Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal

Montreal, Canada

2 - 7 June 2013

Speech Communication

Session 5aSCb: Production and Perception II: The Speech Segment (Poster Session)

5aSCb22. Phonetic properties of [v] in Russian, Serbian and Greek

Christina Bjorndahl*

***Corresponding author's address: Linguistics, Cornell University, 203 Morrill Hall, Ithaca, NY 14853-4701, cjm295@cornell.edu**

This study examines the phonetic properties of the segment [v] in Greek, Serbian and Russian. [v] patterns phonologically like an obstruent in Greek but like a sonorant in Serbian; in Russian it patterns with both obstruents and sonorants. We test the hypothesis that cross-linguistic differences in the phonological status of [v] correlate with phonetic differences. We report on spectral and durational measures of [v] in four environments: word-initial stressed, word-initial unstressed, word-medial stressed and word-medial unstressed. Our results show an association between phonological patterning and phonetic realization. Greek tokens of [v] are produced with significantly more high-frequency spectral energy than those in Serbian, suggesting a relation between phonological status and phonetic realization in these two languages. Tokens of Russian [v] exhibit the same relationship to tokens of Greek [v] in word-initial stressed position; elsewhere, they are produced with relatively little high-frequency spectral energy. Furthermore, the effects of word position and syllable stress are found to be additive in Russian. These results are important because they support the notion that there exist interactions between the phonological status of a segment and its phonetic realization.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4800950]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Russian [v] patterns phonologically with both sonorants and obstruents. Like obstruents, Russian [v] undergoes final devoicing, resulting in alternations such as [prav-a] ~ [praf], 'right (fem./masc.)', and regressive voicing assimilation to [f], so that /v supe/ is realized as [f supe] 'in the soup'. However, like sonorants, it does not trigger regressive voicing assimilation, thus the contrast [tver^j] 'Tver', [dver^j] 'door' (Padgett, 2002). [v] displays the same asymmetry in Slavic languages, such as Bulgarian (Scatton, 1984) and Czech (Hall, 2003), as well as in non-Slavic languages, such as Hebrew and Hungarian (Barkai and Horvath, 1978; Kiss and Bárkányi, 2006). Recent literature on the asymmetric patterning of [v] in Russian, Hungarian and Hebrew have suggested that such behaviour arises from inherent phonetic properties of [v] in these languages (Padgett, 2002; Kiss and Bárkányi, 2006). Nevertheless, as far as the author is aware, there are no cross-linguistic studies designed to ascertain how the phonological status of [v] interacts with its phonetic realization.

The sound transcribed as [v] patterns in some languages solely with obstruents, as in Greek (Joseph and Philippaki-Warburton, 1987), and in other languages solely with sonorants, as in Serbian (Morén, 2006). The present study seeks to understand whether the phonological status of a segment interacts with its phonetic realization, assessed here using spectral and durational measures of [v] tokens in Russian, Serbian and Greek. Specifically, the objective is to determine whether or not there is acoustic evidence for frication during the realization of [v] tokens.

METHOD

Subjects

Seven native speakers of Greek (4F, 3M; aged 26-32), Serbian (3F, 4M; aged 29-47), and Russian (4F, 3M; aged 22-29 and 73) were recorded. None reported any hearing loss or showed evidence of a speech impairment. With the exception of RuF2, who immigrated to the US at 16, all had left their home country after the age of 18 and all but RuF4 had been in North America for 10 years or fewer at the time of recording.

