

AN ANALYSIS OF SPECTRAL ATTRIBUTES, CHARACTERIZING THE INTERACTION OF LITHUANIAN VOICELESS VELAR STOP CONSONANTS WITH THEIR PRE- AND POSTVOCALIC CONTEXT

Arimantas Raškinis, Sigita Dereškevičiute

*Vytauto Didžiojo universitetas
Vileikos St. 8, LT-3035 Kaunas, Lietuva*

Abstract. This work describes an investigation of spectral properties of the voiceless velar stop consonant [k] of Lithuanian. Investigations are based on two Lithuanian corpora: the isolated word corpus VDU-ISO4 and the corpus of short connected phrases VDU-TRI4. Investigations showed that spectra of a plosive [k] exhibit maxima at a frequency corresponding to the second formant of the vowel following plosive [k] and a context independent maximum at 4000 Hz. Properties of these maxima were investigated separately for male and female speech, and for long and short vowel contexts. Results are useful for automatic speech recognition.

1. Introduction

1.1. The aim of the research

Speech technologies is one of the most rapidly developing IT subfields. The research and the analysis of the features describing phonetical units remain an important aspect of the study of this field. Investigation of the stops is notably relevant due to their distinctive feature: the burst release, containing the main acoustical information, is quite short. Accordingly, it is difficult to determine which representation of burst elements – spectra or temporal oscillations better exhibit features, which permit to discriminate in between stop consonants.

The analysis of temporal oscillations of Lithuanian voiceless stop consonants [12], showed that the burst of the stop velar consonant [k] can be divided into two or three shorter bursts, composed of periodical, oscillations decreasing in amplitude. It has been hypothesized that the frequency of the burst oscillation of [k] corresponds to the frequency of the second formant (F_2) of the subsequent vowel (Figure 1). Thus spectra of the consonant [k] should exhibit a prominent peak in the region of the frequency of the second formant (F_2) of the next vowel and spectra of the same consonant [k] should markedly differ for different vowel contexts. Therefore, the main objective of this work is to analyze the spectral features of velar stop consonants, to test the statistical significance of the hypothesis about the impact of the second formant of the following vowel. This work also aims to investigate possible influences of other adjacent sounds and their prosodic features.

Moreover, acoustical studies of interactions of speech sounds are relevant, because smaller-than-word speech units might be constructed on their basis and used for effective speech recognition.

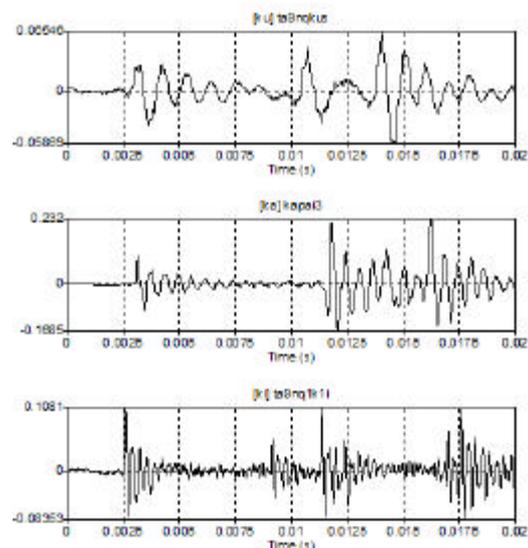


Figure 1. Oscillations of burst elements of stop consonant [k], extracted from Lithuanian words “tānkus”, “kapai”, “tānki”. Vowels [u], [a] and [i] have their second formant (F_2) in the range (500 Hz-900 Hz), (1000 Hz-1700 Hz) and (2000 Hz-3000 Hz) respectively [5], which nicely corresponds to the observed increasing frequency of the burst oscillation

1.2. Spectral analysis of stop consonants

Comprehensive spectral investigation of stop consonants has been made for English language. The main purpose of these investigations usually is to

define which spectral characteristics correspond to different articulatory features of consonants. Kent et al. [9] based their consideration on a spectral template: for bilabials [p] it is characterized as flat or falling, for alveolars [t] – rising and for velars [k], the spectral pattern is determined as compact. Blumstein and Stevens [3] defined the spectra of voiceless stop consonants in a similar way, but also added that the shape of spectrum depends on the following vowel. For instance, a higher spectral energy of a stop consonant is being observed at the frequency corresponding to the second formant of the following vowel. Ali et al. [1] pointed that the frequency of strongest energy in spectrum of velars and alveolars mostly correspond to the frequency of the second formant of the followed vowel. Esposito [6, 7] has explored the amplitudes of the maximum peaks in the spectrum of Italian stop consonants and their dependency on different frequency ranges. She defined a set of acoustic features, in order to discriminate consonants according to their place of articulation. Different sets of acoustic attributes were obtained for [a] and [i] contexts.

