EFFECT OF AGE AND GENDER ON ARTICULATION OF VOICED AND VOICELESS STOP CONSONANTS IN CZECH

Martin Kaňok, Michal Novotný

Department of Circuit Theory, Faculty of Electrical Engineering,
Czech Technical University in Prague, Prague, Czech Republic

Abstract

Evaluation of precision of consonant articulation is commonly used metric in assessment of pathological speech. However, up to date most of the research on consonant characteristics was performed on English while there are obvious language-specific differences. The aim of the current study was therefore to investigate the patterns of consonant articulation in Czech across 6 stop consonants with respect to age and gender. The database used consisted of 30 female and 30 male healthy participants (mean age 51.0 years, standard deviation 18.0 years and range from 20 to 79 years). Four acoustic variables including voice onset time (VOT), VOT ratio and two spectral moments were analyzed. The Czech plosives /p/, /t/ and /k/ were found to be characterized by short voicing lag (average VOT ranged from 14 to 32 ms) while voiced plosives /b/, /d/ and /g/ by long voicing lead (average VOT ranged from -79 to -91 ms). Furthermore, we observed significantly longer duration of both VOT ($p < 0.05$) and VOT ratio ($p < 0.01$) of voiceless plosives in female compared to male gender. Finally, we revealed a significant negative correlation between age and duration of voiceless ($r = -0.36$, $p < 0.05$) as well as voiced VOT ($r = -0.45$, $p = 0.01$) in female but not in male participants.

Keywords

ageing, articulation of consonant, voice onset time, spectral moment, acoustic analysis, Czech

Introduction

The elderly population is growing fast all over the world and therefore, the number of elderly subjects with speech/language disorders has also increased rapidly [1]. The occurrence of voice and speech disorders in adult age is commonly associated with neurological disorders such as Parkinson's disease, Alzheimer's disease, Huntington's disease and many others. Interestingly, it has been previously demonstrated that abnormalities of speech in these disorders may occur several years before the diagnosis is established [2–4] and may even be the earliest indicator of the disease [5]. However, the natural changes of voice and speech also occur in healthy population as a consequence of normal aging process. Therefore, considering that early diagnosis of neurological diseases is essential in improving the patients' life, research focused on speech characteristics of healthy speakers and aging effects is vital for clinical purposes. Without a general model of speech aging for normal healthy, elder individuals, it will be difficult to recognize whether the changes of speech are related to presymptomatic changes due to development of neurodegenerative diseases or if it is just a simple effect of normal aging process.

Evaluation of consonant articulation accuracy is a commonly used metric in the assessment of pathological speech performances [4, 6–8]. Among various classes of consonants, stop plosives are thought to be the most suitable class for investigation. From acoustic point of view, several measurements can be used for their description including various measures of consonant duration, spectral moments or formant transitions [6]. Among them, voice onset time (VOT) was perhaps the most frequently used parameter, and a relatively large amount of data has been published on VOT in normal healthy speakers.

VOT is defined as an interval between the articulatory release of stop and the onset of vocal fold vibration [6]. This temporal parameter is the most reliable acoustic cue for the distinction between voiced and voiceless stops [6, 9]. In most languages, the VOT values for voiced and voiceless stops are produced in discrete duration ranges that correspond to one of the three voicing categories including voicing lead (voicing begins before the stop release; long negative VOT), short voicing lag (short positive VOT) and long voicing lag (long positive VOT) [9]. For example, in English voiceless plosives are predominantly characterized by long voicing lag while voiced plosives by short voicing lag [6, 9]. On the other hand, in Italian, voiceless

plosives are typically characterized by short voicing lag while voiced plosives by long voicing lead [10]. However, up to date most of the research on consonant characteristics was performed on English native speakers [6, 7, 9, 11–13] while there are obvious language-specific differences that have to be considered.