Speech Materials

Data are based on productions of [v, f] in initial and medial position in CVCV(C) words, recorded in a frame sentence, with five randomized repetitions. For all languages, the vowels were either [a] or [ɔ] (except for Greek [pɛzɔ], 'I weigh' and Russian [trefa], 'clubs'), chosen due to their similarity across the three languages, and because they do not trigger either palatalization or fronting. For Russian and Greek, stress varied between initial and final stress, yielding a total of four environments: Word-Initial Stressed (WIS), Word-Initial Unstressed (WIU), Word-Medial Stressed (WMS) and Word-Medial Unstressed (WMU). Serbian does not allow final stress, and thus the WMS environment was unavailable for Serbian. For Serbian only, trisyllabic words with stress on the second syllable were included in order to obtain the WIU environment; to ensure that the number of syllables did not have an effect on spectral properties, trisyllabic words with initial stress were also included. The frame sentences used are given in (1) – (3); the word list is provided in Table 1.

(1) Greek

eɣrapsa ____ tris fores
I wrote ____ three times

(2) Russian

sveta skazala ____ odin ras
Sveta said ____ one time

(3) Serbian

kaʒe jetsa ____ opet
said, Jetsa ____ again

TABLE 1: Word List

	Greek				Russian				Serbian			
WIS	vaθos	**depth**	faros	**lighthouse**	vozit	**to drive**	faza	**phase**	vakuf	**toponym**	fotʃa	**toponym**
	volos	**marble**	fovos	**fear**	vata	**cotton**	fara	**headlight**	voda	**water**	faca	**face**
									voziti	**to drive**	fabrika	**factory**
WIU	vatos	**passable**	foro	**I wear**	voda	**water**	fagot	**bassoon**	vozatʃi	**drivers**	fakira	**fakir**
WMS	ngavos	**cross-eyed**	sofos	**wise**	davat	**to give**	grafa	**blank**				
					zavod	**factory**						
WMU	tafos	**tomb**	fovos	**fear**	pravo	**right**	pafos	**pathos**	lava	**lion**	kafa	**coffee**
			kava	**cellar**	slovo	**word, noun**	trefa	**clubs (playing cards)**	ovaj	**this (m., sg.)**	rafal	**burst**

Recording Setup and Signal Processing

Recordings took place in a sound attenuated chamber in the Phonetics Laboratory at Cornell University and the University of Toronto. The recording device in both locations was a Sony SD722 digital recorder, and recordings were sampled at 44.1kHz with 16-bit quantization. At Cornell, the microphone used was an Electrovoice RE20 dynamic cardioid microphone; at Toronto, a DPA 4011 cardioid shotgun microphone for all speakers except SeM1, for which a Shure SM10-A headworn microphone was used.

All subjects except for GrF1, SeF1 and SeM3 were naïve as to the purpose of the experiment (GrF1, SeF1 and SeM3 are linguists; GrF1 and SeM3 helped design the word lists for Greek and Serbian, respectively). Subjects were asked to read at a comfortable, conversational pace, to skip words they did not know, and were asked to repeat themselves if some disturbance occurred during a particular token (e.g., a cough, rustling of papers, etc.). All subjects were remunerated $10 for their participation.

The recorded signals were hand segmented in PRAAT, based on visual inspection of the waveform and spectrogram. The signal was then resampled to 22050 Hz for processing and analysis in Matlab. Outliers were determined by calculating the spectral centroid on the high-pass filtered signal (using the methodology detailed below), which was then converted to z-scores relative to the mean over all repetitions and speakers, within each environment (WIS, WIU, WMS, WMU); tokens were excluded if they had a z-score greater than 2.0.

Acoustic Analysis

The hypothesis tested is that tokens of [v] with an obstruent specification, as in Greek, are produced with greater frication than those with a sonorant specification, as in Serbian. Acoustically, frication corresponds to the presence of energy in the high frequency domain, which is quantified here by calculating the spectral centroid on the filtered signal.