Although there are lots of studies on the issue of impact of subsequent vowel for stop consonants, it is quite complicated to compare the data, as various authors use different methods for spectral calculation, approximation and analysis in generally.

1.3. Methodological aspects of spectral analysis

It was necessary to compare spectral characteristics of vowels, consisting of periodical oscillations with noise-like release bursts of consonants in this work. Vowel formants are measured by calculating LPC coefficients according to Burg's algorithm [11] which is based on autocorrelation computations of a periodical signal. The spectrum of consonant's release burst may be analyzed directly as it has been done by Esposito [6, 7] or it can be approximated using various methods. Praat software [4] was used in our experiments. It allows to use LPC, cepstrum or excitation (corresponding to sensibility of human's auditory system; evaluated in Barks) type approximations of consonant spectra.

Having chosen appropriate analysis settings, the cepstral approximation appeared to be the best way to visualize the shape of a spectrum envelope.

Cepstral approximation illustrates more precisely the variances of spectral intensity than LPC approximation (Figure 2 and Figure 3). On the one hand, LPC approximation misses some frequency maxima, which are clearly visible via cepstrum spectral approximation (peaks in the proximity of 5600 Hz and 6000 Hz - 8000 Hz frequency regions, (Figure 2) and in the proximity of 500 Hz and 2000 Hz - 4000 Hz frequency regions, (Figure 3)). On the other hand, being divided in two parts, some ordinates of maxima are set by putting them into the side (peaks in the proximity of 2000 Hz - 3000 Hz frequency region, Figure 2 and in the proximity of 4000 Hz - 5000 Hz frequency

region, Figure 3). However, cepstral approximation gives more maxima in comparison with LPC, thus making it more difficult to relate their positions to the formants.

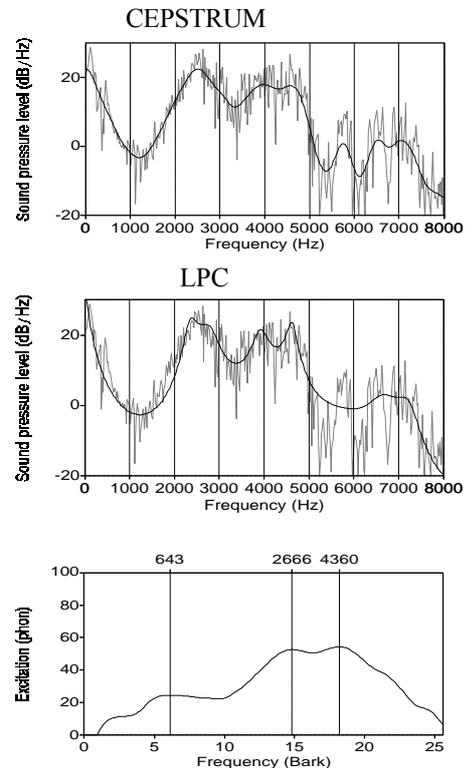


Figure 2. Plots of spectral approximations of the release burst of stop consonant [k], taken from the Lithuanian word “vaikis”: cepstral, LPC and excitation approximations are shown at the top, middle and bottom panels respectively

The approximation of excitation is better at revealing maxima in low frequency region, due to its non-linear scale. However, as maxima appear on a rising slope of the envelope, their frequency position is usually a bit increased (623 Hz, Figure 3). The maxima in a higher frequency region (6000 Hz – 8000 Hz) might be not revealed due to the non-linearity and the falling slope of the envelope. As this work aims to investigate spectral characteristics of release bursts in the proximity of F2 (1000 Hz – 3000 Hz) region, the excitation approximation was chosen as the most appropriate. This approximation was recognized as the most suitable one for revealing the characteristics of stop consonants and was used by other researchers to analyze English consonants having discrimination purposes in mind [1].

