NORMALIZATION OF CZECH VOWELS FROM CONTINUOUS READ TEXTS

Jan Volín and David Studenovský

Institute of Phonetics, Charles University in Prague

jan.volin@ff.cuni.cz, studd4af@jinonice.cuni.cz

ABSTRACT

The effectiveness of vowel normalization methods has been suggested to be language-dependent. Six such methods have been used on Czech vowels to see which of them would lead to the best results in follow-up discriminant analyses while preserving the linguistically informative detail. The discriminant analyses had lower success rates for read continuous texts with multiple tokens from 75 speakers than for the carefully-pronounced monosyllables used previously by other authors, suggesting that the results might also be material-dependent. On the other hand, our variable data offered additional insights into sources of contextual variation and allowed us to identify the so-called enhancing contexts in which identity of a vowel is best preserved.

1. INTRODUCTION

For decades, phoneticians have struggled with the enormous acoustic variation of individual speech segments and especially with the task of relating the richness of acoustic manifestations of segments to a relatively small number of functional units of speech. The search for linguistic invariance in functional units has been an omnipresent challenge in research on spoken language for a very long time. Its importance has been emphasized by the rapid development of speech technologies in recent decades. A large number of models have been developed to group physically different acoustic patterns into the right linguistic classes.

Two major approaches are usually used in the struggle with segmental variation. The linguistic approach tries to explain perceptual normalization of speech segments by listeners in speech communication settings. The technological approach of automatic speech recognition (ASR) seeks formulae that will normalize instrumentally-extracted acoustic parameters and reduce overlap with similar functional units. Obviously, some pre-processing methods of ASR might not contribute directly to understanding fundamentals of human speech behaviour. It is, therefore, fortunate that these two approaches are not isolated from each other: some linguists model listeners' perceptual behaviour mathematically with technical applications as one of their aims (e.g., Johnson [1]), while enlightened speech engineers want to know the extent to which their normalization methods reflect human behaviour. One way or another, "… the output of any adequate normalization procedure must be a correct representation of linguistic fact." [2:253].

The greatest contributor to the overall variance between different tokens of "the same" vowels seems to be anatomical differences between speakers, especially those attributable to gender. People with longer vocal tracts produce lower formant frequencies than speakers with shorter vocal tracts. Formant patterns are also affected by the ratio of pharyngeal to oral cavity lengths. It would be useful for many purposes to be able to factor out this source of variation from the data.

Other sources of variation, however, such as context, habit, and dialect, might be of linguistic interest and it is desirable at times to preserve them in normalized data. In other words, it might be useful to neutralize the influence of vocal tract size, while preserving all linguistically-relevant information about the speech segment itself and sometimes even about the system in which it belongs, i.e., the specific language or accent.

Disner [2] showed that each of the four normalization methods she evaluated worked differently for different languages. Since some of her data sets were not particularly large, she rightly suspected even their size of influencing the results [2:256]. Importantly, the actual normalizing mechanisms (e.g., standardizing to the mean, factorial, standardizing to the range) are influenced by the distribution/dispersion of the vowels in the vocalic space of a given language. This fact has to be taken into consideration especially in languages with asymmetric vowel systems.

One of the motivations for the present study was that up till now, the effectiveness of normalizing methods has not been tested on the vowels of Czech. Another motivation was that most of the earlier studies used relatively small samples. The praiseworthy exception is Adank *et al*. [3], with 160 speakers. Yet as [3] focused on dialectal variation, the vowels were pronounced in a strictly controlled setting: two tokens per speaker of each of the nine Dutch vowels in a /sVs/ non-word placed in a carrier sentence.

In contrast, we explored more natural read texts, with eight tokens of each vowel from each of our speakers. We tested the three most successful normalization methods from previous studies [2, 3] in addition to ERB and Bark transformations, and one combination of ERB transformation with normalization. We used discriminant analyses to evaluate normalization success, and we also examined the extent to which the process of the most successful normalization preserves information about contextual influence.