Admittedly, the measurement of VOT is dependent on some physiological or other speaker-related variables such as age, gender, speaking rate or dialect. In speech performances of normal healthy speakers, VOT varies with the speech rate. In particular, VOT decreases as the speaking rate increases and vice versa, indicating that the overall rate changes are implemented at the segmental level [9, 14]. With respect to age and gender differences in VOT production, the findings in previous literature are rather inconsistent [11–13, 15–19]. While some researchers reported no differences in VOT duration between younger and older speakers [11, 15] as well as between performances of men and women [12, 16] others found both longer [17, 18] or shorter [13] duration of VOT in older speakers as well as gender-dependence of VOT [13, 19]. These discrepancies in the literature may be attributed to several reasons such as the various distributions of age in investigated speakers' groups, small sample size, different languages, speaking tasks or consonants used. Taken together, there are conflicting reports on age- and gender-related changes of VOT in literature and little is known about acoustic characteristics of consonants in Czech.

The aim of the current study was, therefore, to investigate the patterns of consonant articulation in Czech across bilabial, alveolar and velar voiceless and voiced stop plosives. In addition, the effect of age and gender on consonant articulation was assessed.

Methods

Participants

From 2016 to 2017, a total of 60 normal healthy participants of different ages were recruited including 30 healthy women (HW, mean age 51.0, standard deviation (SD) 18.0, range 20−79 years) and 30 healthy men (HM, mean age 49.3, SD 16.9, range 22−77 years). The exclusion criteria for participants were the history of neurological or communication disorders, the serious problems with respiration or hearing, the suspicion of memory deficits and/or the active usage of anti-depressants, antipsychotics or other drugs that have a direct effect on speech or mood. All participants were Czech native speakers born or permanently living in Prague or its surroundings. The study was approved by the Faculty of Biomedical Engineering, Czech Technical University in Prague ethics committee Nr. A004/016 and each participant provided written, informed consent for recording procedure.

Recording procedure

The audio data were recorded in a quiet room with a low level of ambient noise using a head-mounted condenser microphone (Beyerdynamic Opus 55, Heilbronn, Germany) placed approximately 5 cm from the subject's lips. The speech signals were sampled at 48 kHz with 16-bit resolution.

During recording, each participant was instructed to read 18 words presented by the examiner on paper cards. The subjects were further warned not to be surprised as some of the words would be meaningless. As Czech is a language with fixed unambiguous pronunciation rules, no training of reading was performed. The cards were presented with stable pace about 1 card per 2 seconds to ensure speaking task to be as speaking rate-independent as possible. The whole speaking task was performed twice.

Speech stimuli included a series of tokens designed as "CVtka", where the C represents one of the consonants and V corresponds to the corner vowels. Specifically, 3 voiceless stop plosives /p/, /t/ and /k/ as well as 3 voiced stop plosives /b/, /d/, and /g/ were involved. The vowels consisted of /a/, /ɪ/, and /u/. These three vowels were chosen as they are a representative sample with respect to vowel height since vowel height has been reported to have an effect on VOT [7]. The suffix /tka/ was added to ensure the easy pronunciation of tokens. All eighteen words included: *KYTKA; PUTKA; DATKA; TYTKA; KUTKA; DYTKA; GATKA; TATKA; GUTKA; KATKA; BATKA; DUTKA; PATKA; PITKA; BUTKA; BITKA; GYTKA; TUTKA*. For further analyses only the initial-consonants were used.

Acoustic analysis

Audio samples were analyzed using specialized speech software PRAAT® [20]. Four acoustic variables including VOT, voice onset time ratio (VOT ratio), and the first (SM1) and second (SM2) spectral moment were employed to assess the precision of consonant articulation. Time of consonant release, vowel onset, and vowel occlusion were determined from the first syllable of each token by hand using both wide-band spectrogram and oscillographic sound pressure signal displayed on the screen. VOT was defined as an interval between the articulatory release of stop and the onset of vocal fold vibration [6]. If multiple bursts occurred, the initial burst was used to measure VOT [7]. VOT ratio was defined as VOT divided by the duration of whole syllable [7, 8]. Both positive and negative values of VOT were allowed. The negative values of VOT refer to voicing lead, typical for Czech voiced plosives [21]. The first spectral moment reflects the average energy distribution across a defined segment of the spectrum, while the second spectral moment shows the deviation of frequencies represented in the spectrum.

