formant thresholds of 5000 Hz (male) and 5500 Hz (female speakers). Mean f_0 was measured using Praat's autocorrelation algorithm with a range of 50–550 Hz.

The values measured for all six acoustic variables were normalized using Lobanov's speaker-intrinsic, vowel-extrinsic *z*-transformation method.[26] All reported results are based on normalized values. Outliers by vowel and stress condition, defined as values outside the interquartile range (IQR) by 1.5 times IQR, were removed (3.27% for F_1, 3.69% for F_2, 3.12% for F_3 frequency, 2.29% for duration, 2.28% for mean f_0, and 1.29% for mean intensity). The acoustic variables were compared across three positions at first: stressed, (first) pretonic, and other unstressed positions. In total, 11 637 vowel tokens were analysed; Fig. 2 shows the numbers by vowel and position. There were no instances of [ɤ] in the "other unstressed" position. This gap in the data reflects the phoneme's low frequency and morphological distribution in the Bulgarian language in general,[14,18] rather than being an idiosyncrasy of the corpus.

A. First pretonic vs other unstressed vowels

To address the first question of interest, whether vowels have lower realizations in first pretonic than in other unstressed syllables, linear mixed models (LMMs) were constructed for each vowel, where each acoustic variable was the response and position (pretonic vs other unstressed) was the predictor variable. As the Lobanov normalization method factors out physiologically induced acoustic variation while also retaining sociolinguistic differences,[27] speaker was included as a random effect along with phonological context (adjacent consonants) and word length (in syllables).

B. Stressed vs unstressed vowels

To identify the magnitude, dimensions, and direction of stress-conditioned changes, three tests were performed for each vowel. First, LMMs were fitted for each variable with stress (stressed vs unstressed) as the predictor and the same random effects as above. Second, multivariate analysis of variance (MANOVA) was performed for each vowel with all acoustic variables that emerged as significant in the LMMs as the response and stress as the predictor. Pillai's trace from significant MANOVAs was used to quantify reduction for each vowel; this statistic can range from zero to one, where high values indicate strong separation and low values indicate strong overlap.[14,28] Pillai's trace has become established as a reliable metric in the study of vowel contrasts and mergers. However, it can only be obtained from MANOVA, which requires at least two response variables. As will be observed in Sec. III B, the stressed vs unstressed distributions of two of the vowels were significantly distinguished by one acoustic variable only. A dummy variable of random numbers was generated to be used as the second response in the MANOVAs for those vowels. Last, to establish the relative weight or importance of each significant variable in distinguishing stressed from unstressed realizations, linear discriminant analysis (LDA) was conducted with the significant acoustic variables as predictors and stress as the response. Although the normalization procedure transforms all variables such that they are expressed in the same units (SDs), the resulting ranges will still vary. For example, the ranges of normalized F_2 frequency and mean f_0 were $[-2.91, 4.04]$ and $[-4.52, 11.52]$, respectively. For that reason, all LDA input variables were rescaled to the interval $[0, 1]$, thus rendering the absolute values of coefficients directly interpretable as standardized effect size. Results for relative weight are given as percentages in Sec. III, and the nominal values of the coefficients are reported in Appendix A. In addition, to assess the overall extent and direction of spectral reduction, the areas and centroids of mean $F_1 \times F_2$ vowel spaces were calculated using *R* package *geosphere*.[29]

C. Height neutralization

LMM, MANOVA, and LDA were also used to address the third question, which seeks to establish the extent of contrast reduction in unstressed position. LMMs were computed for each height-contrasting pair in stressed and unstressed positions, with the acoustic variables as responses, vowel as predictor, and random effects as above. MANOVAs were then run for each contrastive pair with all significant variables as response and vowel as predictor. In LDA, the predictors and responses of MANOVA are reversed.

D. Acoustic correlates of vowel contrasts

To find significant differences across all contrastive vowels, LMMs were fitted for each acoustic variable as the fixed effect, vowel as predictor, and the same random effects as above. *Post hoc* Tukey's honestly significant difference (HSD) tests were next used for pairwise comparisons.

All statistical analyses and graphs were made in *R*;[30] $p \leq 0.05$ was interpreted as significant.

III. RESULTS

A. First pretonic vs other unstressed vowels

Results from LMM tests comparing vowels in first pretonic and other unstressed syllables are displayed in Fig. 3 along with group means and SDs. There are no significant spectral or durational differences between the two types of unstressed position for any of the vowels, indicating that there is only one degree of reduction in CSB. The two conditions may differ in mean f_0 ([ɛ a u]) or mean intensity ([u]), where "other unstressed" realizations appear to have higher values. Results for [ɤ] are not shown as the vowel does not occur in the other unstressed condition in the data.

B. Stressed vs unstressed vowels

Means, SDs, and results of LMM comparisons of stressed and unstressed vowels are reported in Fig. 4. All three non-high vowels show significant differences in F_1 frequency, duration, mean f_0, and mean intensity, where higher values are associated with the stressed condition. Stressed vs unstressed [ɛ] and [ɔ] also differ significantly in F_2 frequency, stressed realizations being more peripheral. In addition, [ɛ] has higher F_3 frequency in stressed position. In the high vowels, there are no significant differences in F_1 frequency or duration, which demonstrates that high vowels are not lowered when unstressed. There are also no differences in F_3 frequency, and the only significant F_2 difference is in [u], evidencing fronting in unstressed position. [ɤ] and [u] have significantly higher mean intensity in stressed syllables, whereas the only acoustic variable distinguishing stressed and unstressed [i] is f_0, which is higher in unstressed position.