2. METHOD

2.1. Material

Czech has 13 vowels: 10 monophthongs and 3 diphthongs. The monophthongs form a typical five-term system comprising /i, e, a, o, u/. The system is doubled because each articulatory position is occupied by a short and a long vowel which form a phonological opposition. The only asymmetry is the /i:/ - /i/ pair, which has a salient quality as well as a length difference: the shorter /i/ is more open and lax. The front vowels are pronounced with lips spread; the back vowels with lips rounded. Low central /a/ is produced with neutral lip position. We examined only the short vowels, since these are much more frequent in continuous materials, making up 78% of all the vowels occurring in texts [4]. Because of this distributional disparity, our recordings did not provide sufficient number of long vowels.

Seventy-five native speakers of Czech (48 female, 27 male) read two short meaningful texts after familiarizing themselves with their content. They were asked to read in a stylistically unmarked, natural manner. The recordings were carried out in a sound-proof booth with an electret microphone IMG ECM 2000 and sampled at 22050 Hz. Eight examples of each of the vowels from various prosodic, segmental and grammatical contexts were labelled resulting in the set of 3000 tokens (75 speakers, 8 instances of each of the 5 Czech short vowels).

2.2. Method

The frequencies of the first three formants (F1, F2, F3) were extracted automatically by Praat [5] as an arithmetic mean from seven equidistant measurements within the second third of each vowel (Burg method, default settings). The formants were inspected visually in spectrograms, and the automatic measurements were manually corrected where necessary. Corrections principally concerned those cases where a portion of a formant disappeared from the signal or the opposite: where a strong nasal formant was mistakenly extracted as F2. The following paragraphs present formulae of the transformations used. The letter F signals a formant frequency. Its superscript abbreviates the type of transformation; no superscript indicates untransformed Hz. Subscript $_{sf}$ means 'the respective formant of the respective speaker'.

Bark values were computed following the formula by Traunmüller [6]:

$$F_{sf}^{B} = \frac{26.18 \cdot F_{sf}}{1960 + F_{sf}} - 0.53 \quad (1)$$

ERB (equivalent rectangular bandwidth) values were calculated after Moore and Glasberg [7]:

$$F_{sf}^{ERB} = 11.17 \cdot \log\left|\frac{F_{sf} + 0.312}{F_{sf} + 14.675}\right| + 43 \quad (2)$$

Gerstman's method, cited in [2, 3], relates the frequency of each formant's measurements to the speaker's minimum and maximum formant frequencies. The resulting ratio can be optionally multiplied by a constant to produce values more akin to formant frequencies (Gerstman used 999):

$$F_{sf}^{Gerst} = 999 \times \frac{F_{sf} - F_{sf}^{\min}}{F_{sf}^{\max} - F_{sf}^{\min}} \quad (3)$$

Neary, cited in [2], proposed log-transformed values with a speaker-dependent correction term, namely the arithmetic mean of all the F_{sf} values produced by that speaker:

$$F_{sf}^{Near} = \log F_{sf} - \bar{x}_{sf\,(\log)} \quad (4)$$

Lobanov [8] normalized vowels by using the general z-score procedure with the mean and standard deviation in his formula related to all the vowels by the given speaker.

$$F_{sf}^{Lob} = \frac{F_{sf} - \bar{x}_{sf}}{s_{sf}} \quad (5)$$

We used Lobanov's formula with the frequencies in (i) Hertz (Lob-Hz) and (ii) ERBs (Lob-E).

Like [3], we used linear discriminant analysis to assess the normalization procedures. This analysis computes discriminant and classification functions for a set of descriptors (in our case F1, F2, and F3) in order to group the individual cases to the *a priori* known classes. While the discriminant functions serve to describe the role of the descriptors in determining the class membership for the sample cases, classification functions can be used to predict membership of new cases. In practice, it is useful to derive the functions from a training set and verify them on a testing set, although some researchers choose not to validate their results in this manner.

