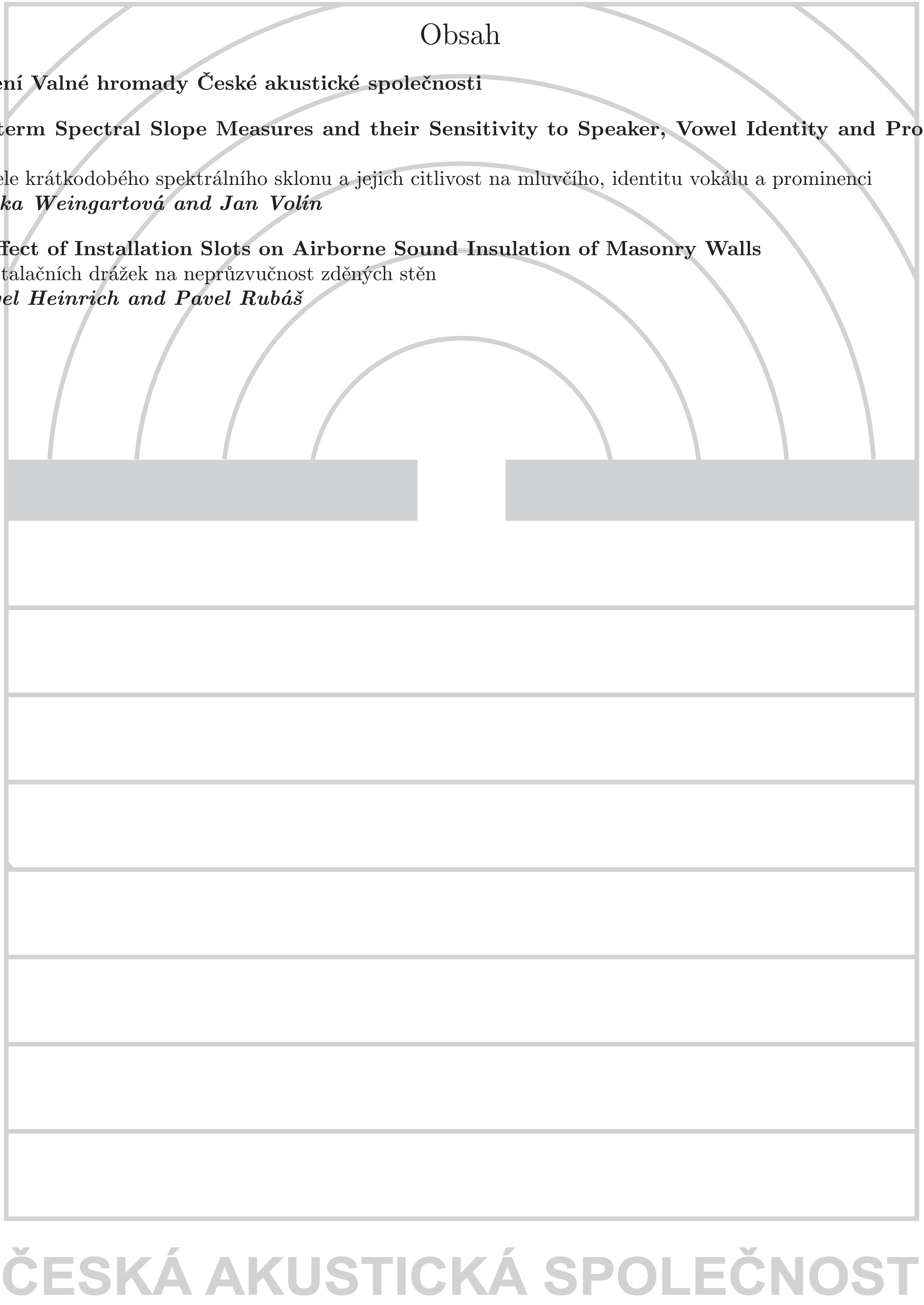
AKUSTICKÉ LISTY

České akustické společnosti
www.czakustika.cz

ročník 20, číslo 1 březen 2014

Obsah



Usnesení Valné hromady České akustické společnosti, konané dne 23. ledna 2014 v prostorách Fakulty elektrotechnické ČVUT

Valná hromada ČsAS bere na vědomí:

1. zprávu o činnosti Rady ČsAS;
2. zprávy o činnosti jednotlivých odborných skupin a o jejich dalším zaměření;
3. zprávu o přípravě 88. akustického semináře;
4. zprávu o výsledcích revize hospodaření společnosti;
5. výsledky voleb do Rady společnosti a výsledky voleb předsedů odborných skupin;
6. zprávu o usnášeníschopnosti Valné hromady, 46/76;
7. informace o plnění úkolů a poslání Akustických listů a vyzývá členy k zasílání příspěvků.

Pro funkční období roku 2014 byli v jednotlivých odborných skupinách zvoleni:

A. Obecná, lineární a nelineární akustika
předseda – M. Červenka zástupce – M. Bednařík

B. Ultrazvuk a akustické emise
předseda – J. Plocek zástupce – R. Bálek

C. Hluk a vibrace
předseda – K. Šnajdr zástupce – J. Kozák

D. Prostorová, stavební a urbanistická akustika
předseda – P. Novák zástupce – M. Meller

E. Zpracování a záznam akustických signálů
předseda – T. Salava

F. Psychoakustika, fyziologická akustika a akustika hudby a řeči
předseda – J. G. Švec

G. Elektroakustika
předseda – Z. Kešner zástupce – B. Sýkora

Valná hromada ČsAS schvaluje:

1. zprávu o činnosti Rady za kalendářní rok 2013 a uděluje Radě absolutorium;
2. zprávu o hospodaření společnosti za kalendářní rok 2013;
3. výši členských příspěvků na rok 2014 (500 Kč pro členy, 150 Kč pro studenty a důchodce);
4. činnost Rady a odborných skupin v roce 2013.

Valná hromada ukládá nově zvolené Radě společnosti na kalendářní rok 2014:

1. pokračovat v odborné a organizační činnosti i v zahraničních kontaktech a v rozvíjení spolupráce;
2. věnovat pozornost pořádání odborných akcí a pravidelných seminářů odborných skupin;
3. nadále rozvíjet vydávání Akustických listů.

Valná hromada ukládá nově zvoleným předsedům odborných skupin na kalendářní rok 2014:

1. publikovat informace o připravovaných aktivitách skupin v Akustických listech a na webové stránce s předstihem tak, aby se zájemci mohli včas na akce přihlašovat.

Valná hromada doporučuje Radě ČsAS:

1. pravidelně se zabývat činností a plánem akcí odborných skupin;
2. pravidelně se zabývat plánem a zaměřením konaných akustických konferencí;
3. poskytovat možnost finančních výhod členům společnosti, např. nižšími sazbami vložného na akcích pořádaných společností;
4. zvážit možnost uložení peněz společnosti na výhodnější termínovaný účet.