The stressed vs unstressed realizations of [i] and [ɤ] are significantly differentiated by only one acoustic variable each. To obtain Pillai's traces for these vowels, a dummy variable of random numbers was used as a second response in the MANOVAs, as explained in Sec. II. Two additional LMMs were computed, one for each vowel, with the dummy variable as response. The results are $p = 0.6788$, $r^2 = -0.04$ for [i], and $p = 0.3282$, $r^2 = -0.04$ for [ɤ]. A negative r^2 means that the model fits worse than a horizontal line (the null hypothesis). In other words, the dummy variable has no relation to stress whatsoever and is, therefore, highly unlikely to affect the MANOVA output.

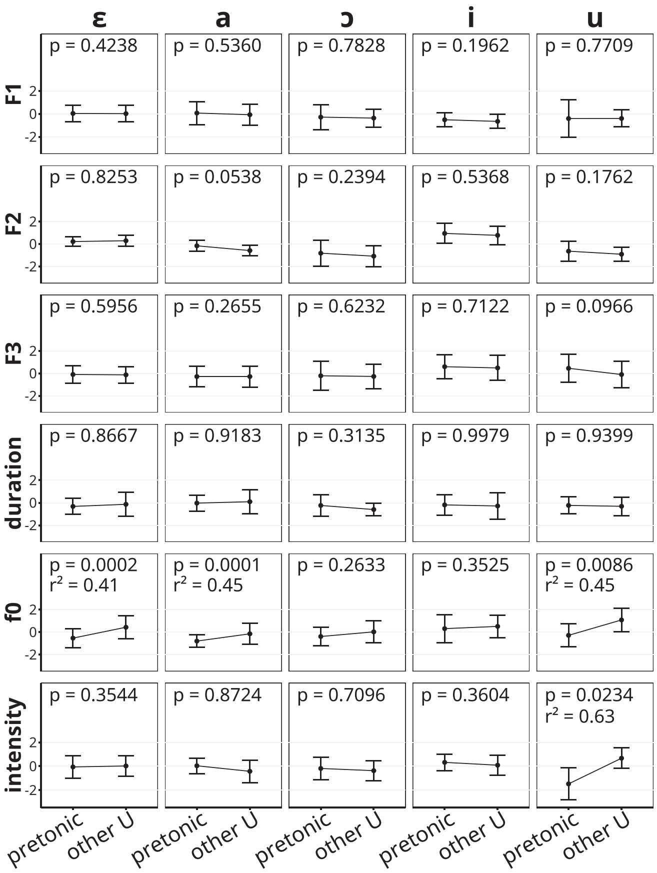

FIG. 3. Results of LMM comparing vowels in first pretonic vs other unstressed (U) syllables. r^2 is shown where $p \leq 0.05$; dots, means; error bars, SDs.

MANOVAs for the effect of stress on all significant acoustic variables taken as a whole yielded significant results for each vowel ($p < 0.0001$). Pillai's traces are plotted in Fig. 5, along with the relative weights of LDA coefficients for each variable that yielded a significant result in the LMM comparisons (Fig. 4). In the non-high vowels, reduction is strongest in [a] and weakest in [ɛ]. For [a], F_1 frequency has the greatest weight, followed by duration, whereas mean f_0 and mean intensity contribute considerably less to distinguishing stressed from unstressed realizations. For [ɔ], F_1 frequency and duration have approximately equal weight; these are followed—in descending order of importance—by F_2 frequency, mean intensity, and mean f_0. For [ɛ], on the other hand, duration outweighs F_1 and F_2 frequencies, which have equal contributions and are followed by mean intensity, mean f_0, and F_3 frequency. There is very little stress-dependent variation in [i], attributable entirely to mean f_0. Pillai's traces are higher for [ɤ] and [u], both differing in mean intensity, whereas F_2 frequency also plays an important role in distinguishing stressed and unstressed [u].

C. Height neutralization

Figure 6 compares non-high and corresponding high vowels with respect to each acoustic variable in both stressed and unstressed positions. In stressed syllables, the vowels in all three pairs are differentiated by F_1 frequency and duration; in addition, stressed [ɛ-i] differ in their F_2 and F_3 frequencies, and [ɔ-u] differ in mean intensity. In unstressed position, all contrast is lost in the non-front pairs [a-ɤ] and [ɔ-u]. CSB unstressed [ɛ-i], on the other hand, remain contrastive in terms of all three formant frequencies and appear to differ additionally in f_0.

Pillai's traces from significant MANOVAs comparing high and non-high vowels in the pairs that emerged as contrastive in LMM (stressed [ɛ-i, a-ɤ, ɔ-u] and unstressed [ɛ-i]) are plotted in Fig. 7, which also shows the relative LDA weight of the acoustic variables that significantly distinguish the vowels in each pair. As we have observed (Fig. 6), all contrast is lost in unstressed position between the vowels in the pairs [a-ɤ] and [ɔ-u]. In [ɛ-i], on the other hand, contrastiveness drops from 0.70 in stressed position to 0.32 in unstressed position (as measured by Pillai's trace): a decrease of 54%.

In the stressed pairs, F_1 frequency has the greatest weight. In stressed [a-ɤ], this is nearly equalled by duration. For stressed [ɛ-i], duration is secondary and followed by the considerably lower weighted F_3 and F_2 frequencies. Mean intensity is secondary for stressed [ɔ-u], followed by duration. F_1 frequency remains primary in distinguishing unstressed [ɛ-i], where it is followed by F_2 frequency, mean f_0, and F_3 frequency.

D. Acoustic correlates of vowel contrasts

As demonstrated in Sec. III C, the height contrast is neutralized in unstressed syllables in two vowel pairs, [a-ɤ] and [ɔ-u], and as a result, the CSB six-vowel stressed inventory, [ɛ a ɔ i ɤ u], is reduced to only four contrastive categories in unstressed position—[ɛ i {a/ɤ} {ɔ/u}]. We now turn to the acoustic differences that obtain among all contrastive vowel categories. LMMs were constructed for each acoustic variable as the response and vowel as the predictor variable. These were followed up by Tukey's HSD pairwise comparisons. Detailed results are given in Appendix B. Figure 8 shows group means and SDs for all contrastive stressed and unstressed vowels and summarizes the pairwise comparison results as compact letter displays: vowels that share at least one letter are statistically indistinguishable by the examined acoustic variable, whereas vowels that share no letters are significantly different.