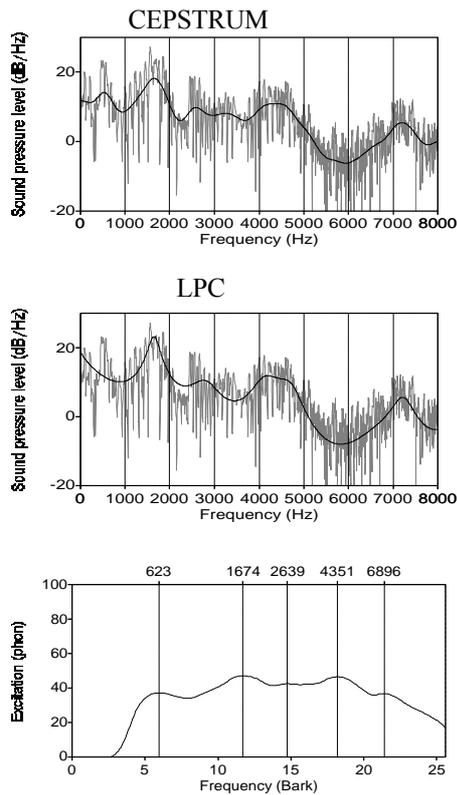


Figure 3. Plots of spectral approximations of the release burst of stop consonant [k], taken from Lithuanian word “kariai”: cepstral, LPC and excitation approximations are shown at the top, middle and bottom panels respectively

2. Data and methods

The data for the analysis of Lithuanian velar stop consonants [k] were obtained from the isolated word speech corpus VMU-ISO4 and the corpus of short connected phrases VDU-TRI4 both compiled at Vytautas Magnus University (VMU) [13]. Praat software [4] was used for data retrieval, speech signal analysis and generation of illustrations. Praat scripting language was used for the extraction of the release burst parts of the consonants and for performing visually-controlled experiment sequences. Automatization allowed carrying out precise large-scale experiments as well as to identify and eliminate annotation mistakes still present in the above-mentioned speech corpora.

3. Experimental results

3.1. The dependency of the release burst spectrum on the subsequent vowel

3.1.1. Initial definition of the F2 range

The spectra of [k] usually have more than one maxima. Thus, in order to test the hypothesis that spectra of [k] show traces of F2 of the subsequent vowel, frequency range has to be defined within which F2 scan needs to be performed. It is well known that F2 is different for male and female speakers as well as

for short and long vowels. This means individual F2 frequency ranges of [k] in the context of 5 Lithuanian vowels: [a], [e], [i], [o], [u] for male and female speakers have to be defined. Vowel duration (long vs. short) was not taken into account during these estimations. Range estimations were based on the VMU-ISO4 speech corpus. First of all, all different CV (consonant vowel) pairs [ka], [ke], [ki], [ko], [ku] for female and male speakers were selected from the corpus and stored separately. Some of the CV pairs, like [ke] were found to be quite few. Secondly, the part of the [k] burst release was extracted, its spectrum and excitation approximation were measured. The cumulative estimation principle was used, i.e. approximation curves were drawn each on the top of the other and visual examination of the dispersion of their maxima let us define the possible range of F2 frequencies (Table 1). Similar results could have been obtained by averaging approximation curves and looking for maxima within an averaged curve.

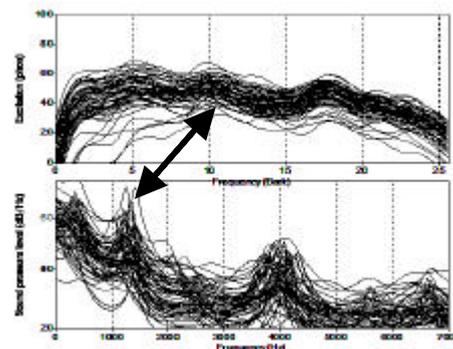


Figure 4. Excitation (top panel) and LPC (bottom panel) approximations of [k], extracted from [ka] samples of male speaker. The arrow shows the maximum corresponding to the F2 of the vowel [a]

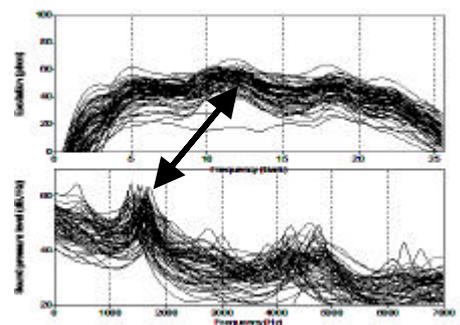


Figure 5. Excitation (top panel) and LPC (bottom panel) approximations of [k], extracted from [ka] samples of female speaker. The arrow shows the maximum corresponding to the F2 of the vowel [a]

Figures 4 through 6 illustrate the different F2 values of male and female speakers in samples of [ka] and [ki]. The maximum peak of F2 of male speaker in comparison with female speaker is moved more to lower frequency range, while the maximum peak of F2 of [ki] sample appears 1500 Hz rightwards in comparison with the [ka] sample.