Výsledky voleb do Rady České akustické společnosti:

Počet vydaných platných hlasovacích lístků: 46
Odevzdáno platných hlasovacích lístků: 46

předseda:	O. Jiříček	46
místopředseda:	V. Kunzl	46
sekretář:	M. Brothánek	46
hospodář:	O. Kudějová	46
revizní komise:	J. Kozák	46
	T. Hellmuth	46
	D. Potužníková	46

Návrh usnesení sestavila návrhová komise ve složení P. Urban a M. Meller.

Akustické listy, **20**(1), březen 2014, str. 5–12 © ČsAS

Short-term Spectral Slope Measures and their Sensitivity to Speaker, Vowel Identity and Prominence

Ukazatele krátkodobého spektrálního sklonu a jejich citlivost na mluvčího, identitu vokálu a prominenci

Lenka Weingartová and Jan Volín

Institute of Phonetics, Charles University in Prague, Náměstí Jana Palacha 2, 116 38 Praha 1

This study examines ten methods of measuring short-term spectral slope with the primary aim of quantifying vowel prominence. A highly controlled dataset consisting of 45 pseudowords with varied placement of prominent vowel read by 12 speakers was used. The investigated methods included Hammarberg index, α-measure, linear regression, differences of spectral energies in separate frequency bands, difference between amplitudes of the first two formants and spectral emphasis. All of the methods showed substantial sensitivity to speaker and vowel identity in particular. Nevertheless, most of the methods were also able to discriminate among prominent and non-prominent realizations of the vowels. The difference between sums of energy in two adjacent frequency bands separated by a floating pivot locked to the value of the second formant proved to be the most successful measure. On the contrary, linear regression of the spectrum performed worst. Also, exclusion of the fundamental frequency band decreased performance of the methods significantly. The vowel /i/ was shown to be the most sensitive to prominence, whereas /u/ appeared to be the least useful for quantifying both prominence and speaker identity. A considerable speaker discrimination potential of the methods was also corroborated.

1. Introduction

Spectral slope (also spectral tilt or spectral balance) is a speech phenomenon resulting from the combination of the source signal originating in the glottis and filtering characteristics of the vocal tract, which create different energy composition of the spectrum. For instance, while fricatives exhibit a flat spectrum with energy more or less evenly spread, vowels will show a characteristic decay of energy towards higher frequencies. This manifests as a gradual decline of the spectral envelope curve. The decline, however, is not monotonous: the so-called formants produce local maxima.

Previous research established that spectral slope can be influenced by numerous speech factors. A flatter spectrum with a greater proportion of energy in higher frequencies can be a consequence of creaky phonation [1], greater vocal effort [2], [3], the so-called Lombard effect [4], greater linguistic prominence [3], [5], [6], [7], or emotions, particularly anger [8], [9]. Contrariwise, a breathy, silent or calm voice exhibits a steeper spectral slope, i.e., less energy in higher frequencies.

It is apparent from the definition of spectral slope, that it is a two-dimensional feature (frequency × energy). Therefore, to be able to compare the slope of different utterances, speakers or speech sounds, it is necessary to quantify it in a replicable manner with one or more values. Researchers have been trying to design various methods, often serving different applicational purposes. The measurements began with introduction of the Long-Term Average Spectra (LTAS) and were used for characterization of vocal pathologies or abnormal phonation types [1] or for speaker identification [10]. Short-term measurements on individual segments for the purposes of quantifying prominence or word stress were introduced later [3], [5], [6], [7].

1.1. Overview of spectral slope measurements

One of the first established methods was that of Hammarberg *et al.* [1]. Their goal was to find parameters correlating with voice quality and phonatory settings. According to their results, spectral slope plays a role in almost all of their investigated factors. Consequently, the term *Hammarberg index* was coined for this method of quantifying spectral slope as a difference between sound pressure level maxima in LTAS frequency bands.

The so-called α-measure originates also from LTAS measurements to diagnose vocal pathologies for medical purposes. It was first used by [11] who discovered that the values of α were correlated with change in voice quality. This method was taken up by numerous researchers in the field, e.g., [12], [13], [14], [15] who confirm the relationship between α-measure and phonation type, vocal effort or, in the last case, type of singing voice.

Another possible approach seems to be modelling the spectrum by linear regression. This was used, for instance, by [16] and [9]. The latter confirm the usefulness of the regression model for recognition of spectral slope changes in emotionally expressive speech.

So far, all the methods were used primarily to assess long-term spectra extracted from longer utterances. However, there are methods that have been applied for

Přijato 30. prosince 2013, akceptováno 7. února 2014.

short-term spectral assessment. One of them was developed by Sluijter and van Heuven [3] who established the now widely accepted assumption that spectral slope is an acoustic correlate of vowel prominence. They measured the sums of spectral energy in four frequency bands, which were supposed to contain fundamental frequency (F0), and the first three formants (F1, F2 and F3), respectively. In their dataset, non-prominent (unstressed, unaccented) vowels exhibited a steeper spectral slope with less energy in higher frequencies.

Another short-term method of measuring spectral characteristics exploits differences between various peaks in the spectrum (e.g., first or second harmonics, or first three formants): [17], [18]. Contrary to the previously mentioned methods, this type of measurement disregards higher frequencies and concentrates on the properties of the source. It has been used successfully to distinguish phonation types or modes [18], [19] or vowel articulation types [17].

Yet another group of methods is technically based on the α-measure and has been explicitly used for quantifying prominence by calculating differences of the sums of energies in two frequency bands with various low and high boundaries. In [20], this method was successfully used to distinguish stressed from unstressed vowels in different intonation contexts. Another modification was proposed by [21] who compared stressed and unstressed vowels in Czech and excluded the F0 and F2 frequency bands. The former was left out because it was hypothesized that the massive energy of F0 could obscure the fine spectral detail. F2 was excluded because of its strong link to vowel identity. This method was later used in [22] to differentiate Czech and English production of the central mid vowel *schwa*, and in [23] speaker identification potential of these methods was investigated.

Finally, a method called spectral emphasis has also been used to differentiate stressed and unstressed vowels in various languages. It takes into account the energy in the whole spectrum versus the energy in F0 band (see, e.g., [26], [27]).

All these methods show sensitivity not only to prominence, but also to vowel identity and voice characteristics of individual speakers. The aim of this paper is to compare existing spectral slope measurements with regard to their ability to detect vowel prominence using identical (directly comparable) material. Furthermore, we want to see how much the methods are influenced by the factors of speaker and vowel quality and what their speaker discriminating potential is.

2. Method

2.1. Speech corpus

Since prominence can be an elusive phenomenon and its manifestations in natural speech can be extremely variable, a highly controlled laboratory corpus was used. 12 male native Czech speakers aged 20 to 30 read a list of 45 pseudowords. These consisted of three syllables and their form was CVCVCV, where C was one of the consonants /h, m, t/ and V one of Czech short vowels /a, e, i, o, u/, e.g., "hahaha", "mememe" or "tototo" (15 possible combinations). The five Czech short vowels /a, e, i, o, u/ are considered to be particularly suitable for this kind of analysis, since they represent the most frequent vowel types in the languages of the world. Moreover, they display remarkable phonetic stability, i.e., no systematic reduction in non-prominent positions [28: 324].