Vowels are grouped into four distinct categories by F_1 frequency in stressed position: [a], [ɛ ɔ], [ɤ], and [i u]. In unstressed syllables, only two F_1 categories remain: [ɛ {a/ɤ}] vs [i {ɔ/u}]. There are also four categories with regard to F_2 frequency in stressed syllables—[i], [ɛ], [a ɤ], and [ɔ u]—which stay unchanged in unstressed position: [i], [ɛ], [a/ɤ], and [ɔ/u]. F_3 frequency sets apart [i] from the rest of the vowels in both stressed and unstressed syllables, while also distinguishing stressed [ɛ] from stressed [a ɔ]. Two durational categories emerge in stressed position—the longer non-high [ɛ a ɔ] vs the shorter high [i ɤ u]—while the unstressed vowels are undifferentiated by duration. There

https://doi.org/10.1121/10.0025293

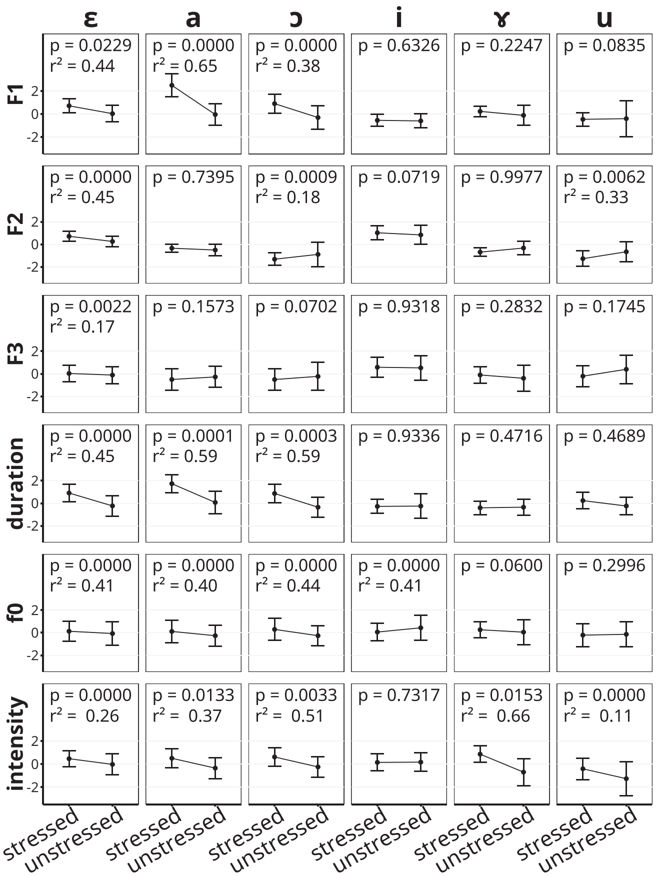

FIG. 4. Results of LMM comparing stressed vs unstressed vowels. r^2 is shown where $p \leq 0.05$; dots, means; error bars, SDs.

are no significant differences in mean f_0 between any two of the stressed vowels. Unstressed [i], on the other hand, has significantly higher mean f_0 than unstressed [a/ɤ]. In terms of mean intensity in stressed position, [ɤ] is higher than [i u], whereas [ɛ ɔ] are higher than [u]; no significant differences in mean intensity were found among the unstressed vowels.

IV. DISCUSSION

A. First pretonic vs other unstressed vowels

The comparison of vowels in first pretonic and other unstressed syllables (Fig. 3) confirmed that CSB first pretonic vowels are not lower—or less reduced—than other unstressed vowels, and that there exists only one degree of vowel reduction in all unstressed syllables. This finding is valid for the material studied, connected read speech, but is also consistent with earlier work based on careful speech.[14,18]

The vowels [ɛ a u] appear to have significantly lower mean f_0 in first pretonic than in other unstressed syllables. We should point out, however, that position within the phrase and focus were not controlled for, and such linguistic variables may have affected the results for f_0. For example, because many of the vowels in the "other unstressed" condition are post-tonic, they may be in the scope of phrase-final continuation rises. The vowel [u] appears to have significantly lower mean intensity in first pretonic syllables. It should be noted that there are very few tokens of [u] in "other unstressed" syllables ($N = 62$) and in many of them, the vowel occurs early in the word, which is a position that is associated with greater intensity than syllables closer to the end (see Ref. 8, p. 161).

B. Stressed vs unstressed vowels

There is a significant and considerable overall difference between the stressed and unstressed realizations of all non-high vowels. Reduction is strongest in [a] and somewhat weaker in [ɔ]. The reduction of [ɛ] is weaker still but, nonetheless, substantial: the separation or non-overlap between stressed and unstressed distributions amounts to 44%. The three vowels differ in terms of how important each significant variable is as a marker of word stress: in [a], F_1 frequency is primary and duration is secondary, in [ɛ], duration is primary and the first two formant frequencies are secondary, whereas for [ɔ], duration and F_1 frequency play a primary role while F_2 frequency is secondary. Mean intensity plays a smaller yet significant part, and the role of mean f_0 is even smaller but still significant. [ɛ] is the only vowel for which F_3 frequency plays a small but significant part.