The spectral centroid measures the concentration of energy in the frequency domain, and has been used extensively in studies of fricative discrimination (see Jongman *et al.* (2000) for review). Nevertheless, the calculation of the spectral centroid over the entire frequency domain is unsuitable for voiced fricatives. Voicing corresponds to a concentration of energy in the low frequency range, whereas noise corresponds to a concentration in the high frequency range; thus the calculation of the spectral centroid amounts to calculating a weighted mean over a bimodal distribution. To circumvent this problem, and because the goal of this study is to determine the distribution of energy in the high-frequency range, the signal first underwent a 1500 Hz high-pass 4th order butterworth filter, which had the effect of removing the effect of voicing and the first several harmonics. The spectral centroid of the filtered signal was then calculated on a moving 20 ms Hann window with 10 ms overlap over the duration of the segment, with the first window centered at the start of the segment. The values obtained for the middle three windows were then averaged to yield the centroid value for each token.

In order to explicitly compare the energy distributions of [f] and [v] in the three languages, an additional measure was used, whereby all the segments had their centroid measures relativized to [f], and speaker-specific differences were controlled for. More specifically, for each speaker, the mean centroid of [f] was computed; this value was then subtracted from each centroid measure taken for that speaker (for both [f] and [v]). These *relative centroid* measures were then averaged over speakers within a language. In this measure, the higher the value (i.e., the closer to zero), the more similar the distribution of energy is to that of [f] in that environment.

Finally, relative duration of [v] (as a percentage of the syllabel) was measured to assess whether or not there is a correlation between the spectral centroid and duration of the segment. Duration results are reported for only a subset of the data to control for word length and syllable structure; thus, the analysis is restricted only to words that are two syllables, and with [v] in the WIS environment.

RESULTS

Relative Spectral Centroid

Figure 1 displays the results of the relativized spectral centroid for each language, in each condition. A one-way ANOVA on the relativized centroids showed a main effect of language in the WIS and WMU conditions, but not in the WIU condition; a paired t-test indicates that we fail to reject the null hypothesis for the WMS condition; the F-statistics and p-values are given in Table 2. Post-hoc Tukey tests indicate the groupings indicated in Figure 1.

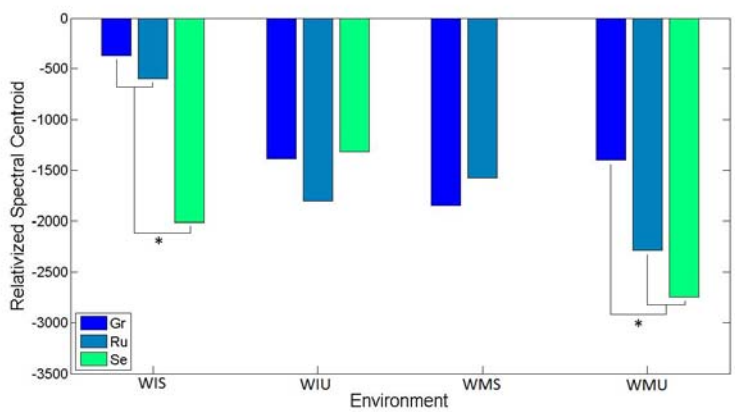

FIGURE 1: Relative spectral centroid of [v].

TABLE 2: Statistical Significance for Relativized Spectral Centroid

Environment	*F*	*p*
WIS	53.36	8.06549 e-20
WIU	0.64	0.5285
WMS	N/A	0.4816
WMU	14.35	1.54785 e-006

In the WIS and WMU environments, the relative centroid of Greek [v] tokens indicate a distribution of energy more similar to Greek [f] than Serbian [v] token are to Serbian [f]. This suggests that Greek [v] is phonetically realized as the voiced counterpart to [f], while Serbian [v] is realized with a more approximant manner of articulation, and supports the hypothesis that there is some interaction between the phonological status of [v] and its phonetic realization.

The difference in patterning across environments suggests a more complicated relationship mediated by both stress and position in the word.

Spectral Centroid

In order to evaluate the effect of stress and word position on the distribution of energy in tokens of [v] in each language, we present results in terms of the non-relativized spectral centroid.