In addition to that, one of the syllables was typographically marked as prominent (e.g., "taTAta"). The subjects were instructed to pronounce the marked syllable as stressed or prominent. Examples from natural Czech words or phrases were supplied and the subjects were encouraged to rehearse it before recording. Acceptability of the production was checked. Three sets of pseudowords were thus formed, with prominence on the first, second and third syllable, respectively. The ratio of prominent to non-prominent syllables in the sets was, therefore, 1:2.

The recordings were made in the sound-treated studio of the Institute of Phonetics with an electret microphone IMG ECM 2000, soundcard SB Audigy 2 ZS, 32-kHz sampling frequency and 16-bit resolution.

In this way we obtained 1620 individual vowels (12 speakers, 45 words and three vowels in each word). Afterwards, all the vowels were auditorily assessed and tokens containing external noise, non-standard vowel quality or tokens pronounced with incorrect prominence placement were excluded from the analyses; ultimately 1536 vowels were analyzed. The vowels were segmented automatically using the Prague Labeller software [24] in the Praat environment [25] and boundaries were manually corrected by experienced phoneticians. All measurements were taken in the middle third of the vowel duration to reduce potential transition effects from the consonants.

In preliminary investigations, some divergences were found for different analysis methods and particularly different softwares. Therefore, with comparability in mind, we aimed to use the same software, Praat, for all of the measurements with the exception of one (A1*-A2*).

2.2. Measuring the spectral slope

The procedure first included creating a spectral slice from the selected middle third of the vowel (i.e., an average frequency spectrum in the time selection). Several of the measurements were taken directly from the spectrum object, the Hammarberg index and the linear regression had to be measured from an LTAS spectrum which expresses the logarithmic power spectral density as a function of frequency using a 100 Hz bandwidth.

The Hammarberg index (**HI**) was measured as a difference between maxima in frequency bands of 0–2 kHz and 2–5 kHz; contrary to [1], the third band (5–8 kHz) was not used since it is hypothesized not to contain any relevant speech information and rather be influenced by characteristics of the microphone and the recording situation.

The α-measure was computed as a difference between sums of energies in the bands 0–1 kHz and 1–5 kHz (the highest boundary value is taken from [11]).

Linear regression (**LinReg**) was computed in the frequency band of 500–3000 Hz, following [16]. The logarithmic frequency scale and robust fit method were used.

Sluijter and van Heuven's method [3] involves sums of energies in four frequency bands: 0–0.5 kHz, 0.5–1 kHz, 1–2 kHz and 2–4 kHz. Since we wanted a one-dimensional feature to compare it with the others, a least squares linear fit to the four values was used (**4B-LinFit**).

Band energy differences were measured using four settings. **BgNoF0** (band difference with gap and excluded F0, following [21]): 0.35–1.1 kHz and 2.3–5.5 kHz; **Bg**: 0–1.1 kHz and 2.3–5.5 kHz. We also tried a floating pivot determined by the F2 frequency, hence **BpNoF0**: 0.35–F2, F2–5.5 kHz and **Bp**: 0–F2, F2–5.5 kHz.

Spectral emphasis (**SpEm**) was calculated as the difference between energy in the whole spectrum up to the Nyquist frequency, i.e., 0–16 kHz and the energy in the band 0–1.43*F0.

Both F0 and F2 were also measured with Praat in the middle third of the corresponding vowel duration.

With regard to the spectral peak differences, it is the only method which cannot be effectively computed in Praat as yet. For characterizing prominence, the difference between amplitudes of the first two formants has been used by [29], corrected for formant frequency effects using the formula from [17] (abbreviated as **A1*-A2***). These values were extracted with help of the VoiceSauce software [30] using default settings (LPC formant detection).

Table 1: Overview of methods used in this study for quantifying spectral slope. F0$_{\mathrm{freq}}$, F1$_{\mathrm{freq}}$ and F2$_{\mathrm{freq}}$ denote frequencies of F0, F1 and F2, respectively

Method	Frequency bands used (Hz)	Computation method
HI	0–2 000, 2 000–5 000	difference of max.
α	0–1 000, 1 000–5 000	difference of sums
LinReg	500–3 000	lin. regr. of spectr.
4B-LinFit	0–500, 500–1 000, 1 000–2 000, 2 000–4 000	linear fit of sums
BgNoF0	350–1 100, 2 300–5 500	difference of sums
Bg	0–1 100, 2 300–5 500	difference of sums
BpNoF0	350–F2$_{\mathrm{freq}}$, F2$_{\mathrm{freq}}$–5 500	difference of sums
Bp	0–F2$_{\mathrm{freq}}$, F2$_{\mathrm{freq}}$–5 500	difference of sums
SpEm	0–16 000, 0–1.43·F0$_{\mathrm{freq}}$	difference of sums
A1*-A2*	F1$_{\mathrm{freq}}$, F2$_{\mathrm{freq}}$	difference

Table 1 above shows the overview of all methods considered in this paper. All parameters were statistically assessed by one- or two-way ANOVAs for independent measures with Tukey HSD post-hoc tests.

Table 2: Results of ANOVAs for all examined methods and factors Speaker, Vowel and Prominence. The least sensitivity to Speaker and Vowel and the highest sensitivity to Prominence is highlighted in **bold**, the opposite result is marked by *italic*

Method	Factor in ANOVAs					
	Speaker		Vowel		Prominence	
	F	p	F	p	F	p
HI	24.3	<0.001	361.3	<0.001	19.9	<0.001
α	11.7	<0.001	712.7	<0.001	14.5	<0.001
LinReg	**5.7**	<0.001	*1 167.7*	<0.001	*3.5*	n. s.
4B-LinFit	23.5	<0.001	470.2	<0.001	24.8	<0.001
BgNoF0	20.4	<0.001	400.5	<0.001	4	0.04
Bg	24.9	<0.001	414.3	<0.001	22.5	<0.001
BpNoF0	15	<0.001	194.3	<0.001	4.4	0.03
Bp	19.4	<0.001	339.7	<0.001	**27**	<0.001
SpEm	*115.6*	<0.001	**19.4**	<0.001	6	0.01
A1*-A2*	7.1	<0.001	257.3	<0.001	4.8	0.03

3. Results

3.1. Sensitivity of the methods to speaker, vowel and prominence

Table 2 shows the results of one-way ANOVAs for all ten examined methods and three different factors – Speaker (M1–M12), Vowel (a, e, i, o, u) and Prominence (prominent, non-prominent).

It is obvious from the displayed values that all of the methods are highly sensitive to both speaker and vowel identity with p values being lower than 0.001 in all cases. LinReg is least sensitive to speaker and at the same time most sensitive to vowel identity (by far), while exactly the opposite is true for spectral emphasis (SpEm).