There are no stress-related differences in F_1 frequency in the high vowels, which confirms that the received view that high vowels are lowered in unstressed position must be rejected. The only spectral difference here is the second formant frequency of [u], which points to fronting in unstressed position. There are no significant durational differences between stressed and unstressed high vowels, which is admittedly somewhat surprising, and may be attributable to

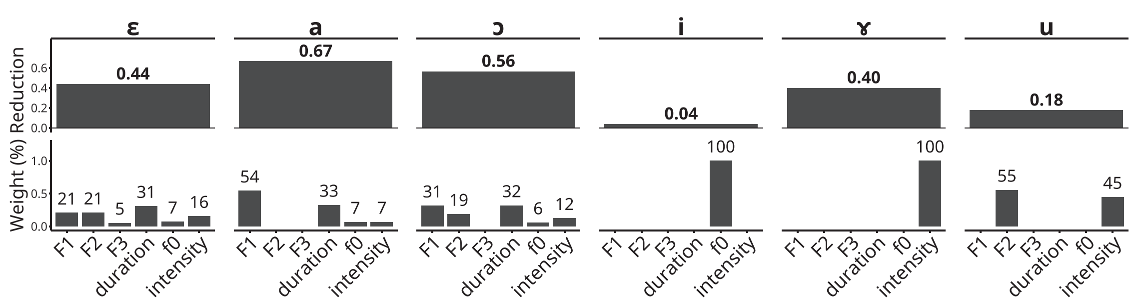

FIG. 5. Pillai's traces for stressed vs unstressed vowels (top) and relative LDA weights (as %) of significant acoustic variables (bottom). Actual LDA coefficients are given in Appendix A.

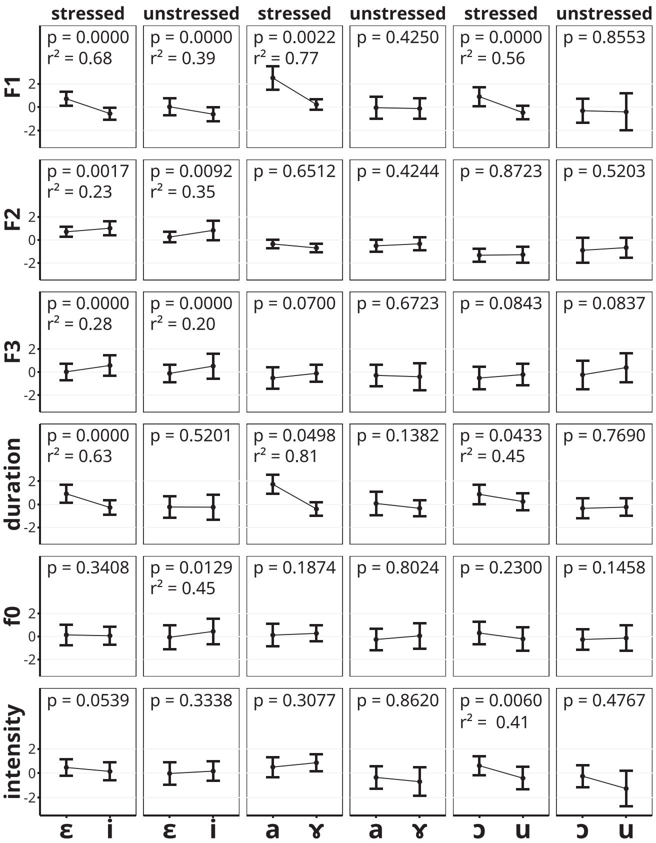

FIG. 6. Results of LMM comparing non-high vs high vowels in stressed and unstressed syllables. r^2 is shown where $p \leq 0.05$; dots, means; error bars, SDs.

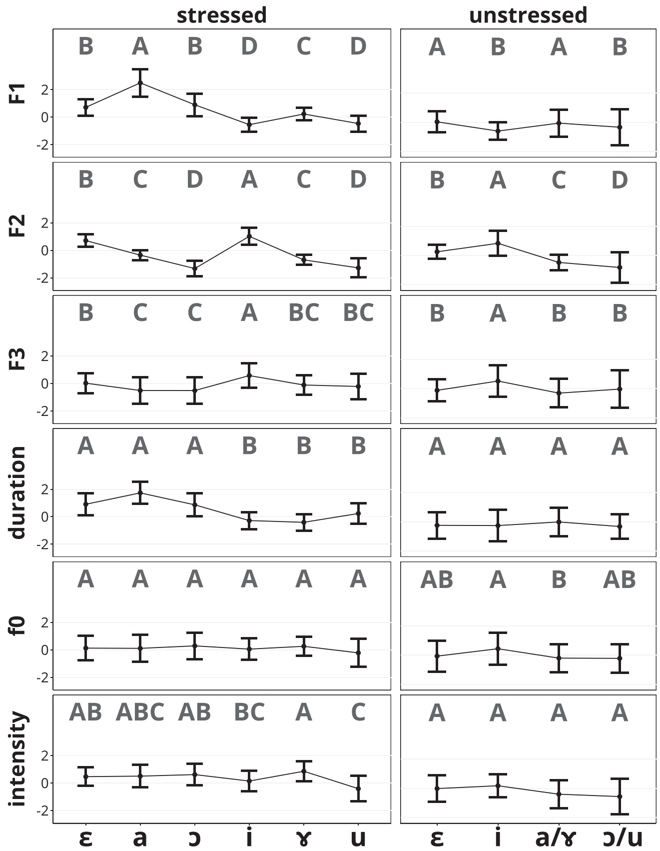

FIG. 8. Results of Tukey's HSD pairwise comparisons of all contrastive vowels. Dots, means; error bars, SDs. (See Appendix B for detailed results.)

a confounding effect of prosodic position, a variable that was not controlled for in this study. Stress is realized by mean f_0 in [i], mean intensity in [ɤ], and F_2 frequency and intensity in [u].

It is hard to speculate whether mean f_0 and mean intensity are sufficient sole markers of stress for [i] and [ɤ], respectively. However, there are other prosodic devices, not examined or controlled for here, that are likely to be employed in signalling word stress, in particular, pitch contours and prosodic peak alignment.