We chose to split our sample of 3000 vowels into the training and the testing set. The discriminant and classification functions were computed for 2000 randomly selected vowels in the training set. Naturally, computations for the raw data in Hz and for each of the normalizing modes were independent. The resulting functions were validated on the testing set, which comprised the remaining 1000 vowels. As can be seen in Table 1, however, our sample proved to be large and representative enough to ensure only negligible differences between the sets, and, actually, in some cases the testing set exhibited even slightly higher success rates. All the differences were insignificant. Such a result clearly indicates that the discriminant and classification functions found are reliable and the results for any new cases from the same population will be stable. For this reason, we carried out most of the remaining analyses on the training and testing sets merged together again.

3. RESULTS

3.1. Overall vowel discrimination

Table 1 shows that the Hertz, Bark and ERB formulae all produced about 72% success rates in discrimination. Gerstman increased the success rate by a further 4% and Neary by 6-7%. Lobanov with Hz values (Lob-Hz) was very slightly better still: about 80% success. Lobanov with ERB values (Lob-E) showed no further gain.

Table 1: Success rates for each transformation method in percentages of correctly recognized vowels.

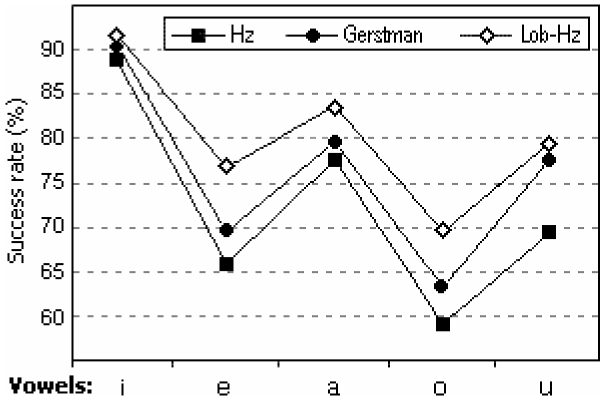
Transformation Method	*Success Rate (%)*	
	Training	*Testing*
Hz	72.5	71.8
Bark	72.7	71.6
ERB	72.3	72.0
Gerstman	76.1	76.5
Neary I	78.8	79.2
Lob-Hz	80.2	80.5
Lob-E	79.3	79.6

3.2. Individual vowels

Although individual vowels made unequal contributions to the overall results, the pattern was about the same for all the transformation methods. As shown in Fig. 1, /i/ was discriminated best, followed by the open central and close back vowels. These three vowels form the vertices of the vocalic space. Both mid vowels, /e, o/, had worse recognition scores. The difference between the best and worst discriminated vowels, /i/ and /o/, was 20-30%. (For clarity, only the Hz, Gerstman, and Lob-Hz rates are shown. Others can be seen in Figure 2).

Figure 1 also indicates that performances of the individual transformation methods were vowel-dependent. For instance, Gerstman was amongst the best for /u/, but amongst the worst for /e/. Therefore, we examined patterns of success rates for all the methods within individual vowels.

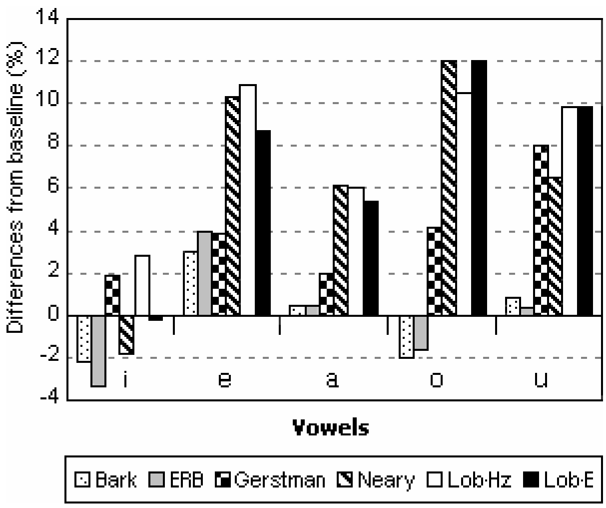
Figure 1: Success rates in discrimination of individual vowels. Since the pattern was similar for all the methods, only the baseline condition in Hz and the Gerstman and Lob-Hz methods are displayed.