With respect to prominence, all but one of the methods were able to distinguish the two levels of prominence with statistical significance. The most successful was band energy difference with floating F2 pivot (Bp: 0–F2$_{\mathrm{freq}}$, F2$_{\mathrm{freq}}$–5 500 Hz). The second best method seems to be the linear fit to the four frequency band energies (4B-LinFit), although compared to Bp, it is more sensitive to both speaker and vowel identity. Third ranked the Bg method, i.e., band energy difference with a gap between 1 100 and 2 300 Hz. Both band energy differences with excluded F0 (BgNoF0 and BpNoF0) performed worse by an order of magnitude.

On the other hand, LinReg performed worst for differentiating prominence, it was the only parameter with statistical significance above $p = 0.05$. It seems to be dependent mainly on vowel identity rather than on speaker or prominence.

Table 3: Correlation matrix of all variables with statistically significant and meaningful coefficients ($r < -0.1$ or $r > 0.1$)

	HI	α	LinReg	4B-LinFit	BgNoF0	Bg	BpNoF0	Bp	SpEm	A1*-A2*
HI		−0.76	−0.45	**−0.94**	−0.84	**−0.95**	−0.56	−0.64	−0.27	0.61
α	−0.76			0.88	0.62	0.78	0.68	0.81	0.33	−0.69
LinReg	−0.45			0.36	0.57	0.44				−0.10
4BLinFit	**−0.94**	0.88	0.36		0.85	**0.96**	0.63	0.69	0.32	−0.64
BgNoF0	−0.84	0.62	0.57	0.85		**0.91**	0.59	0.42		−0.51
Bg	**−0.95**	0.78	0.44	**0.96**	**0.91**		0.59	0.64	0.26	−0.61
BpNoF0	−0.56	0.68		0.63	0.59	0.59		0.86		−0.65
Bp	−0.64	0.81		0.69	0.42	0.64	0.86		0.38	−0.69
SpEm	−0.27	0.33		0.32		0.26		0.38		−0.17
A1*-A2*	0.61	−0.69	−0.10	−0.64	−0.51	−0.61	−0.65	−0.69	−0.17	

Table 3 shows the correlation matrix of all variables used in the statistical analyses. As we can see, some of the variables correlate highly. In the case of 4B-LinFit, band difference with gap (Bg) and Hammarberg index the correlation with one another explains more than 90 % of the variance, although their results in Table 2 were quite different. Also in the cases of Bg vs. BgNoF0 and Bp vs. BpNoF0 we can see a high correlation, but at the same time a substantial difference in success in prominence detection. The least correlated parameter is LinReg followed by spectral emphasis.

3.2. Contribution of individual vowels to prominence differences

Since the five Czech vowels have distinct spectral qualities, it can be assumed that not all of them contribute to the same extent to differentiating prominent from non-prominent vowels. Therefore, two-way ANOVAs were conducted using the factors Vowel and Prominence, to see which vowels are enabling the prominence difference to be measured by the examined methods. The results are shown in Table 4 below.

It is obvious from the table that the vowel /u/ does not show any differences between its prominent and non-prominent realization in any of the measurements taken. /o/ is also problematic: only two of the methods were able to differentiate the prominence. On the other hand, the most useful vowel for measuring prominence seems to be /i/ with seven effective methods, followed by /e/ with five and /a/ with four.

The two most successful methods were 4B-LinFit and Bp (the same as in one-way ANOVA with all vowels pooled), they both distinguished prominence on four vowels, i.e., /a/, /e/, /i/ and /o/. The α-measure also seems to be quite reliable: it distinguished prominent from non-prominent realizations of /a/, /e/ and /i/.

Table 4: Tukey post-hoc test results on two-way ANOVAs with factors Vowel and Prominence. Stars indicate significance of the differences between prominent and non-prominent realizations of the corresponding vowel: *** → $p <$ 0.001; ** → $0.001 < p < 0.01$; * → $0.01 < p < 0.05$; n. s. → not significant

Method	Vowel				
	a	e	i	o	u
HI	n. s.	***	**	n. s.	n. s.
α	***	**	*	n. s.	n. s.
LinReg	n. s.	n. s.	**	n. s.	n. s.
4B-LinFit	***	***	**	*	n. s.
BgNoF0	n. s.	n. s.	n. s.	n. s.	n. s.
Bg	n. s.	***	**	n. s.	n. s.
BpNoF0	n. s.	n. s.	n. s.	n. s.	n. s.
Bp	**	***	***	*	n. s.
SpEm	*	n. s.	n. s.	n. s.	n. s.
A1*-A2*	n. s.	n. s.	*	n. s.	n. s.

3.3. Speaker discriminating potential of the methods

Since the methods have been shown to be sensitive to speaker factor, it seems reasonable also to assess their speaker discriminating potential. In this set of analyses, LinReg was a priori excluded, since in the first tests it did not seem to differentiate among speakers as reliably as the other methods.

In the first analysis we decided to express the identity of the speaker as a vector of his average values in individual measurement methods. In order to control for vowel differences, we chose only values of /e/, since it is the most frequent Czech vowel and has been shown to have the highest speaker identification potential in Czech [31].

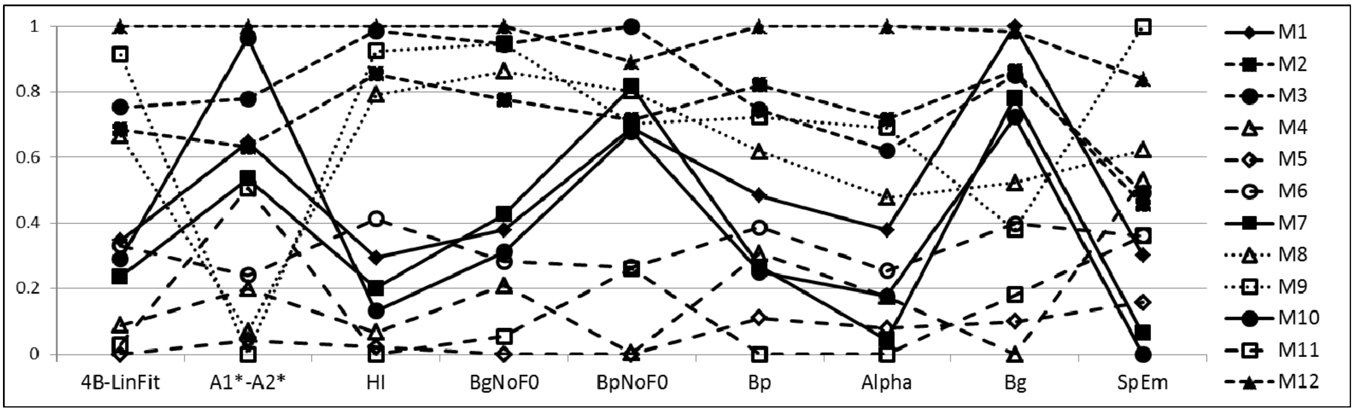


Figure 1: Speaker vectors of normalized values from nine methods used over the vowel /e/. Value 1 is assigned to speaker with the highest value, 0 to lowest. Speakers are clustered into four groups signalled by different line types