Statistical analysis

Before statistical analysis, the assigned values for each participant obtained from two vocal task runs were averaged. In addition, the values for each stop plosive were averaged across all corner vowels. The acoustic parameters were assigned for a subset of voiceless (defined as average value from /p/, /t/, and /k/) and voiced (defined as average value from /b/, /d/, and /g/) consonants. As all acoustic variables were normally distributed (Kolmogorov-Smirnov test), group differences for each acoustic parameter and gender were calculated using the two-sample t-test. The Pearson coefficient was calculated to determine correlations between speech variables and age. Due to the exploratory nature of the study, adjustment for multiple comparisons was not performed and the level of significance was set to $p < 0.05$.

Results

Considering voiceless plosives, significantly longer duration of VOT and VOT ratio was found in men compared to women for stop consonant /p/ ($p < 0.05$; $p < 0.01$) as well as /t/ ($p < 0.01$; $p < 0.001$). In addition, we revealed lower SM1 in men compared to women across all 3 plosives /p/ ($p < 0.05$), /t/ ($p < 0.05$) and /k/ ($p < 0.001$) and lower SM2 in consonant /k/ ($p < 0.01$). With respect to voiced plosives, the significant differences between genders were found only for measure of SM1 where men manifested lower SM1 for /b/ ($p < 0.05$), /d/ ($p < 0.01$) and /g/ ($p < 0.01$) compared to women. The means and standard deviations across each speaker's group and acoustic measure are shown in Table 1 and Table 2.

Table 1: Results of acoustic analyses for plosives /p/, /t/ and /k/. Statistical comparison between genders: *p < 0.05, **p < 0.01, ***p < 0.001.

	/p/ Mean/SD	/t/ Mean/SD	/k/ Mean/SD
VOT (ms)			
HW	13.6/3.6*	18.0/4.7**	27.8/8.2
HM	16.8/5.7*	23.1/6.7**	32.1/9.2
VOT ratio (-)			
HW	0.20/0.06**	0.25/0.06***	0.42/0.16
HM	0.26/0.10**	0.33/0.10***	0.48/0.11
SM1 (Hz)			
HW	2910/400*	3798/498*	3132/299***
HM	2638/456*	3484/663*	2600/420***
SM2 (Hz)			
HW	4662/321	4859/334	4558/289**
HM	4485/399	4748/430	4258/404**

Figure 1 illustrates the results of correlation analysis between speech variables and age across both genders and a subset of voiceless as well as voiced plosives. VOT of voiceless plosives correlated with age in women ($r = -0.36$, $p < 0.05$). Considering voiceless plosives, we also observed a negative correlation between VOT ratio and age in women ($r = -0.41$, $p < 0.05$) and between SM1 and age in men ($r = -0.44$, $p < 0.05$). Regarding voiced plosives, we revealed correlation between age and voiced VOT only in women ($r = -0.45$, $p = 0.01$). In addition, we found a negative correlation between SM1 elicited from voiced consonants and age in both genders (women: $r = -0.40$, $p < 0.05$; men: $r = -0.48$, $p < 0.01$) and between SM2 and age in women ($r = -0.39$, $p < 0.05$).

Table 2: Results of acoustic analyses for plosives /b/, /d/ and /g/. Statistical comparison between genders: *p < 0.05, **p < 0.01, ***p < 0.001.

	/b/ Mean/SD	/d/ Mean/SD	/g/ Mean/SD
VOT (ms)			
HW	-90.6/23.4	-84.3/27.7	-90.2/30.6
HM	-85.7/23.4	-79.1/20.0	-81.9/31.6
VOT ratio (-)			
HW	-0.91/0.24	-0.80/0.22	-0.81/0.25
HM	-0.93/0.27	-0.83/0.18	-0.81/0.29
SM1 (Hz)			
HW	2008/450*	2215/314**	2119/358**
HM	1785/297*	1968/350**	1892/289**
SM2 (Hz)			
HW	4219/474	4046/276	4066/349
HM	4090/284	4060/326	3968/377

Discussion

In previous literature [11–13, 15–19], there are conflicting reports on the effects of age and gender on acoustic characteristics of consonants such as VOT. Furthermore, little is known about acoustic characteristics of stop consonants in Czech. Therefore, the aim of the current study was to investigate the patterns of consonant articulation in Czech across 6 stop plosives and to assess the effect of age and gender on common acoustic features.