The received view that intensity is the primary marker of Bulgarian word stress (see Ref. 8, p. 160) does not hold true for any of the vowels apart from [ɤ]. It is also important, although not primary, for [u].

Figure 9 shows the mean F_1 and F_2 frequencies of the contrastive vowel categories in stressed versus unstressed position. The ratio of the stressed-to-unstressed area of the vowel space, defined as the area of the convex hulls enclosing the contrastive vowels, is $1:0.12$. Unstressed vowel reduction, therefore, results in a severe contraction of the spectral vowel space. Although the contraction is very dramatic in the vertical dimension, there is also some notable horizontal shrinkage as a result of the significant retraction of [ɛ] and fronting of [ɔ u]. It should be noted that no similar evidence of centralization was found in recent experimental work on Bulgarian vowel reduction. The data in most of those studies, however, are from careful speech,[14,16–18,20] which suggests that centralization is likely to be a function of speech rate and, more generally, speaking style[31,32] rather than a categorical process, in which case, even greater horizontal contraction may be expected in spontaneous speech.

The centroid, or centre of gravity, of the vowel space is considerably raised and also fronted to some extent in unstressed position.

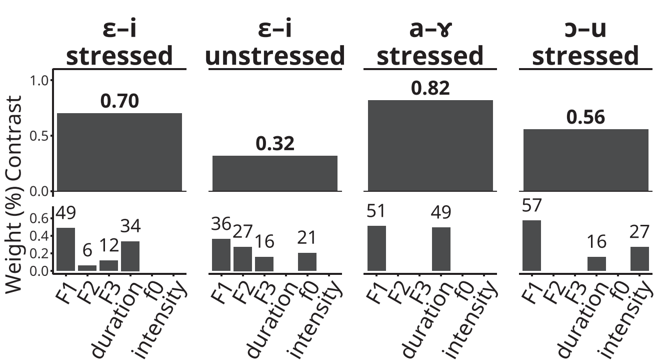

FIG. 7. Pillai's traces for the height contrast, where present (top), and relative LDA weights (as %) of significant acoustic variables (bottom). Actual LDA coefficients are given in Appendix A.

C. Height neutralization

In stressed position, all three vowel pairs [ɛ-i, a-ɤ, ɔ-u] are highly contrastive, although a notably higher degree of

JASA
https://doi.org/10.1121/10.0025293

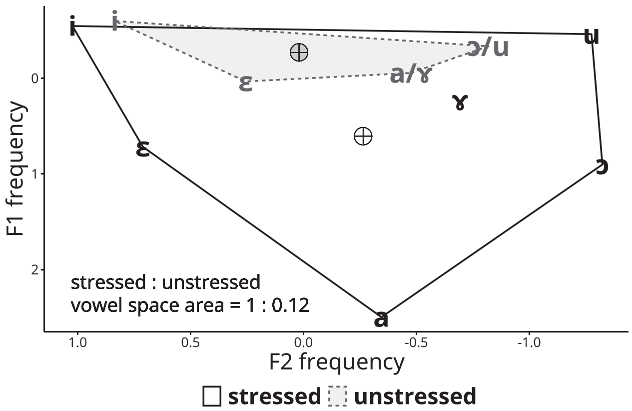

FIG. 9. Contrastive vowels in the $F_1 \times F_2$ frequency space (means). ⊕, centroids.

overlap is tolerated in [ɔ-u] than in [ɛ-i] and [a-ɤ]. F_1 frequency plays a primary role in realizing the height contrast in all three pairs, which is to be expected. Duration is virtually as important in [a-ɤ] and secondary in [ɛ-i]. In [ɔ-u], mean intensity is secondary while duration comes third. In addition, F_2 and F_3 frequencies play a small but significant part in distinguishing stressed [ɛ] from stressed [i]. In none of the pairs do the vowels differ in mean f_0.

In unstressed position, there are no significant differences in any of the acoustic variables for [a-ɤ] and [ɔ-u], which shows beyond doubt that the vowels in both of these pairs are completely merged. The received view that [a-ɤ] merge more readily than [ɔ-u] is, therefore, incorrect with regard to the present state of the language.

Unstressed [ɛ-i], on the other hand, remain contrastive, which is in line with the received view. The unstressed vowels in this pair are significantly distinguished primarily by F_1 and F_2 frequency and additionally by mean f_0 and F_3 frequency. F_2 frequency is noticeably more important for the contrast in unstressed than in stressed position as a result of the centralization of unstressed [ɛ].

A body of work in phonology has previously assumed that Bulgarian [ɛ a ɔ] are all raised in unstressed position and, as a result, merge with [i ɤ u], respectively.[6,7,33–36] While this has been demonstrated to hold good for some eastern dialects,[14,18] in CSB, only the non-front [a ɔ] are raised to the extent of merger with [ɤ u]. Although [ɛ] is affected by reduction and does undergo raising in unstressed syllables, it remains distinct from unstressed [i]. As a result, there are four contrastive vowel categories in unstressed position, [ɛ i {a/ɤ} {ɔ/u}], which are plotted in Fig. 9.

D. Acoustic correlates of vowel contrasts

Four distinct vowel heights in stressed position emerged from comparisons of F_1 frequency, which closely correspond to the four IPA vowel heights: close [i u], close-mid [ɤ], open-mid [ɛ ɔ], and open [a]. As we have observed, the CSB pattern of unstressed vowel reduction results in the raising of [a ɔ] and their merger with [ɤ u]. In addition, in some Eastern Bulgarian dialects, the unstressed front [ɛ] similarly raises and merges with unstressed [i].[14] This parallelism has led various researchers[6–8] to recognize only two contrastive phonological heights in Bulgarian: the high(er) [i ɤ u] and the non-high [ɛ a ɔ], an approach that has been followed here and one that is similar to the standard treatment of Modern Turkish as a language with a two-height vowel system, ignoring the greater variation in phonetic height. In unstressed syllables, two statistically distinct F_1 frequency groups remain: the higher [i {ɔ/u}] and the lower [ɛ {a/ɤ}]. However, if a relational view of vowel height is to be adopted, the unstressed vowels would be regrouped into high(er) [i {a/ɤ}{ɔ/u}] and non-high [ɛ]. Any more fine-grained phonological interpretation of Bulgarian vowel height makes it challenging to formalize vowel reduction. It should also be noted that although [ɤ] is not phonetically a high vowel, it is not raised any further in unstressed position; in other words, it behaves exactly like the true high vowels [i u] and not like the mid [ɛ ɔ], which are raised.