In view of the fact that the methods yielded different values in terms of their dimension and variability, we chose to normalize the data – speakers with highest and lowest value of the given measurement method were set to constitute 1 and 0, respectively, with the remaining speakers ordered in between according to the original relations between the values. The resulting vectors are shown in Figure 1. Clustering into groups by visual inspection was attempted, resulting in four quite distinct speakers groups. Group 1 consists of speakers M2, M3 and M12 who tend to vary around high values, 0.6–1. M12 is moreover the speaker with the highest values in six cases and in the remaining three ranks second. On the opposite side there is group 2 with values ranging between 0 and 0.4 and members M4, M5, M6 and M11. Group 3 consists of speakers M1, M7 and M10 and can be characterized by high values (above 0.5) in A1*-A2*, BpNoF0 and Bg and low values (below 0.4) in the other methods. Group 4 contains only two speakers, M8 and M9, who exhibit very low values in A1*-A2* (below 0.1) and rather high values in other methods, but in some cases not as high as group 1 (Bp and Bg).

An orthogonal grouping of speakers was also attempted, where we used only one method but represented a speaker by a vector of his values for individual vowels. This time, a simple normalization by grand mean of the individual vowels was used.

In this manner we investigated SpEm, Bg, HI, 4B-LinFit and BgNoF0 which were the methods with the highest F criterion resulting from the one-way ANOVAs with the factor Speaker. Results for SpEm and Bg are displayed in Figures 2 and 3. SpEm was able to cluster the speakers into three groups – Group 1 exhibited high above average values (M5, M9 and possibly also M12) while group 2 stayed consistently below average (M7 and M10). The rest of the speakers formed group 3 which fluctuated around the grand mean. Bg, HI and 4B-LinFit classified the speakers into two groups where M2, M3, M8, M9 and M12 formed the above average set and the others range below the grand mean. The vectors of BgNoF0 revealed no apparent grouping of the speakers.

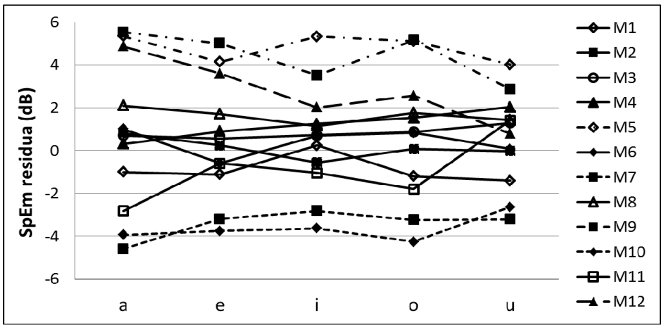


Figure 2: Speaker vectors of SpEm values for five vowels normalized by grand mean of each vowel

All of the above mentioned methods with the exception of SpEm showed a marked higher spread of the values of /i/ which produced substantial differences between groups of speakers in BgNoF0, too. /a/, /e/ and /o/ also showed inter-speaker variability while /u/ was in all methods in disaccord with values of the other vowels.

4. Discussion and conclusions

In this study we have measured short-term spectral slope of five Czech vowels using ten different methods found in current literature. It was found that the sensitivity of all investigated methods to speaker and vowel identity is indeed substantial. For the purposes of quantifying prominence, Bp exhibited the best performance in differentiating prominent from non-prominent realizations of the vowels. Hammarberg index, α-measure, 4B-LinFit and Bg also belong to successful procedures for quantifying prominence. The remaining measures (A1*-A2*, SpEm, LinReg, BgNoF0 and BpNoF0) were either not significant at all (as in the case of LinReg) or less successful by an order of magnitude, see Table 2. Since our speech material con-

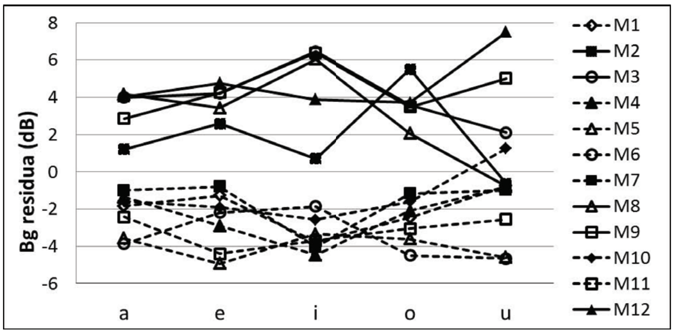


Figure 3: Speaker vectors of Bg values for five vowels normalized by grand mean of each vowel

tained explicit prominence, it can be argued that on natural speech the differences might be smaller and the less successful methods could fail completely.

Bp was also the best method when investigating prominent and non-prominent realizations of individual vowels, it was able to differentiate among them in all instances with the only exception of /u/ (see Table 4).

Our results of 4B-LinFit seem to be in accord with findings of [3] and [5], although contrary to their propositions, the importance of the F0 band to the spectral slope differences has to be emphasized. Their own illustrations indicate that it is precisely the first frequency band (0–500 Hz) which makes the spectral slope dissimilar (see [3: 2481]). In the higher bands, the difference lies only in relative energy level and the slope of the line is very similar.

Excluding the F0 frequency band also decreases the performance of the band energy methods. Both BgNoF0 and BpNoF0 were less effective than Bg and Bp, which could most prominently be seen in the individual vowel analysis, where BgNoF0 and BpNoF0 could not discriminate prominent from non-prominent realizations of any vowel.

In agreement with [29], the results of A1*-A2* regarding prominence are not very satisfactory, although its speaker and vowel sensitivity is also relatively low. The authors obtained better results when they abandoned the idea of formant amplitude corrections for frequency effects, but they infer that this method is influenced rather by the change in vowel timbre (e.g., centralization) than by changes in prominence.

Interestingly, LinReg performed worst for quantifying prominence. It seems that taking all parts of the spectra into account is not beneficial in this regard. On the other hand, it appears to be extremely well suited for vowel identification and moreover not that much influenced by speaker identity.

Contrariwise, spectral emphasis has performed best for speaker identification in the first set of tests (Table 2) since only 12 out of the possible 66 pairs of speakers were found non-significantly different in post-hoc tests.

Concerning individual vowels, prominence is manifested most on /i/ and least on /u/ (see Table 4: none of the methods was effective on /u/). This is probably a result of different inherent spectral properties of the vowels: /i/ has more energy clustered in higher frequencies (around 2 000 to 4 000 Hz), while /u/ is the opposite with energy clustered in lower frequencies. It may be hypothesized that having less energy in high frequencies leads also to smaller possibility of prominence marking in these frequencies. And as /u/ is also a rather rare vowel in Czech texts, it can be advisable to exclude it from spectral slope measurements with the aim of quantifying prominence.