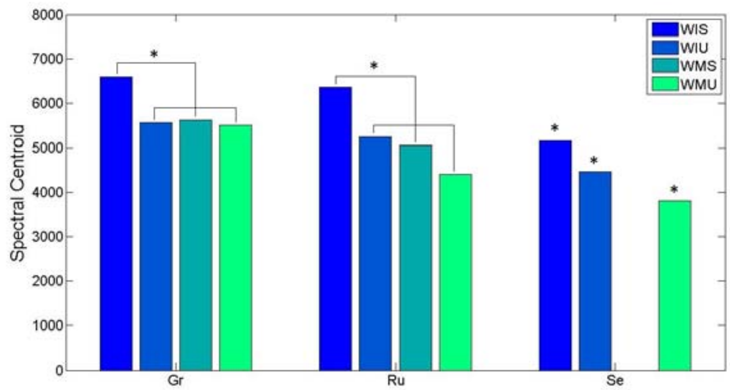

FIGURE 2: Spectral centroid of [v].

A two-way ANOVA (stress × position) was performed on the centroid values for tokens of Greek and Russian [v]. For Greek, there was a main effect of stress [$F = 9.24$, $p = 0.0027$], position [$F = 7.19$, $p = 5.76$] and a slight interaction of stress and position [$F = 5.76$, $p = 0.0173$]; post-hoc Tukey tests indicate that the centroid values in WIU, WMS and WMU position do not differ significantly. For Russian, there was a main effect of stress [$F = 15.27$, $p = 0.0001$] and position [$F = 22.37$, $p = 0$], but no interaction between stress and position [$F = 0.94$, $p = 0.334$]; post-hoc Tukey tests indicate that the centroid values in WIU, WMS and WMU position do not differ significantly. A one-way ANOVA testing environment (WIS, WIU, WMU) on the centroid values for Serbian [v] showed that environment had a significant effect [$F = 21.63$, $p = 3.17346$ e-009]. Post-hoc Tukey tests indicate that centroid values in all three environments differ significantly.

Duration

A two-way ANOVA on the relative duration of [f, v] in the WI condition found a significant effect for segment [$F = 188.11$, $p = 0$], but not for language [$F = 1.94$, $p = 0.1452$]; however, a significant effect was found for the interaction of segment and language [$F = 32.57$, $P = 0$]. Post-hoc Tukey tests showed that the relative duration of [v] in Greek and Russian were not significantly different from each other, but that the duration of Serbian *v* was significantly shorter.

The correlation between the relative duration and centroid was calculated for the labiodental continuants. As can be seen from Figure 3, the correlation between duration and centroid is significant for [v] in all three languages in the WI condition. Nevertheless, the effect is small, as r for Russian is only 0.413.

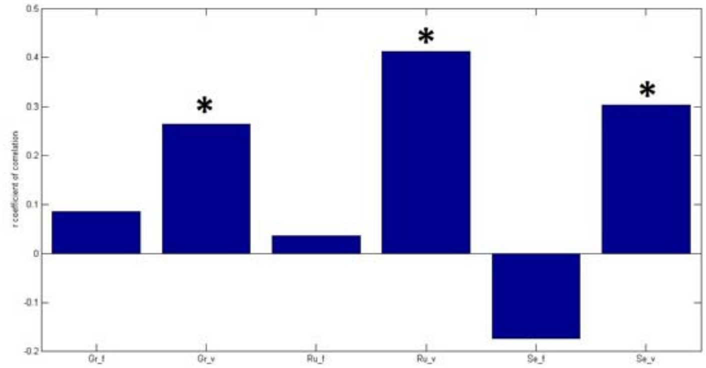

FIGURE 3: Correlation between centroid and relative duration of [v].