Differences from the baseline Hz condition are shown in Figure 2. For /i/ all methods perform around the zero line which represents the baseline Hz condition. For the other vowels, Neary and both versions of Lobanov always outperform the other methods except that Gerstman achieves very good results for /u/.

In recognition tasks, speech engineers very often appreciate improvements by fractions of per cent without reporting the significance of the difference. Just for a rough idea, however, it is useful to realize that, for example, the differences between Gerstman and Neary (the chequered and striped bars in Fig. 2) are not significant for vowel /u/: χ^2 (1) = 0.38; p = 0.45, marginally significant for /a/: χ^2 (1) = 3.49; p = 0.062, significant for /e/: χ^2 (1) = 6.44; p = 0.011, and highly significant for /o/: χ^2 (1) = 8.36; p = 0.004. Other comparisons can be estimated accordingly.

Figure 2: Differences in success rates (% correctly classified vowels) of individual methods, for each vowel. The zero line represents the Hz condition. Negative differences mean worse performance and positive ones better performance.



3.3. Overlap in vowel distributions

While traditional descriptions of the Czech vocalic system maintain that vowel distributions (in the sense of vocalic space distributions) do not overlap [9, 10], the reality of everyday speech points towards quite a different picture [11]. Non-overlapping vocalic territories would obviously lead to 100% success rates in linear discriminant analysis. This was clearly not the case even in our relatively carefully read texts. Therefore, we took a closer look at the existing sources of confusion.

It is easy to hypothesize greater overlap between neighbouring vowels and smaller overlap between more distant ones. However, the way individual vowels contribute to the results is important for the phonetic description of Czech and deserves more explicit elaboration. That can be facilitated by Table 2, which is a confusion matrix of success rates averaged across all the methods used.

Table 2: Classification matrix with mean success rates (%) averaged across all the methods used. Observed vowels (i.e. intended by speakers) are in rows, while predictions (by discriminant analysis) are in columns (n = 3000). Zeros can represent values up to 0.49% (i.e. 0.5% is rounded up).

Predict. / Observ.	i	e	a	o	u
i	**88**	11	0	0	0
e	7	**72**	17	2	1
a	0	11	**81**	8	0
o	0	2	7	**64**	26
u	9	1	0	15	**75**

The most prominent trends were, indeed, mutual confusions of the vowels that are adjacent in the vowel space. Quite unexpected was the erroneous classification of 9% of /u/s as /i/s, which we will return to later with some explanations. This confusion is not mirrored by mistakes in the opposite direction: no observed (intended) /i/ was predicted (i.e. classified) as /u/. There is some more asymmetry in the matrix, even if perhaps less conspicuous. For example, the front mid /e/ is confused with the neighbouring /i/ in about 7% of the cases, but with /a/ to a much greater extent (17%). This outcome actually corresponds to our informal observations of the trend in modern Czech: the front mid /e/ is pronounced as open-mid, most probably under the influence of the Prague accent.

The observed low /a/s were mistakenly classified as front /e/ and back /o/ to about the same extent (11 and 8% respectively). Listening to randomly selected misclassified items confirmed that the direction of the confusion is reflected in the authors' auditory impression.

A remarkable disproportion occurred in the case of back mid /o/, which was mistakenly classified as its higher neighbour /u/ in about 26% of its instances, while only 7% of its instances were incorrectly classified as /a/. Finally, more real, observed /o/s are mistakenly classified as /u/s than vice versa. This trend reflects the perceptual reality as informally ascertained by the authors.