The speaker discriminating potential of the methods was also evaluated. A vector of normalized values from nine methods (LinReg was excluded) and the vowel /e/ distinguished four groups of speakers (Figure 1). The vector of normalized vowel values for spectral emphasis was also effective (Figure 2); nevertheless, it separated only three groups of speakers. The other methods in this analysis categorised the speakers into two groups.

It is noteworthy that some of the speakers were clustered together by all classifications: M4, M6 and M11; M2 and M3; M7 and M10. It seems reasonable to argue that these speakers share some common spectral properties of their vowels and it would be interesting to investigate whether this clustering coincides, for example, with any perceived similarities of the speaker's voices.

The role of including F0 energy for speaker identification is not as clear as in the case of prominence. On the one hand, SpEm which emphasizes F0 performed best and LinReg with excluded F0 band worst, on the other hand, the two other methods with excluded F0 (BgNoF0 and BpNoF0) display average results.

Furthermore, it appears that the vowel /u/ is not useful for speaker identification either. This may resonate with the reasons stated above concerning its inherent spectral characteristics.

The main limitation of the study rests in the nature of the investigated material – results obtained for pseudowords may not be easily generalizable to natural speech. In particular, conversational word stress could be realized in a different way from the prominence obtained in the described way from our speakers. However, the performance of the selected methods needed to be tested on a more transparent, controlled material. Their usefulness and suitability will be further ascertained on spontaneous speech.

To conclude, when computing short-term spectral slope with the aim of quantifying prominence, vowel and speaker identity plays a significant role. It is, therefore, necessary to control for these effects and compare vowels and speakers separately. For future research a method of measurement which would normalize the differences between vowels and/or speakers seems to be indispensable.

Acknowledgements

The first author was funded by Internal Grant of the Faculty of Arts, VG192, and the second author by a grant of the Czech Science Foundation (GACR 406/12/0298). The support of the Programme of Scientific Areas Development at Charles University in Prague (PRVOUK), subsection 10 – Linguistics: Social Group Variation is also acknowledged.

References

[1] Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., Wedin, L.: Perceptual and acoustic correlates of abnormal voice qualities, *Acta Otolaryngologica*, 90, p. 441–451, 1980.

[2] Doddington, G.: Speaker Recognition – Identifying People by their Voices, *Proc. IEEE*, 73, p. 1651–1664, 1985.

[3] Sluijter, A., van Heuven, V.: Spectral balance as an acoustic correlate of linguistic stress, *Journal of the Acoustical Society of America* 100, p. 2471–2485, 1996.

[4] Bořil, H., Hansen, J. H. L.: Unsupervised Equalization of Lombard Effect for Speech Recognition in Noisy Adverse Environments, *IEEE Transactions on Audio, Speech, and Language Processing*, 18(6), p. 1379–1393, 2010.

[5] Sluijter, A., van Heuven, V., Pacilly, J.: Spectral balance as a cue in the perception of linguistic stress, *Journal of the Acoustical Society of America*, 101(1), p. 503–513, 1997.

[6] Campbell, N., Beckman, M.: Stress, prominence, and spectral tilt, *ESCA Workshop on Intonation: Theory, Models and Applications*, p. 67–70, Athens: University of Athens, 1997.

[7] Heldner, M.: Spectral emphasis as a perceptual cue to prominence, *TMH-QPSR*, 42, p. 51–57, 2001.

[8] Banse, R., Scherer, K. R.: Acoustic profiles in vocal emotion expression, *Journal of Personality and Social Psychology*, 70(3), p. 614–636, 1996.

[9] Tamarit, L., Goudbeek, M., Scherer, K. R.: Spectral slope measurements in emotionally expressive speech, *ISCA Tutorials and Research Workshop*, Aalborg, Denmark, 2008.

[10] Nolan, F.: *The Phonetic Bases of Speaker Recognition*, Cambridge: Cambridge University Press, 1983, reissued in 2009.

[11] Frøkjær-Jensen, B., Prytz, S.: Registration of voice quality, *Technical Review*, 3, p. 3–17, 1976.

[12] Kitzing, P.: LTAS-criteria pertinent to the measurement of voice quality, *Journal of Phonetics*, 14, p. 477–482, 1986.

[13] Sundberg, J., Nordenberg, M.: Effects of vocal loudness variation on spectrum balance as reflected by the alpha measure of long-term-average spectra of speech, *Journal of the Acoustical Society of America*, 120(1), p. 453–457, 2006.

[14] Boersma, P., Kovači, G.: Spectral characteristics of three styles of Croatian folk singing, *Journal of the Acoustical Society of America*, 119(3), p. 1805–1816, 2006.

[15] Plag, I., Kunter, G., Schramm, M.: Acoustic correlates of primary and secondary stress in North American English, *Journal of Phonetics*, 39, p. 362–374, 2011.

[16] Kochanski, G., Grabe, E., Coleman, J., Rosner, B.: Loudness predicts prominence: fundamental frequency lends little, *Journal of the Acoustical Society of America*, 118(2), p. 1038–1054, 2005.

[17] Fulop, S. A., Kari, E., Ladefoged, P.: An acoustic study of the tongue root contrast in Degema vowels, *Phonetica*, 55, p. 80–98, 1998.

[18] Hanson, H. M., Stevens, K. N., Kuo, H-K. J., Chen, M. Y., Slifka, J.: Towards models of phonation, *Journal of Phonetics*, 29, p. 451–480, 2001.

[19] Mills, T.: *Speech Motor Control Variables in the Production of Voicing Contrasts and Emphatic Accent*, PhD thesis, Edinburgh: University of Edinburgh, 2009.

[20] Prieto, P., Ortega-Llebaria, M.: Stress and Accent in Catalan and Spanish: Patterns of duration, vowel quality, overall intensity, and spectral balance, *Proceedings of Speech Prosody 2006*, p. 337-340, Dresden, 2006.

[21] Volín, J., Zimmermann, J.: Spectral slope parameters and detection of word stress, *Technical Computing Prague*, p. 125–130, Praha, 2011.

[22] Volín, J., Weingartová, L., Skarnitzl, R: Spectral Characteristics of Schwa in Czech Accented English, *Research in Language*, 11(1), p. 31–39, 2013.

[23] Weingartová, L., Volín, J.: Spectral Measurements of Vowels for Speaker Identification in Czech, *Studies in Applied Linguistics*, 1, p. 21–36, 2013.

[24] Pollák, P., Volín, J., Skarnitzl, R.: HMM-based phonetic segmentation in Praat environment, *Proceedings of the XIIth International Conference Speech and computer – SPECOM 2007*, p. 537–541, 2007.

[25] Boersma, P., Weenink, D.: *Praat: doing phonetics by computer* [computer program], version 5.3.60, retrieved from: http://www.praat.org/, 2013.

[26] Eriksson, A., Thunberg, G. C., Traunmüller, H.: Syllable prominence: A matter of vocal effort, phonetic distinctness and top-down processing, *Proceedings of EuroSpeech 2001*, p. 399–402, 2001.

[27] Barbosa, P. A., Eriksson, A., Åkesson, J.: On the Robustness of some Acoustic Parameters for Signalling Word Stress across Styles in Brazilian Portuguese, *Proceedings of Interspeech 2013*, Lyon: ISCA, 2013.