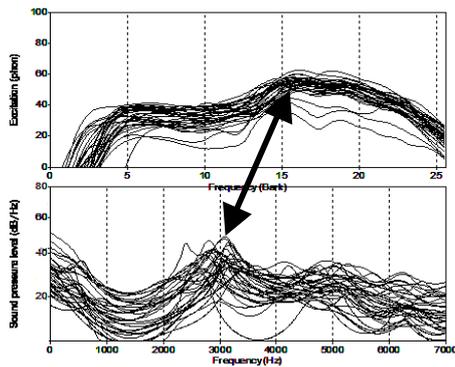


Figure 6. Excitation (top panel) and LPC (bottom panel) approximations of [k], extracted from [ki] samples of female speaker. The arrow shows the maximum corresponding to the F2 of the vowel [i]

Table 1 summarizes the frequency range estimations (for all samples and for both male and female speakers). This range was further used for scanning [k] spectra for maxima which, if found, were considered to be possibly related to the second formant of the subsequent vowel.

Table 1. The frequency ranges to be scanned for peaks within spectra of the release burst of the consonant [k], and possibly corresponding to the F2 of the following vowel

Vowel context	Frequency range in Hz	
	Male	Female
[ki]	2500-3500	2500-3500
[ke]	2500-3100	2700-3300
[ka]	1000-1500	1300-1700
[ko]	500-1100	700-1200
[ku]	400-1000	600-1200

Table 2. Mean and standard deviation estimates for maxima observed within a defined scan range of spectra of the release burst of [k] uttered before vowel [u]

Vowel following the consonant [k]	Speaker's id	Number of utterances		F2 of consonant [k] (Hz)			
				Per speaker		Grouped by gender	
		Per speaker	Grouped by gender	Mean	Standard Deviation	Mean	Standard Deviation
The short [u]	Ak1	18	36	917	91	917	234
	Bj1	18		817	96		
	LK1	18	34	814	66	825	68
	Tk1	16		838	70		
The long [u]	Ak1	4	8	765	56	755	51
	Bj1	4		745	53		
	LK1	4	8	749	89	754	61
	Tk1	4		760	27		
Semidiphthongs [ul,um,ur]	Ak1	10	20	899	82	850	205
	Bj1	10		731	38		
	LK1	8	17	843	71	823	73
	Tk1	9		806	74		

3.1.2. Comprehensive analysis of the dependence of the release burst spectrum on the F2 of the following vowel

The studies of the vowels formants for Lithuanian [2] and other languages [8] show that values of formants differ among male and female speakers, long and short vowels. Therefore, the [k] burst spectra were grouped according to these characteristics. Considering the tentative analysis data, it was decided to treat semidiphthongs as a context separate from others.

Table 2 – Table 6 present summary statistics about that frequency of the spectrum's maximum of the consonant [k], which correspond to the frequency of F2 of the following vowel. Ak1 and Bj1 represent female speakers, while Lk1 and Tk1 represent male speakers. Means and standard deviations are shown separately for each speaker, for males and for females.

Frequency positions of the spectral maxima of the consonant [k] shown in Table 2 closely correspond to the research data about the frequency of the vowel's F2. Formant values for long and short vowels of Šakiai subdialect of Lithuanian are shown in Table 7. Similar F2 values are found for Latvian language [8].

Table 8 gives general values of the second formant of 8 American English vowels [10] for comparison purposes. It seems that English and Lithuanian are quite similar in the sense above.

It should be noted that F2 frequency ranges but not individual F2 values are compared and their relationships examined (Table 2-Table 6) Each individual consonant [k] and the subsequent vowel are in fact different sounds and may have slightly different maxima within F2 range. However sometimes both individual subsequent F2 maxima coincide quite well as it is shown in figure 7. All 5 vowel formants can be seen in the dotted line graphic.