Thus, while the phonological layout of the Czech monophthongal system is neatly symmetrical, its phonetic representation does not seem to be such, at least not from the viewpoint provided by the discriminant analysis and informal examination of randomly chosen items.

For comparison, Lob-Hz, the most successful normalization method, produced a similar classification matrix to that in Table 2, but with all numbers off the diagonal decreased by 1 or 2%. The most dramatic improvement was the reduction of /i/ with /e/ confusions from 11% to 8%, and of /u/ with /i/ from 9% to 6%.

3.4. Contextual influences

Each of the five vowels occurred in eight arbitrary contexts. The natural continuous texts used as the source of the material were relatively short and, importantly, were not designed specifically for this study. Since phonetic contexts are, by their nature, multidimensional (e.g., position in a word, in a stress-group, in an intonation phrase, presence of a pause, manner and place of articulation of the neighbouring consonants, features of the closest vowels, etc.), they cannot be made uniform in all dimensions if we want to preserve the naturalness of the linguistic material. The primary criteria for the original numbering of the contexts were prosodically motivated, and based mainly on positions within a stress-group and on the distances from an intonation phrase boundary. For example, context 1 contains word-final (zero-coda) vowel, last in a three-syllable stress-group, followed by an intonation phrase boundary but not a silent pause. (Stress-groups in Czech are defined as stretches of speech beginning with a stressed syllable and reaching to another, but not including it.) Table 3 shows percentages of correctly discriminated vowels in individual contexts.

Table 3: Discrimination success rates (%) for individual vowel contexts 1-8. Left half: normalization by Hz method. Right half: by Lob-Hz procedure. Grey cells highlight the best result, and black cells the worst, for each vowel.

	Hz					***Lob-Hz***				
	[i]	[e]	[a]	[o]	[u]	[i]	[e]	[a]	[o]	[u]
1	77	57	89	68	75	87	85	93	68	84
2	91	79	77	53	49	93	**93**	91	71	68
3	95	79	68	64	**41**	96	79	77	59	**61**
4	**67**	**33**	88	63	55	**75**	**44**	93	65	72
5	85	73	**95**	**39**	81	85	89	**97**	**37**	85
6	**100**	77	80	67	87	**100**	83	81	83	87
7	97	**80**	**51**	52	**89**	97	88	**56**	87	85
8	97	52	79	**72**	76	99	55	80	**92**	**89**

What seems to be the case, however, is that the resulting informal context classes do not reflect the trends which we arrived at with the discriminant analysis. For example, context 5 (line numbered 5 in Table 3) hosts both best recognized /a/ and worst recognized /o/. This context involves the unstressed vowel in two-syllable words forming two-syllable phrase-final stress-groups. While beneficial to /a/, this context does not support the discriminability of /o/. Almost 90% of the incorrectly classified cases were mistaken for /u/. The variable which was not controlled in this context was the semantic class of the word: /o/ occurs in a pronoun (a synsemantic word), whereas /a/ occurs in a noun (an autosemantic word). Whether this difference might play any role remains to be seen, but to our knowledge, it is not mentioned in Czech phonetic literature.

Context 4 suggests that an important role is played by the immediate segmental context: both the worst classified /i/ and /e/ were preceded by the liquid /l/ and followed by silence. This moved the resulting phones downward in the vocalic space: the intended /i/ was classified as /e/ and intended /e/ as /a/.

Another interesting case is the worst success rate for /u/ of context 3. Context 3 vowels were all in the middle of a 3-syllable stress-group. Only /u/, however, was followed by the voiceless palatal plosive /c/. Anticipatory coarticulation caused massive fronting, which resulted in the sounds [ɨ] and [y] rather than [u]. Naturally, many of these items were confused with /i/.

On the other hand, if we take only our 'best' data (i.e., subsets i6, e2, a5, o8, and u8) and perform discriminant analysis on those, we can observe the success rate rising to 98.1%. These vowels occur in what could be called *enhancing environments* (i.e., /i/ next to palatal /j/, /e/ between two alveolar occlusives, /a/ after /r/ and before silence, /o/ with the closest vowels to it also /o/s plus a neighbouring labiodental consonant, /u/ surrounded by labial consonants flanked by rounded vowels).