[28] Palková, Z.: Fonetika a fonologie češtiny. Praha: Karolinum, 1994.

[29] Ortega-Llebaria, M., Prieto, P.: Acoustic correlates of stress in Central Catalan and Castilian Spanish, *Language and Speech*, 54(1), p. 73–97, 2011.

[30] Shue, Y.: VoiceSauce: *A program for voice analysis* [computer program], version 1.14, retrieved from: http://www.seas.ucla.edu/spapl/voicesauce/, 2013.

[31] Skarnitzl, R., Vaňková, J., Weingartová, L.: Speaker discrimination using short- and long-term segmental information in vowels, *Proceedings of IAFPA 2012,* Santander, 2012.

Akustické listy, **20**(1), březen 2014, str. 13–16 © ČsAS

The Effect of Installation Slots on Airborne Sound Insulation of Masonry Walls

Vliv instalačních drážek na neprůzvučnost zděných stěn

Pavel Heinrich[a] and Pavel Rubáš[b]

[a]HELUZ cihlářský průmysl, v. o. s., Dolní Bukovsko čp. 295, heinrich@heluz.cz
[b]Technický a zkušební ústav stavební Praha, s. p., Praha 9 – Prosek, Prosecká 811, rubas@tzus.cz

The demands of acoustic properties of buildings are continually changing with human development and technical activities. Of particular concern in this respect are the requirements for suppressing the adverse effects of sound and, in certain cases, the opposite: the requirements to ensure sound's audibility and clarity. Increasingly, intense vehicular traffic is the main culprit for increased noise levels in the environment, and thus the requirement emerges for sound insulation of building envelopes. Users of buildings require acoustic comfort, so that they cease to perceive the source of disturbing noise from adjacent rooms or apartments as a normal phenomenon. Disturbing sound is created by oscillating environmental particles and spreads by wave motion, in which the environment is affected by shared oscillatory motion. Building dwellers feel the potency, height, duration, information content and nonspecific effects of exposure to sound pressure levels, both physiologically and psychologically. Sound insulation is an important feature of building structures, leading to the loss of sound power during its transmission via the air through the structure.

1. Sound reduction theory of single structures with slots

In terms of acoustics, single-layer building structures are structures which oscillate as a whole. They consist of either one type of construction material, or multiple layers of similar materials with similar acoustic characteristics (e.g., brick walls). For monolayer structures, it is a constant that their apparent sound reduction index increases with their surface density and the frequency of incidental acoustic energy. Intervention of masonry walls (e.g., wall sewage or water pipes, electrical box installations, etc.) can decrease their apparent sound reduction index due to a local weakening of the wall. It is recommended that installation ducts be appropriately adjusted using insulating materials, or not implemented at all. The effect of installation slots on the apparent sound reduction index of inter-apartment masonry walls has not yet been satisfactorily addressed; therefore, a case study using a HELUZ AKU 25 MK brick wall, intended for load-bearing and non-bearing structures with a sound insulating feature, was carried out within the construction acoustics laboratory of the firm HELUZ Brick Industry, a public company.

A simple wall with slots can be considered as a composite structure, as opposed to an original wall with no slots, inhomogeneous in its surface area and consisting of more than one part each with a different level of sound insulation. The weighted apparent sound reduction index of a structure's inhomogeneous surface area can be calculated thus:

$$R'_{\mathrm{w}} = 10 \cdot \lg(S) - 10 \cdot \lg \sum_{i=1}^{n} S_i \cdot 10^{-R'_{\mathrm{w},i}/10}, \qquad (1)$$

in which
S is the surface area of the composite construction (m^2);
S_i the surface area of the sub-element (m^2);
$R'_{\mathrm{w},i}$ the weighted apparent sound reduction index of the sub-element (dB).

Considering the aforementioned facts with an inter-apartment HELUZ AKU MK 25 with an area of 12 m^2, an apparent sound reduction index is achieved of $R'_{\mathrm{w}} = 55$ dB, weakened by slots in 10 % (1.2 m^2) of its surface area, with a decline in the R'_{w} value of 2 dB ($R'_{\mathrm{w}} = 53$ dB) in these slots; it would then be theoretically concluded that, using the above formula, the R'_{w} throughout the entire structure is reduced to 54 dB.

2. Practical verification on the structure in question

The theory has been verified by a series of measurements within construction laboratory conditions. The measurements were performed in a laboratory equipped with flanking transmission paths. It was measured in the form of airborne sound insulation according to EN ISO 140-4 "Acoustics – Measurement of sound insulation in buildings and of building elements – Part 4: Field measurement of airborne sound insulation between rooms". The main result of the test, which is objectively applied to the measured structure, is the apparent sound reduction index R'_{w}.

The tested structure in various weakened states was placed between the source and receiving rooms. The airborne sound transmission loss is expressed by (apparent sound reduction) R', which is determined by the formula:

$$R' = L_1 - L_2 + 10 \cdot \lg \frac{S}{A}, \qquad (2)$$

Přijato 17. února 2014, akceptováno 9. března 2014.

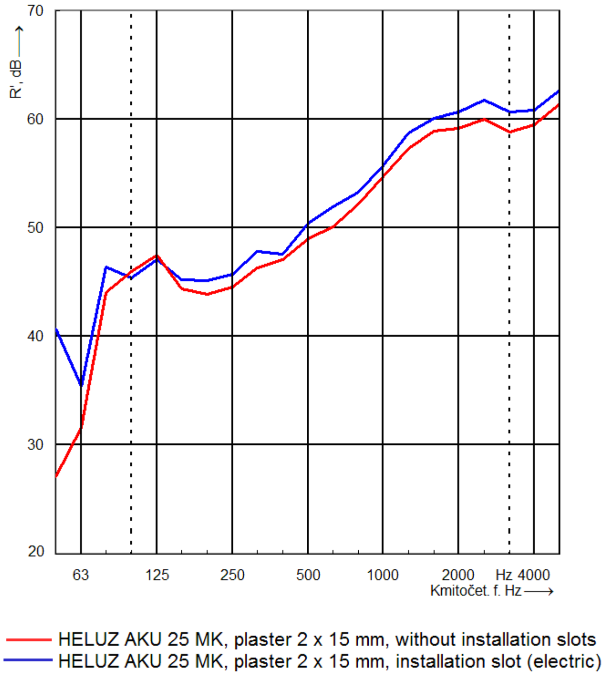


Figure 1: Comparison of apparent sound reduction R' (option 1 and option 2)

in which
L_1 is the sound pressure level in the source room (dB);
L_2 the sound pressure in the receiving room (dB);
S the surface area of the tested partition structure (m^2);
A the equivalent absorption area of the receiving room (m^2).

It is determined from the measured reverberation time according to the relationship:

$$A = 0{,}16 \cdot \frac{V}{T}, \tag{3}$$

in which
V is the volume of the receiving room (m^3);
T the reverberation time of the receiving room (s).