Table 3. Mean and standard deviation estimates for maxima observed within a defined scan range of spectra of the release burst of [k] uttered before vowel [o]

Vowel following the consonant [k]	Speaker's id	Number of utterances		F2 of consonant [k] (Hz)			
		Per speaker	Grouped by gender	Per speaker		Grouped by gender	
				Mean	Standard Deviation	Mean	Standard Deviation
Vowel [o]	Ak1	16	31	860	80	804	88
	Bj1	15		744	46		
	LK1	15	30	942	83	936	76
	Tk1	15		930	72		

Table 4. Mean and standard deviation estimates for maxima observed within a defined scan range of spectra of the release burst of [k] uttered before vowel [a]

Vowel following the consonant [k]	Speaker's id	Number of utterances		F2 of consonant [k] (Hz)			
		Per speaker	Grouped by gender	Per speaker		Grouped by gender	
				Mean	Standard Deviation	Mean	Standard Deviation
The short [a]	Ak1	31	59	1557	147	1505	139
	Bj1	28		1501	127		
	LK1	26	54	1395	152	1307	147
	Tk1	28		1267	114		
The long [a]	Ak1	22	38	1598	75	1600	98
	Bj1	16		1604	100		
	LK1	17	38	1418	34	1376	83
	Tk1	21		1341	33		
Semidiphthongs [al,am,an,ar]	Ak1	16	30	1595	185	1551	150
	Bj1	14		1501	78		
	LK1	17	33	1404	64	1350	78
	Tk1	18		1294	46		

Table 5. Mean and standard deviation estimates for maxima observed within a defined scan range of spectra of the release burst of [k] uttered before vowel [e]

Vowel following the consonant [k]	Speaker's id	Number of utterances		F2 of consonant [k] (Hz)			
		Per speaker	Grouped by gender	Per speaker		Grouped by gender	
				Mean	Standard Deviation	Mean	Standard Deviation
Vowel [e]	Ak1	25	37	3038	265	3046	243
	Bj1	12		3063	200		
	LK1	17	29	2856	190	2822	166
	Tk1	12		2777	120		

Table 6. Mean and standard deviation estimates for maxima observed within a defined scan range of spectra of the release burst of [k] uttered before vowel [i]

Vowel following the consonant [k]	Speaker's id	Number of utterances		F2 of consonant [k] (Hz)			
		Per speaker	Grouped by gender	Per speaker		Grouped by gender	
				Mean	Standard Deviation	Mean	Standard Deviation
The short [i]	Ak1	21	40	2949	302	3105	307
	Bj1	19		3287	198		
	LK1	15	34	2960	257	2859	221
	Tk1	19		2778	151		
The long [i]	Ak1	20	30	2990	251	3034	226
	Bj1	10		3121	140		
	LK1	14	24	2972	191	2981	199
	Tk1	10		2993	218		
Semidiphthongs [il,im,in,ir]	Ak1	19	36	2984	212	3093	294
	Bj1	17		3216	331		
	LK1	15	30	2902	236	2891	199
	Tk1	15		2880	136		

Table 7. F2 frequency of short and long Lithuanian vowels (Hz) [2]

Vowel	Long	Short
[u]	720	930
[o]	940	1140
[a]	1340	1290
[e]	1580	1770
[i]	2280	1900

Table 8. F2 frequency of American English vowels [10]

Vowel	Non-short	Short
[u]	870	1030
[o]	880	-
[a]	-	1100
[e]	1660	-
[i]	2250	1920

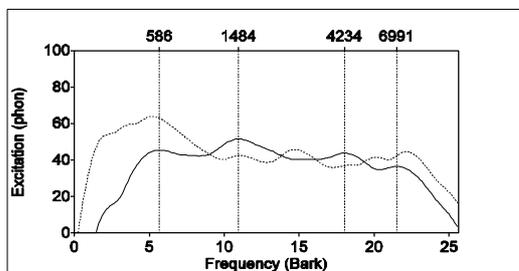


Figure 7. Excitation spectra of consonant [k] (plain line) and of the subsequent vowel [a] (dotted line), extracted from Lithuanian word “vaikas” uttered by a female speaker. The numbers above represent four frequency values (Hz) of maxima of consonant [k]. Vowel’s F2 value is 1510 Hz

F2 maxima of a stop consonant [k] seem to exhibit features proper to the F2 of vowel sounds. Female F2 appears in a higher frequency region than F2 of male speakers. Similarity is observed in F2 relations between short and long vowels (for example, F2 of long [u] is lower than of the short one, while for a vowel [a] a reverse case is observed). Separate analysis of stop consonants [k] followed by semi-diphthongs was probably not very well grounded as all observed differences with respect to other contextual groups are not statistically significant. The results obtained for [o] and [e] contexts should be taken with caution as well. Although the F2 values of consonant [k] closely correspond to F2 of the subsequent vowel [o] and [e], the amount of data is inadequate. Therefore, their averaged statistics are not statistically reliable. Due to this reason, present vowel contexts were kept from being subdivided into the even more specific contexts.