To test whether this finding is generalizable for Czech, we would need either a larger database or strictly controlled experimental material: the texts used in this study come from the Prague Phonetic Corpus (founded by Janota and Palková, [12]) and though they are both phonetically rich, they are too short to accommodate all the possible combinations of segments.

However, we did find /i/ in a context parallel to the enhancing environment of the above mentioned i6. It occurred after a palatal consonant and before /s/ just like i6. The difference was that the palatal consonant was not an approximant but an obstruent, which should not really matter too much. The prosodic context was also different. The Lob-Hz normalized data for this new segment (we shall call it *sim*-i6) was obtained following the procedure described above in Section 2 – Method.

In the next step we took the classification functions calculated in the original analysis (see Table 4). These functions are in a general form of

(6) $$C_V = a_V + b_{V1} \cdot \mathrm{F1} + b_{V2} \cdot \mathrm{F2} + b_{V3} \cdot \mathrm{F3},$$

where C_V is the criterion for the given class (i.e., vowel /i/, /e/, /a/, /o/, or /u/), a_V is the constant for the given class, and b_{V1}, b_{V2}, and b_{V3}, are the respective weights for the frequencies F1, F2, and F3. There are as many classification functions as there are classes (i.e., five in our case). Each set of formant frequencies is processed through the classification functions for the individual vowels. The unknown vowel is grouped into the class for which it has reached the highest criterion.

Table 4: The constants (a) and weights (b_1, b_2, b_3) for the classification functions calculated from the Lob-Hz data. Individual vowel classes are in columns.

	/i/	/e/	/a/	/o/	/u/
a	-7,96	-2,81	-4,17	-3,67	-5,64
b_1	-1,65	2,04	4,00	-1,03	-3,35
b_2	7,82	2,41	-0,81	-4,53	-4,89
b_3	0,44	0,28	-0,13	-0,07	-0,52

When the data of our seventy-five *sim*-i6 vowels were processed through the previously calculated classification functions, they were all classified unambiguously as /i/s. The criteria for all the other vowels were considerably lower.

To make sure there were no unexpected artefacts, a new discriminant analysis was carried out, in which the original i6 items were replaced with the *sim*-i6 items. This new analysis, which calculated new discriminant and classification functions for the modified set, confirmed that all the *sim*-i6 items were unambiguous /i/s. The overall results (success rates for individual vowel) also stayed practically unchanged. Thus, the enhancing effect of the preceding palatal and following alveolar consonant with respect to /i/ appears to be robust.

4. DISCUSSION

It has to be stressed at this point that the purpose of this paper was not to describe contextual variation of vocalic formants. Instead, we wanted to identify a suitable vowel-normalization method for Czech and find out whether various coarticulatory phenomena of linguistic interest are preserved after the normalization. In our case the most successful normalization technique was apparently Lob-Hz. Comparison of the values in Table 3 indicates that Lob-Hz does not distort context-specific information about vowels. The results from the baseline Hz and Lob-Hz data are highly correlated ($r = 0.86$, $p < 0.001$, $n = 40$). In other words, the trends in the normalized data are very similar to those in the baseline data: all the worst groups stay the worst after normalization; the best groups change in two cases, but even this change is far from dramatic. Thus, we may conclude that the Lob-Hz procedure increases classification success without hindering linguistic analysis of the vowels. The degree to which it corresponds with human perceptual normalization, however, cannot be ascertained without proper perceptual experiments. Informal inspections of individual items by the authors suggest an uncontroversial relationship: incorrectly classified cases sound ambiguous in the directions indicated by the discriminant analysis.