The Czech voiceless plosives /p/, /t/ and /k/ were found to be characterized by short voicing lag while voiced plosives /b/, /d/ and /g/ by long voicing lead. The average duration of VOT in voiceless stops ranged from 14 to 32 ms while in voiced stops ranged from -79 to -91 ms. These findings are the most similar to Italian and Spanish [10, 22] (for overview on comparison with other languages see Ogut et al. [23]). Interestingly, the durations of VOT about 14 to 25 ms were also measured from fast syllable repetition of /pa/-/ta/-/ka/ using the database of 24 Czech native speakers [8]. Thus, it appears that VOT does not differ significantly between reading words and diadochokinetic speaking task.
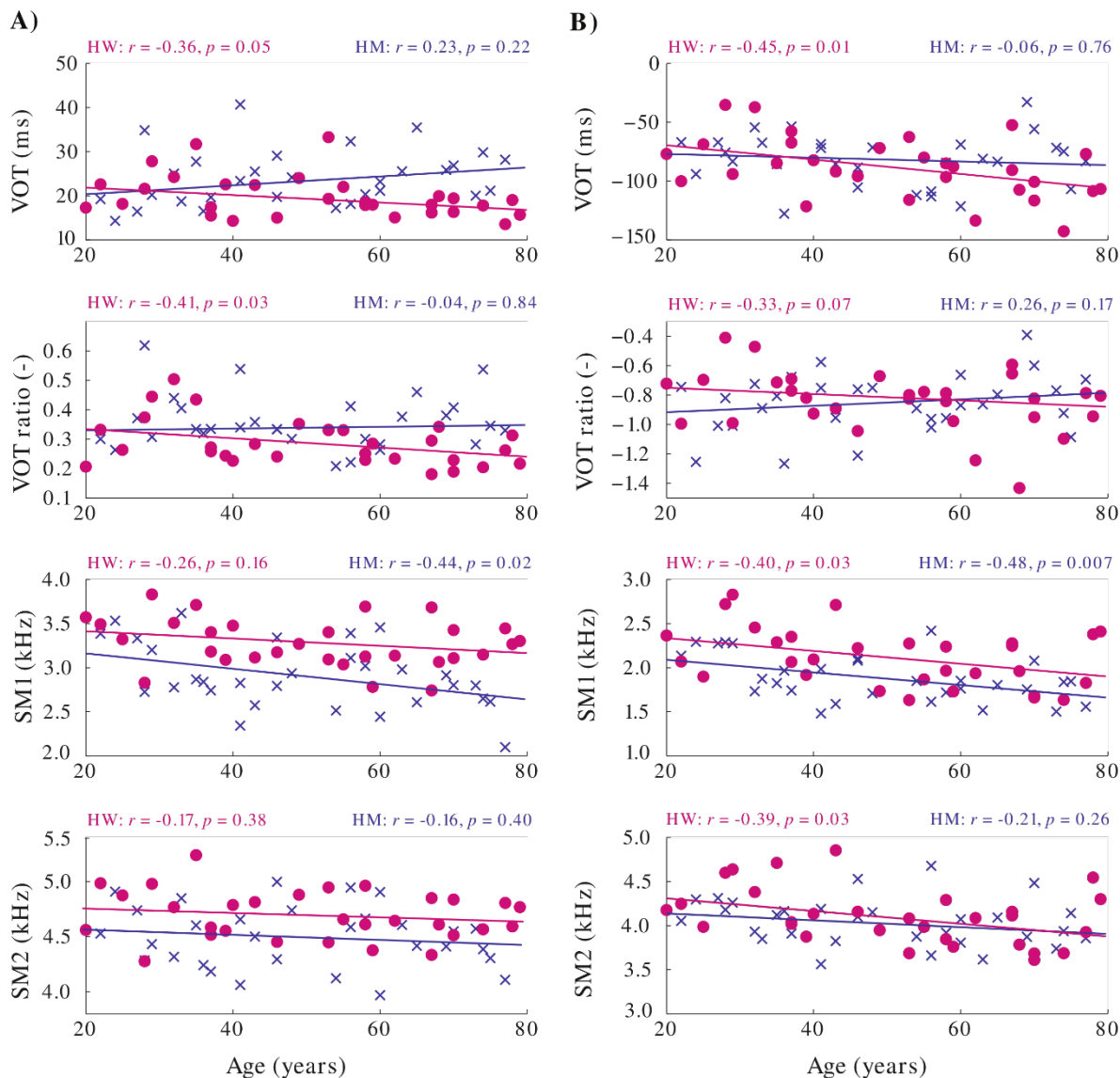
Fig. 1: Results of correlation analysis between speech variables and age for A) voiceless plosives and B) voiced plosives. The purple color and circlets are used for women while the blue color and cross are assigned to men.