3.2. Investigation of context-independent spectral maximum in the proximity of 4000 Hz of the stop consonant [k]

Preliminary analysis (Figure 4 – Figure 6) showed the presence not only of the peak of [k] spectra related to the F2 of the subsequent vowel, but also another significant peak in the region of 4000 Hz-4500 Hz which seemed to be, independent on the vowel context. In order to verify this tentative hypothesis, the same methodology as in case of the context-dependent maxima was applied. The data (spectra of release burst of consonant [k] with different vocalic environment) related to this maximum were collected and analogous statistics were estimated. The results are shown in Tables 9 through 13.

Table 9. Mean and standard deviation estimates for maximum observed in the region of 4000 Hz of spectra of the release burst of [k] uttered before the vowel [u]

Vowel following the consonant [k]	Speaker’s id	Number of utterances		F2 of consonant [k] (Hz)			
				Per speaker		Grouped by gender	
		Per speaker	Grouped by gender	Mean	Standard Deviation	Mean	Standard Deviation
The short [u]	Ak1	16	34	4164	222	4256	218
	Bj1	18		4338	185		
	LK1	18	35	3865	125	3914	121
	Tk1	17		3965	97		
The long [u]	Ak1	4	8	3805	177	4070	326
	Bj1	4		4336	166		
	LK1	4	8	3965	139	3938	124
	Tk1	4		3911	122		
Semi-diphthongs [ul,um,un,ur]	Ak1	10	20	4049	258	4186	266
	Bj1	10		4322	202		
	LK1	8	16	3930	92	3870	124
	Tk1	8		3809	128		

Table 10. Mean and standard deviation estimates for maximum observed in the region of 4000 Hz of spectra of the release burst of [k] uttered before the vowel [o]

Vowel following the consonant [k]	Speaker's id	Number of utterances		F2 of consonant [k] (Hz)			
		Per speaker	Grouped by gender	Per speaker		Grouped by gender	
				Mean	Standard Deviation	Mean	Standard Deviation
Vowel [o]	Ak1	16	33	4089	104	4184	144
	Bj1	17		4274	116		
	LK1	15	30	3841	150	3882	137
	Tk1	15		3924	112		

Table 11. Mean and standard deviation estimates for maximum observed in the region of 4000 Hz of spectra of the release burst of [k] uttered before the vowel [a]

Vowel following the consonant [k]	Speaker's id	Number of utterances		F2 of consonant [k] (Hz)			
		Per speaker	Grouped by gender	Per speaker		Grouped by gender	
				Mean	Standard Deviation	Mean	Standard Deviation
The short [a]	Ak1	27	53	4440	224	4401	192
	Bj1	26		4361	145		
	LK1	23	52	3982	150	3991	135
	Tk1	29		3998	124		
The long [a]	Ak1	21	38	4189	246	4242	214
	Bj1	17		4335	150		
	LK1	19	44	3970	155	4025	139
	Tk1	25		4067	111		
Semi-diphthongs [il,im,in,ir]	Ak1	15	28	4466	181	4446	141
	Bj1	13		4424	74		
	LK1	12	30	4033	161	3998	127
	Tk1	18		3974	96		

Table 12. Mean and standard deviation estimates for maximum observed in the region of 4000 Hz of spectra of the release burst of [k] uttered before the vowel [e]

Vowel following the consonant [k]	Speaker's id	Number of utterances		F2 of consonant [k] (Hz)			
		Per speaker	Grouped by gender	Per speaker		Grouped by gender	
				Mean	Standard Deviation	Mean	Standard Deviation
Vowel [e]	Ak1	9	18	4681	215	4488	290
	Bj1	9		4296	222		
	LK1	9	28	4248	131	4026	232
	Tk1	19		3920	191		

Table 13. Mean and standard deviation estimates for maximum observed in the region of 4000 Hz of spectra of the release burst of [k] uttered before the [i]