The second formant frequency also divides the stressed vowels into four significantly distinct categories: [i], [ɛ], [a ɤ], and [ɔ u]. It is well-known that frontness increases as openness decreases in front vowels, as reflected in the IPA quadrilateral. Such differences in frontness, however, are normally ignored as phonologically irrelevant, and CSB stressed vowels are typically classed as front [ɛ i], central [a ɤ], and back [ɔ u] or, alternatively, front, back unrounded, and back rounded. Notice that although stressed [ɤ] appears to be somewhat retracted compared to stressed [a] (Fig. 9), the difference in F_2 frequency is not statistically significant. It would, therefore, appear to be more appropriate to transcribe the vowel as [ə] or [ɘ]. We have retained the "ram's horns" symbol introduced in the IPA illustration for Bulgarian[11] to emphasize that the vowel occurs in both stressed and unstressed syllables. Alternative transcriptions with schwa seem to have misled some researchers to assume that the vowel is merely a reduced realization of [a] and not a phoneme in its own right.[35,36] Whereas some degree of horizontal shrinkage of the vowel space is observed in unstressed position, the number of significantly distinct F_2 categories remains unchanged.

F_3 frequency does not appear to play a systematic role in vowel contrasts; it certainly does not correlate with roundness. What can be noted is that the third formant frequency enhances the phonetic distinctness of [i] from the rest of the vowels and that of stressed [ɛ] from the other stressed non-high vowels [a ɔ].

In stressed position, the non-high vowels [ɛ a ɔ] are systematically longer than the high vowels [i ɤ u]. This reflects a cross-linguistic tendency for more open vowels to have longer durations.[37] In unstressed syllables, on the other hand, there are no significant durational differences. This is only to be expected for [a-ɤ] and [ɔ-u], which are completely merged. It is noteworthy, however, that unstressed [ɛ-i] are not distinct in duration either, although a significant difference in F_1 frequency is preserved, which highlights the important role duration plays in CBS vowel reduction.

No significant differences in mean f_0 were found between any of the vowels in stressed position, which shows that although different vowel qualities may be associated with different typical or intrinsic fundamental frequencies,[8,21] such differences are overridden or masked by prosody.

There are only a few significant differences in mean intensity between stressed vowels and none between unstressed vowels. The previously observed fact that more open vowels tend to have higher intensity than closer vowels[8,22] also appears to be largely masked by prosodic factors.

TABLE I. LDA coefficients of the acoustic variables significantly distinguishing *stressed from unstressed* vowels. (Coefficients for [i ɤ] are not calculated: single significant variable for each.)

	F_1	F_2	F_3	Duration	f_0	Intensity
ɛ	−0.69	−1.04	0.18	−0.76	0.17	−0.44
a	−0.85			−0.61	0.15	0.16
ɔ	−0.75	0.50		−0.82	−0.16	−0.34
u		0.88			−0.45	

V. CONCLUSIONS

A current and comprehensive examination of Bulgarian vowel acoustics has, thus far, been lacking, and the present article fills this gap. We have presented a clear and detailed picture of the vocalic phonetics and phonology of CSB after a careful analysis of six acoustic variables—the first three formant frequencies, duration, mean f_0, and mean intensity—of 11 637 vowel tokens from the speech of 140 CSB speakers.

A. First pretonic vs other unstressed vowels

We have confirmed that there are only two levels of word stress—stressed vs unstressed—and there is no difference between unstressed vowels in first pretonic and other unstressed syllables. It has previously been claimed that Bulgarian first pretonic vowels are more open, or less reduced, than other unstressed vowels,[8] echoing a pattern that is standardly associated with Russian.[38,39] No such two-degree reduction system is in place in CSB.

B. Stressed vs unstressed vowels

In unstressed position, the non-high vowels [ɛ a ɔ] are raised, shortened, less loud, and lower in pitch. In addition, unstressed [ɛ] and [ɔ] are centralized. There are no stress-conditioned differences in F_1 frequency or duration in any of the high vowels [i ɤ u] and, thus, the received view that high vowels are lowered when unstressed is once again disproven.

C. Height neutralization

The reduction of CSB [a ɔ] results in complete merger with [ɤ u], respectively, which refutes an earlier claim that [a-ɤ] are more likely to merge than [ɔ-u]. Unstressed [ɛ-i] remain spectrally distinct, which is in line with recent experimental work and the received view but at odds with a common assumption in the phonological literature that all three non-high vowels [ɛ a ɔ] merge with their high counterparts [i ɤ u]. Caution should be exercised when using Bulgarian as an example or case study for vowel reduction with regard to the generalization of the process and the specific variety referred to.