With respect to gender differences, we observed the longer duration of VOT in /p/ and /t/ in men compared to women. It should also be mentioned, that these differences between genders were not suppressed using the measurement of VOT ratio, a rate-independent variation of VOT, thus prolonged voiceless VOTs in men cannot be interpreted as a simple effect of slower speaking rate in males. On the other hand, neither VOT nor VOT ratio of voiced plosives was revealed to be gender-dependent. Finally, SM1 was found to be lower in men compared to women across both voiced and voiceless consonants probably mainly due to lower fundamental frequency exhibited by male speakers.

Considering the effects of age, our results showed significant correlation between age and voiceless VOT in female subjects but not in male subjects. The older women exhibited shorter duration of VOT compared to younger ones. Interestingly, the shortening of consonant

duration in women was preserved even when using VOT ratio thus cannot be interpreted as a simple effect of faster speaking rate. Our results can be compared to two previous studies investigating voiceless plosives in Dutch [17] and Hungarian [18] speaking subjects as they reported their speakers to produce short voicing lag, i.e. the same VOT category as was presented by Czech speakers. In general, our findings are quite well in agreement with these studies [17, 18] as a group of Hungarian speakers composed predominantly of women demonstrated significantly shorter VOT of /p/ and /t/ in older (mean age 76.9 years) compared to younger group (mean age 25.3 years) [18] and a group of 130 male and 135 female Dutch-speaking participants demonstrated no significant change in duration of VOT of /p/ and /k/ [17].

Finally, considering voiced plosives, we observed the age-dependence of voiced VOT in women, SM1 in

men as well as in women and SM2 in women. However, to the best of our knowledge, we are not aware of any study that would investigate the effect of aging on articulation of "truly" voiced stop consonants such as those presented in initial-word position in Czech.

Conclusion

We showed that the commonly used measurements for evaluation of precision of consonant articulation such as VOT can be both age- and gender-dependent. Thus, during the design of cross-sectional study or binary classification experiments using assessment of consonant articulation, authors should always keep in mind to carefully select participants in order to ensure gender- and age- balanced groups since an inappropriate design of speakers' groups may easily lead to false results [24].

Acknowledgement

References

[1] Mueller PB. What is normal aging?. Geriatric Medicine Today. 1985; 41: 48–57.

[2] Ramig LA and Scherer R. Acoustic analysis of voice of patients with neurological disease: rationale and preliminary data, Annals of Otology, Rhinology & Laryngology. 1988; 97: 164–172. DOI: 10.1177/000348948809700214

[3] Rusz J, Saft C, Schlegel U, Hoffman R and Skodda S. Phonatory Dysfunction as a Preclinical Symptom of Huntington Disease. Plos One. 2014; 9: e113412. DOI: 10.1371/journal.pone.0113412

[4] Hlavnicka J, Cmejla R, Tykalova T, Sonka K, Ruzicka E and Rusz J. Automated analysis of connected speech reveals early biomarkers of Parkinson's disease in patients with rapid eye movement sleep behaviour disorder. Scientific Reports. 2017; 7: 12. DOI: 10.1038/s41598-017-00047-5

[5] Postuma RB, Lang AE, Gagnon JF, Pelletier A, Montplaisir JY. How does parkinsonism start? Prodromal parkinsonism motor changes in idiopathic REM sleep behaviour disorder. Brain. 2012; 135: 1860–1870. DOI: 10.1093/brain/aws093

[6] Kent RD and Read C. The acoustic analysis of speech. San Diego: Singular Pub. Group 1992.