Vowel following the consonant [k]	Speaker's id	Number of utterances		F2 of consonant [k] (Hz)			
		Per speaker	Grouped by gender	Per speaker		Grouped by gender	
				Mean	Standard Deviation	Mean	Standard Deviation
The short [i]	Ak1	11	19	4363	264	4406	251
	Bj1	8		4477	228		
	LK1	9	21	4018	203	4037	158
	Tk1	12		4052	122		
The long [i]	Ak1	8	15	4559	181	4574	161
	Bj1	7		4590	147		
	LK1	13	27	3957	199	3965	191
	Tk1	14		3973	190		
Semi-diphthongs [al,am,an,ar]	Ak1	10	17	4538	143	4478	204
	Bj1	7		4392	257		
	LK1	10	22	4247	88	4061	221
	Tk1	12		3907	173		

Like in the foregone analysis, the amount of data with [e] and [o] contexts was quite small; therefore, these samples were not divided into smaller sub-contexts. The results with [a], [u] and [i] contexts show that maximum frequency values reliably differ only among female (higher) and male (lower) speakers. Differences among groups corresponding to contexts of long and short vowels and diphthongs are not significant.

4. Conclusions

Spectral representation of the consonant [k] is mainly influenced by two the most prominent maxima. The frequency of the first maximum corresponds to the second formant F2 of the subsequent vowel. This maximum is observed in the range from 500Hz to 3500Hz. The frequency of the second maximum is about 4000Hz and seems not to depend on the subsequent vowel.

Female voices have both maxima shifted to the upper frequencies by 200-300 Hz with respect to male voices. More speech data are necessary to be able to investigate possible relationships of spectral maxima in the contexts described by different vowel duration (short vs. long), accent and semidiphthong presence.

Because of the spectral shape of plosive [k] strongly depends on the subsequent vowel, speech recognition research should consider looking for features describing plosive-vowel pairs simultaneously instead of features describing both sounds in isolation.

References

- [1] **A.M. Abdelatty Ali, J. Van der Spiegel, P. Mueller.** Acoustic-phonetic features for the automatic classification of stop consonants. *IEEE Transactions on Speech and Audio Processing*, Vol.9, Issue 8, November 2001, 833-841.
- [2] **R. Baceviciute.** Šakiu šnektos prozodija ir vokalizmas. *Opera linguistica Lituanica*. Vilnius: Lietuviu kalbos instituto leidykla, 2004, (in Lithuanian).
- [3] **S.E. Blumstein, K.N. Stevens.** Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. *J. Acoustical Society of America*, 1979, 66.
- [4] **P. Boersma, D. Weenink.** Praat: doing phonetics by computer. (Version 4.3.14) [Computer program]. Retrieved May 26, 2005. <http://www.praat.org/>.
- [5] **M. Clark** Spectrograph Frequencies. <http://www.vocalist.org.uk/frequencies.html>.
- [6] **A. Esposito.** The amplitude of the peaks in the spectrum: data from [a] context. In Kokkinakis G. (ed), *Proceeding of EUROSPEECH97, University of Patras, Vol.1*, 1997, 1015-1018.
- [7] **A. Esposito, M.Di. Benedetto.** Acoustic analysis and perception of classes of sound (vowels and consonants). *Speech processing, recognition and artificial neural networks*. London: Springer-Verlag, 1999, 54-84.
- [8] **J. Grigorjevs.** Latviešu valodas patskanu sistemas akustiskas raksturojums. *Acta Baltica*. Kaunas: 2001, 15-40.
- [9] **R.D. Kent, Ch. Read** The acoustic analysis of speech. London, Whurr Publishers, 1992.
- [10] **P. Ladefoged.** A Course in Phonetics. Fourth Edition. Thomson Wadsworth, 2000.
- [11] **W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery.** Numerical Recipes in C: the art of scientific computing. Second Edition Cambridge University Press, 1992.
- [12] **A. Raškinis, S. Dereškevičiute** The acoustical analysis of the features of the voiceless stop consonants from the phonograms. *Proceedings of the conference "Information Technologies 2006"*, KTU, Kaunas, 2006, 99-103, (in Lithuanian).
- [13] **A. Raškinis, G. Raškinis, A. Kazlauskienė.** Universal Annotated VDU Lithuanian Speech Corpus. *Proceedings of the conference "Information Technologies 2003"*, KTU, Kaunas, 2003, IX 28-34, (in Lithuanian).

Received November 2006.