D. Acoustic correlates of vowel contrasts

Acoustic comparisons of all contrastive vowels have revealed four statistically significant levels of height in stressed and two in unstressed position. The observed reduction patterns, however, argue in favour of recognizing only two phonologically significant heights: [ɛ a ɔ] vs [i ɤ u] in stressed and [ɛ] vs [i {a/ɤ} {ɔ/u}] in unstressed syllables. Four statistically significant levels were also identified for backness: [i], [ɛ], [a ɤ], and [ɔ u]. From a phonological perspective, however, [ɛ i] should be collapsed into a single category as the two vowels are maximally front for their respective heights. F_3 did not prove to be a revealing variable in examining Bulgarian vowel quality, with only a handful of significant differences across contrastive vowels. Duration significantly distinguishes high from non-high vowels in stressed syllables, in line with a cross-linguistic tendency for more open vowels to be longer. At the same time, duration is an important ingredient of Bulgarian vowel reduction and, as a result, any durational differences in height pairs are lost in unstressed position. There are only sporadic and unsystematic differences in mean f_0 and mean intensity across vowel phonemes: any intrinsic properties that the individual vowels might have appear to be masked by prosody. We conclude, therefore, that F_1, F_2 frequency, and duration are sufficient acoustic variables for the study of CSB vowels when higher-level prosody is not controlled for, even though mean intensity and mean f_0 do play a significant, if relatively small, part in differentiating stressed from unstressed vowels.

ACKNOWLEDGMENTS

This research was funded by the German Research Foundation (DFG Project No. 491553503) and the Bulgarian National Science Fund (Project No. Кп-06-Н40/11). The authors are grateful to two anonymous reviewers and the associate editor for comments which led to improvements in the text.

TABLE II. LDA coefficients of the acoustic variables significantly distinguishing contrastively *high from non-high* vowels.

	F_1	F_2	F_3	Duration	f_0	Intensity
Stressed						
ɛ-i	1.18	−0.15	−0.28	0.81		
a-ɤ	1.29			1.25		
ɔ-u	1.08			0.30		0.51
Unstressed						
ɛ-i	0.75	−0.57	−0.33		−0.42	

https://doi.org/10.1121/10.0025293

TABLE III. LMM for the effect of vowel on the acoustic variables and *post hoc* Tukey's HSD pairwise comparisons: *stressed* vowels.

	F_1	F_2	F_3	Duration	f_0	Intensity
LMM p	0.0000	0.0000	0.0000	0.0429	0.5429	0.0010
LMM r^2	0.70	0.80	0.24	0.82		0.37
ɛ-a	0.0000	0.0000	0.0142	0.4067		0.9983
i-a	0.0000	0.0000	0.0000	0.0000		0.6208
ɔ-a	0.0000	0.0000	1.0000	0.5851		0.9992
u-a	0.0000	0.0000	0.2654	0.0000		0.0635
ɤ-ɛ	0.0000	0.0000	0.8234	0.0000		0.2399
i-ɛ	0.0000	0.0203	0.0001	0.0000		0.4642
ɔ-ɛ	0.0517	0.0000	0.0015	1.0000		0.9220
u-ɛ	0.0000	0.0000	0.8941	0.0004		0.0115
i-ɤ	0.0000	0.0000	0.0000	0.9998		0.0060
ɔ-ɤ	0.0000	0.0000	0.2052	0.0000		0.9333
u-ɤ	0.0000	0.0000	1.0000	0.5038		0.0001
ɔ-i	0.0000	0.0000	0.0000	0.0000		0.1922
u-i	0.5540	0.0000	0.0000	0.5038		0.4627
u-ɔ	0.0000	0.9639	0.1481	0.0079		0.0055

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX A: LDA COEFFICIENTS

See Tables I and II for nominal values of LDA coefficients.

APPENDIX B: PAIRWISE COMPARISONS

See Tables III and IV for LMM and Tukey's HSD results.

[1]D. Hirst and A. Di Cristo, "A survey of intonation systems," in *Intonation Systems: A Survey of Twenty Languages*, edited by D. Hirst and A. Di Cristo (Cambridge University Press, Cambridge, UK, 1998), pp. 1–44.

[2]D. R. Ladd, *Intonational Phonology* (Cambridge University Press, Cambridge, UK, 1996).

TABLE IV. LMM for the effect of vowel on the acoustic variables and *post hoc* Tukey's HSD pairwise comparisons: *unstressed* vowels.

	F_1	F_2	F_3	Duration	f_0	Intensity
LMM p	0.0000	0.0000	0.0000	0.1785	0.0285	0.0744
LMM r^2	0.22	0.47	0.13		0.43	
ɛ-{a/ɤ}	0.6921	0.0000	0.5198		0.9279	
i-{a/ɤ}	0.0000	0.0000	0.0000		0.0220	
{ɔ/u}-{a/ɤ}	0.0026	0.0042	0.1025		0.9658	
i-ɛ	0.0000	0.0000	0.0000		0.0544	
{ɔ/u}-ɛ	0.0191	0.0000	0.6199		0.9996	
{ɔ/u}-i	0.1533	0.0000	0.0000		0.0536	

[3]W. J. Barry, B. Andreeva, M. Russo, S. Dimitrova, and T. Kostadinova, "Do rhythm measures tell us anything about language type?," in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona (2003).

[4]S. Dimitrova, "Bulgarian speech rhythm: Stress-timed or syllable-timed?," J. Int. Phonetic Assoc. **27**(1-2), 27–33 (1997).

[5]B. Andreeva, W. J. Barry, and J. Koreman, "Local and global cues in the prosodic realization of broad and narrow focus in Bulgarian," Phonetica **73**, 256–278 (2016).

[6]N. S. Trubetzkoy, *Grundzüge der Phonologie* (*Principles of Phonology*) (Akciová moravská knihtiskárna, Prague, 1939).

[7]R. Jakobson, "The phonemic concept of distinctive features," in *Proc. 4th ICPhS*, edited by A. Sovijärvi and P. Aalto, Helsinki (1961) (Mouton and Co., The Hague, 1962), pp. 440–455.

[8]D. Tilkov, T. Bojadžiev, E. Georgieva, J. Penčev, and V. Stankov, *Gramatika na săvremennija bălgarski knižoven ezik* (*Grammar of Contemporary Standard Bulgarian*), *Fonetika* (*Phonetics*) (Bulgarian Academy of Science Press, Sofia, 1982), Vol. 1.