The material also provided an opportunity to compare the influence of prosodic and segmental contexts on the vowel formants. The segmental impact is undoubtedly quite strong while the prosodic factors seem to be much weaker, perhaps with the exception of a silent pause. This issue, however, would definitely require a dedicated study of its own.

Several other issues raised by our study still deserve to be addressed. For instance, the discrimination success rate for our normalized data was less than that reported in the literature. Adank *et al*. [3] report 92% correct in discrimination of their Lobanov processed data, and Gerstman (cited in [3]) reports 97.5% for his set. Our best results on the Lobanov processed data were just over 80%. We believe that this difference is due to the different materials and speech style. Adank *et al*. used monosyllables in identical segmental and prosodic contexts (a carrier sentence), while Gerstman used isolated monosyllables. Our continuous texts with their multidimensional variation of segmental and suprasegmental influences obviously provide greater overlap of individual vowels' territories.

However, do we have to expect a normalization method leading to a 100% success rate? We have already started perception tests with humans using the same contextless vowels as the input. It seems most likely that human speech perception does not rely on individual speech segment recognition and many individual segments might be identified only after a higher unit (syllable, morpheme, word) has been recognized. If, for some reason, we need perfect recognition of individual segments, then various contextual influences have to be taken into account: plain normalizing formula will not do the trick for natural connected speech.

The second point concerns the importance of the third formant. Wilks' Lambda values from our discriminant analyses indicated that the most important formant for the discrimination of vowels was F2, while F3 had by far the lowest contribution to the discrimination. This finding parallels that of Evans and Iverson [13] who dropped F3 from their analyses after establishing that it had little effect on their results. Nevertheless, little effect still means some effect. A closer look at the F3 contribution in our case suggests that it improves the distinction between /o/ and /u/: for /u/ F3 is on average lower. Again, one must warn against drawing direct parallels between human perception and computational transformations.

Future work will investigate the degree of similarity between Lobanov/Neary procedures on the one hand and human perception on the other. We have started perception experiments to see how human recognition errors correspond with the confusions resulting from discriminant analysis of the normalized data. At the same time, we will continue to investigate coarticulatory and prosodic influences on formant configurations in spoken Czech.

5. ACKNOWLEDGMENTS

The authors would like to express gratitude to Sarah Hawkins for her valuable comments over the draft of this paper. All the shortcomings are naturally ours. Supported by the grants GACR 405/05/0436 of the Czech Academy of Sciences and VZMSM0021620825 of the Czech M.Ed.

6. REFERENCES

[1] Johnson, K. 1990. Contrast and normalization in vowel perception. *Journal of Phonetics* 18, 229-254.

[2] Disner, S.F. 1980. Evaluation of vowel normalization procedures. *J. Acoust. Soc. Am.* 67/1, 253-260.

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

[4] Ludvíková, M., Kraus, J. 1966. Kvantitativní vlastnosti soustavy českých fonémů. *Slovo a slovesnost* 27, 334-344.

[5] Boersma, P., Weenink, D. 2004. *Praat 4.1.28: Doing phonetics by computer*. (www.praat.org).

[6] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am.* 88, 97–100.

[7] Moore, B.C.J., Glasberg, B.R. 1983. Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *J. Acoust. Soc. Am.* 74/3, 750–753.

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

[9] Hála, B. 1975. *Fonetika v teorii a v praxi*. Praha: SPN.

[10] Palková, Z. 1997. *Fonetika a fonologie češtiny*. Praha: Karolinum.

[11] Hedbávná, B. 2003. Czech vocalic system in realizations of individual speakers. In: *Proc. 15th ICPhS*, 679-682. Barcelona: UAB.

[12] Janota, P., Palková. Z. 1991. Contemporary Czech pronunciation: a database study. In: *Proc. 12th ICPhS*, 178-181. Aix-en-Provence: UP.

[13] Evans, B.G., Iverson, P. 2003. Vowel normalization for accent: A comparison of northern and southern British English speakers. In: *Proc. 15th ICPhS,* 829-832. Barcelona: UAB.