DISCUSSION

Whether the segment phonologically transcribed as [v] is realized as a fricative or as an approximant is, in articulatory terms, a question of labiodental aperture and the gradient between oral and atmospheric pressure. Acoustically, this is most likely to manifest in the presence of high frequency energy, which results from the turbulence generated both at the place of stricture between the upper teeth and lower lip, and where the airstream is impeded downstream from the stricture at the upper lip (Stevens, 1988). In order to test this acoustically, this study calculated the spectral centroid on the high-pass filtered signal as a metric for the presence or absence of high frequency energy.

This study presents evidence that the phonological status of a segment may interact with its phonetic realization. In both the WIS and WMU environments, the relativized centroid values for Greek [v] are higher than the relativized centroid values for Serbian [v]. This suggests that, modulo the effect of voicing, the distribution of energy in tokens of Greek [v] is comparable to that of Greek [f], but that the distribution of energy in Serbian [v] tokens is unlike that of Serbian [f]. This supports the phonological classification of Greek [v] as an obstruent and the voiced counterpart to [f], and the classification of Serbian [v] as a sonorant.

In all three languages, [v] was manifested with more frication in the strongest prosodic environment, the WIS condition, evidenced by its higher (relativized) centroid values; in the weakest prosodic environment, the WMU condition, [v] was manifested with less high frequency energy in all three languages. This study does not find evidence of inherent phonetic properties that distinguish [v] in Russian from [v] in Greek or Serbian. Rather, Russian [v] appears to be distinguished by a greater difference in realization across different environments. Moreover, as the interaction between stress and position was not found to be significant in Russian, this suggests that the effect of these factors is additive.

Additionally, this study explored the possibility that differences in phonetic parameters may arise due to gestural undershoot (Lindblom, 1983). A significant difference was found in the duration of [v] in Greek and Serbian, in which a slight correlation was found between duration and spectral centroid. However, since R^2 is less than 0.4 in either case, less than 16% of the data are explained in this way. A more likely explanation is that the differences in phonetic parameters in Greek and Serbian arise from different articulatory representations, and do not exhibit a case of gestural undershoot. Of course, further studies examining the interaction of duration and fricative articulation are required to provide more conclusive results.

To conclude, the results of this study suggest that in Greek, the articulatory representation of [v] includes as a gestural target a fairly narrow labiodental aperture, while in Serbian, it includes a less constricted labiodental aperture.

ACKNOWLEDGMENTS

The author wishes to thank the phonologists and phoneticians at Cornell University for useful feedback and discussion, and to Alexei Kotchetov at the University of Toronto for allowing access to the recording facilities there.

REFERENCES

Barkai, M. and Horvath, J. (**1978**). "Voicing assimilation and the sonority hierarchy: evidence from Russian, Hebrew and Hungarian", Linguistics **212**, 77–88.

Hall, D. C. (**2003**). "A formal approach to /v/: Evidence from Czech and Slovak", in *Formal Approaches to Slavic Linguistics: The Ottawa Meeting*.

Jongman, A., Wayland, R., and Wong, S. (**2000**). "Acoustic characteristics of English fricatives", Journal of the Acoustical Society of America **108**, 1252–1263.

Joseph, B. and Philippaki-Warburton, I. (**1987**). *Modern Greek* (Croom Helm).

Kiss, Z. and Bárkányi, Z. (**2006**). "A phonetically-based approach to the phonology of [v] in Hungarian", Acta Linguistica Hungarica **53**, 175–226.

Lindblom, B. (**1983**). *The Production of Speech*, chapter Economy of speech gestures, 217–246 (Springer-Verlag, New York).

Morén, B. (**2006**). "Consonant-vowel interactions in Serbian: Features, representations and constraint interactions", Lingua **116**, 1198–1244.

Padgett, J. (**2002**). "Russian voicing assimilation, final devoicing, and the problem of [v] (*or*, The mouse that squeaked)", Unpublished paper.

Scatton, E. (**1984**). *Reference Grammar of Modern Bulgarian* (Slavica Publishers, Columbus, Ohio).

Stevens, K. N. (**1988**). *Acoustic Phonetics* (MIT Press).