[7] Fischer E and Goberman AM. Voice onset time in Parkinson disease. Journal of Communication Disorders. 2010; 43(1): 21–34. DOI: 10.1016/j.jcomdis.2009.07.004

[8] Novotny M, Rusz J, Cmejla R and Ruzicka E. Automatic Evaluation of Articulatory Disorders in Parkinson's Disease. IEEE-ACM Transactions on Audio Speech and Language Processing. 2014; 22(9): 1366–1378. DOI: 10.1109/TASLP.2014.2329734

[9] Auzou P, Ozsancak C, Morris RJ, Jan M, Eustache F and Hannequin D. Voice onset time in aphasia, apraxia of speech and dysarthria: a review. Clinical Linguistic & Phonetics. 2000; 14(2): 131–150. DOI: 10.1080/026992000298878

[10] Bortolini U, Zmarich C, Fior R and Bonifacio S. Word-initial voicing in the production of stops in normal and preterm Italian infants. International Journal of Pediatric Otorhinolaryngology. 1995; 31: 191–206. DOI: 10.1016/0165-5876(94)01091-B

[11] Petrosino L, Colcord RD, Kurcz KB and Yonker RJ. Voice onset time of velar stop productions in aged speakers. Journal of Perceptual and Motor Skills. 1993; 76: 83–88. DOI: 10.2466/pms.1993.76.1.83

[12] Morris RJ, McCrea CR and Herring KD. Voice onset time differences between adult males and females: Isolated syllables. Journal of Phonetics. 2008; 36(2): 308–317. DOI: 10.1016/j.wocn.2007.06.003

[13] Torre III P and Barlow JA. Age-related changes in acoustic characteristics of adult speech. Journal of Communication Disorders. 2009; 42: 324–333. DOI: 10.1016/j.jcomdis.2009.03. 001

[14] Baum SR and Ryan L. Rate of speech in aphasia: Voice onset time. Brain and Language. 1993; 44: 431–445. DOI: 10.1006/brln.1993.1026

[15] Sweeting PM, Baker RJ. Voice onset time in a normal aged population. Journal of speech, hearing Research. 1982; 25: 129–134. DOI: 10.1044/jshr.2501.129

[16] Lundeborg I, Larsson M, Wiman S and McAllister AM. Voice onset time in Swedish children and adults. Logopedics Phoniatrics Vocology. 2012; 37(3): 117–122. DOI: 10.3109/14015439.2012.664654

[17] Decoster W and Debruyne F. Changes in spectral measures and voice-onset time with age: a cross-sectional and a longitudinal study. Folia Phoniatricaet Logopaedica. 1997; 49 (6): 269–80. DOI: 10.1159/000266467

[18] Bona J. Voice onset time and speakers' age: data from Hungarian. Clinical Linguistic & Phonetics. 2014; 28(5): 366–72. DOI: 10.3109/02699206.2013.875593

[19] Robb M, Gilbert H and Lerman J. Influence of gender and environmental setting on VOT. Folia Phoniatricaet Logopaedica. 2005; 57: 125–133. DOI: 10.1159/000084133

[20] Boersma P. Praat, a system for doing phonetics by computer. Glot International. 2002; 5: 341–345.

[21] Simackova S, Podlipsky VJ and Chladkova K. Czech spoken in Bohemia and Moravia. Journal of the International Phonetic Association. 2012; 42(2): 225–232. DOI: 10.1017/S0025100312000102

[22] Rosner BS, Lopez-Bascuas LE, Garcia-Albea JE and Fahey RP. Voice-onset times for Castilian Spanish initialstops. Journal of Phonetics. 2000; 28: 217–224. DOI: 10.1006/jpho.2000.0113

[23] Ogut F, Kilic MA, Engin EZ and Midilli R. Voice onset times for Turkish stop consonants. Speech Communication. 2006; 48(9): 1094–1099. DOI: 10.1016/j.specom.2006.02.003

[24] Rusz J, Novotny M, Hlavnicka J, Tykalova T and Ruzicka E. High-accuracy voice-based classification between patients with Parkinson's disease and other neurological diseases may be an easy task with inappropriate experimental design. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2017; 25: 1319–1321. DOI: 10.1109/TNSRE.2016.2621885

Martin Kaňok
Department of Circuit Theory
Faculty of Electrical Engineering
Czech Technical University in Prague
Technická 2, 166 27, Prague 6, Czechia

E-mail: kanokma1@fel.cvut.cz
Phone: +420 224 352 234