[9]E. A. Scatton, *A Reference Grammar of Modern Bulgarian* (Slavica, Columbus, OH, 1984).

[10]T. Bojadžiev, I. Kucarov, and J. Penčev, *Săvremenen bălgarski ezik. Fonetika, leksikologija, slovoobrazuvane, morfologija, sintaksis* (*Contemporary Bulgarian Language. Phonetics, Lexicology, Word-Formation, Morphology, Syntax*) (Petăr Beron, Sofia, 1998).

[11]E. Ternes and T. Vladimirova-Buhtz, "Bulgarian," J. Int. Phonetic Assoc. **20**(1), 45–47 (1990).

[12]V. Žobov, *Zvukovete v bălgarskija ezik* (*The Sounds of Bulgarian*) (Sema RŠ, Sofia, 2004).

[13]S. Stojkov, *Bălgarska dialektologija* (*Bulgarian Dialectology*) (Marin Drinov Publishing House, Sofia, 1962).

[14]M. Sabev, "Unstressed vowel reduction and contrast neutralisation in Western and Eastern Bulgarian: A current appraisal," J. Phon. **99**, 101242 (2023).

[15]D. Tilkov, *Le vocalisme bulgare. Les mouvements articulatoires et leur effet acoustique dans la formation des voyelles bulgares* (*Bulgarian Vocalism. Articulatory Movements and their Acoustic Effect in the Formation of Bulgarian Vowels*) (Librarie C. Klincksieck, Paris, 1970).

[16]B. Andreeva, W. Barry, and J. Koreman, "The Bulgarian stressed and unstressed vowel system. A corpus study," in *Proc. 14th Interspeech*, Lyon (ISCA, 2013), pp. 345–348.

[17]M. Dokovova, M. Sabev, J. Scobbie, R. Lickley, and S. Cowen, "Bulgarian vowel reduction in unstressed position: An ultrasound and acoustic investigation," in *Proc. 19th ICPhS*, edited by S. Calhoun, P. Escudero, M. Tabain, and P. Warren, Melbourne (2019) (Australasian Speech Science and Technology Association Inc., Canberra, 2019), pp. 2720–2724.

[18]M. Sabev, "Spectral and durational unstressed vowel reduction: An acoustic study of monolingual and bilingual speakers of Bulgarian and Turkish," Ph.D. thesis, University of Oxford, 2020.

[19]M. Sabev, B. Andreeva, C. Gabriel, and J. Grünke, "Bulgarian unstressed vowel reduction: Received views vs corpus findings," in *Proc. 24th Interspeech*, Dublin (ISCA, 2023), pp. 2603–2607.

[20]M. Sabev, "Reduction of unstressed central and back vowels in Contemporary Standard Bulgarian," in *Proc. 18th ICPhS*, Glasgow (2015).

[21]C. H. Shadle, "Intrinsic fundamental frequency of vowels in sentence context," J. Acoust. Soc. Am. **78**(5), 1562–1567 (1985).

[22]I. Lehiste and G. E. Peterson, "Vowel amplitude and phonemic stress in American English," J. Acoust. Soc. Am. **31**(4), 428–435 (2005).

[23]N. Hateva, P. Mitankin, and S. Mihov, "BulPhonC: Bulgarian speech corpus for the development of ASR technology," in *Proc. 10th International Conference on Language Resources and Evaluation*, Portorož (2016), pp. 771–774.

[24]P. Mitankin, S. Mihov, and T. Tinchev, "Large vocabulary continuous speech recognition for Bulgarian," in *Proc. Recent Advances in Natural Language Processing*, Borovets (2009), pp. 246–250.

[25]P. Boersma and D. Weenink, "Praat: Doing phonetics by computer (version 6.3.02) [computer program]," available at http://www.praat.org/ (Last viewed December 1, 2022).

[26]B. M. Lobanov, "Classification of Russian vowels spoken by different speakers," J. Acoust. Soc. Am. **49**, 606–608 (1971).

[27]P. Adank, R. Smits, and R. van Hout, "A comparison of vowel normalization procedures for language variation research," J. Acoust. Soc. Am. **116**(5), 3099–3107 (2004).

https://doi.org/10.1121/10.0025293

[28]J. Nycz and L. Hall-Lew, "Best practices in measuring vowel merger," Proc. Mtgs. Acoust. **20**(1), 060008 (2014).
[29]R. Hijmans, C. Karney, E. Williams, and C. Vennes, "Geosphere: Spherical trigonometry," available at https://CRAN.R-project.org/package=geosphere (Last viewed January 12, 2023).
[30]R Core Team, "*R*: A language and environment for statistical computing," version 4.2.2, available at https://www.R-project.org/ (Last viewed November 14, 2022).
[31]B. Lindblom, "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781 (1963).
[32]B. Lindblom, "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic, Dordrecht, 1990), pp. 403–439.
[33]E. A. Scatton, *Bulgarian Phonology* (Slavica, Cambridge, MA, 1975).
[34]K. Crosswhite, *Vowel Reduction in Optimality Theory* (Routledge, London, 2001).
[35]J. Harris, "Vowel reduction as information loss," in *Headhood, Elements, Specification and Contrastivity*, edited by P. Carr, J. Durand, and C. J. Ewen (Benjamins, Amsterdam, 2005), pp. 119–132.
[36]A. Nevins, "Triumphs and limits of the contrastivity-only hypothesis," Linguist. Var. **15**(1), 41–68 (2015).
[37]L. Lisker, "On 'explaining' vowel duration variation," Glossa **8**, 233–246 (1974).
[38]N. J. Shvedova, *Russkaia grammatika* (*Russian Grammar*) (Nauka, Moscow, 1980), Vol. 1.
[39]A. Timberlake, *A Reference Grammar of Russian* (Cambridge University Press, Cambridge, UK, 2004).