The essence of the test is to measure the difference in sound pressure levels in the source and receiving rooms, during the activity of a sound source emitting a broadband noise signal. The absorption in the receiving room is taken into account by the correction element, which was determined from measurements of reverberation time in the receiving room. Measurements were carried out in construction laboratory conditions according to ČSN EN ISO 140-4 in third octave frequency bands within the 50 Hz to 5 000 Hz range. The measured frequency-dependent transmission loss values R' were compared with the values of the reference curve, as defined by ČSN EN ISO 717-1. The result of the evaluation is a single-digit value – the apparent sound reduction index R'_w. Furthermore, the spectrum adaptation terms (C, C_tr) were determined; which can, according to the type of spectrum of the noise source in real conditions, be attributed to the value R'_w. The value of C is a factor for pink noise weighted by function A, which roughly corresponds to the noise spectrum of apartment activity or highway traffic noise. The factor C_tr refers to the weighted spectrum of urban noise in towns and villages. These factors (C, C_tr) are recorded simultaneously with the variable R'_w and true for a basic frequency range of 100–3 150 Hz. Also determined as additional factors were the spectrum adjustment factors for the extended frequency range $C_{100\text{–}5\,000}$ and $C_{\mathrm{tr}100\text{–}5\,000}$, which are related to the frequency range of 100–5 000 Hz.

In total, five variants were measured within construction laboratory conditions. The first measured data shows the apparent sound reduction index of the original HELUZ AKU MK 25 walls without intervention. Subsequently, installation slots were gradually added and the sound insulation of the modified wall was measured. Even the very first modification – a one-sided wiring slot – reduced the R'_w value by 1 dB. Further modifications reduced the construction insulation only by a few tenths of a decibel. This finding has not been reflected in the weighted value, since, in terms of the ČSN EN ISO 717-1 standard, the resulting value is rounded down to a whole number, meaning that detected values of 55.9 dB and 55.0 dB are identically rounded to 55 dB. Setting aside the implications of the standard and the using of non-rounded figures, we concluded that the difference between Option 1 (the original wall) and Option 2 (with maximum wall installation slots) is 1.5 dB, thus one can reasonably argue that the result of additional installation slots is approaching a loss of 2 dB!

Table 1: The $R'_\mathrm{w}(C, C_\mathrm{tr})$ recorded values of various installation slot options of weakened walls

Option		$R'_\mathrm{w}(C, C_\mathrm{tr})$
1	no modifications to wall	55 (−1, −3) dB
2	single-sided wiring slot	54 (−1, −3) dB
3	double-sided wiring slot, electri-cal box slot is not on the same side	54 (−1, −3) dB
4	double-sided wiring slot, electrical box slot is not on the same side + second slot for room wiring	54 (−1, −3) dB
5	double-sided wiring slot, electrical box slot is not on the same side + water installation slot	54 (−1, −3) dB

3. General guidelines for the implementation of slots

Generally, technical installation slots in brickwork are made in such a way that the masonry is affected to the

Figure 2: Model slot/switch slot-making and brick cutting using a grinder

minimum extent. The implementation of slots affects the wall's structural and acoustic properties; this was the main focus of the study performed. In practice, cable wiring slots are preferably made by a groove planer. Alternatively, slots can be made using a diamond wheel angle grinder and, subsequently, knocking out the brickwork. Various electrical box sockets are made using trepans. The model wiring slot is shown in Figure 2. Improper work procedures in slot making, such as the use of a demolition hammer, result in a significant weakening of the cross sectional area of the wall, which has the effect of reducing the local load-bearing capacity of the masonry and the deterioration of its acoustic or thermal insulation properties and only annoyingly adds extra work to encase these slots. The permissible slot dimensions with no required static assessment are given in Table 2. Other measures recommended for slot-making can be obtained from the literature (see [3], [4]).

Table 2: Masonry slot dimensions without structural assessment

	Vertical slots		Horizontal and diagonal slots	
wall thickness (mm)	max. depth (mm)	max. width (mm)	max. depth at unlimited length (mm)	max. depth at length up to 1 250 mm
up to 115	30	100	0	0
116–175	30	125	0	15
176–225	30	150	10	20
226–300	30	175	15	25
over 300	30	200	20	30

4. Masonry slot-making during testing

The procedure of slot-making in masonry tested in this experiment was different from usual construction practices, where the technical installation slots are made before plastering. In this case, due to technical testing procedures, slots were made in an already plastered wall, which served as a reference sample. The slots remained exposed during measurement, validating the credibility of the results. The slots were made by cutting the wall with a diamond-blade angle grinder and knocking out the masonry using a chisel. The making of a slot is shown in Fig. 3. The geometry of the slots made is shown in Fig. 4. Fig. 5 captures the side of the wall in the source room after completion of the tests, and Fig. 6 is a detail of the wiring slot finish. The three slots made simulate electrical duct slots in their execution. The fourth slot was given larger dimensions to determine the point of significant wall weakening due to ill-planned slot-making, or when conducting technical installations of larger sizes (water, cable bundle, etc.).

5. Conclusion

Certification of sound insulation of components or building partition elements is not mandatory; CE marked products are guaranteed free movement in the EU and can be sold on the single market with no further trade barriers. The same situation does not apply in the design of buildings and incorporation of products in construction. Here, the regulations (typically local building acts and building codes) are currently not harmonized, and the requirements for sound insulation of built-in structures are different in every EU state. In this country, the standard ČSN 73 0532:2010 "Acoustics – Protection against noise in buildings and evaluation of acoustic properties of building elements – Requirements" can be used for making a proper design. Specific criteria can be found in the national standard - minimum requirements for structural airborne sound insulation of internal and external walls.

During the design and project preparation stages, the measured or calculated laboratory values of sound insulation of building structures R'_w are used, and their approximate conversion to the apparent sound reduction index R'_w is carried out according to the formula $R'_\mathrm{w} = R_\mathrm{w} - k_1$, where k_1 is the correction dependant on flanking transmitting (min. 2 dB, for more detail see ČSN 73 0532:2010 clause 5.1). Our wall measurements were carried out with the slots open. One situation common on actual construction sites is that, after placing the wire in the slot, the slot is filled with mortar, which generally has a greater bulk density than the original ceramic material (often with cavities) that has been cut out of the slot. The problem remains in case of distribution boxes, especially if they are situated against each other, and light plastic pipe distributions. Therefore, the authors plan further research in this area.

References

[1] Čechura, J.: *Structural/constructive Physics 1: Acoustic of Building Structures*, Prague: CTU Publishing House, 1996

[2] Beranek, L.: *Noise Reduction*, Prague: State publishing house of technical literature, 1965.

[3] ČSN EN *1996-1-1 Eurocode 6: Masonry Structures Design – Part 1-1*

[4] General Guidelines for Masonry Structures, ČNI 2007.

[5] *Technical Manual for Designers and Builders. HELUZ*, 2013.

Figure 3: Slot-making – cutting the wall and knocking out the slot

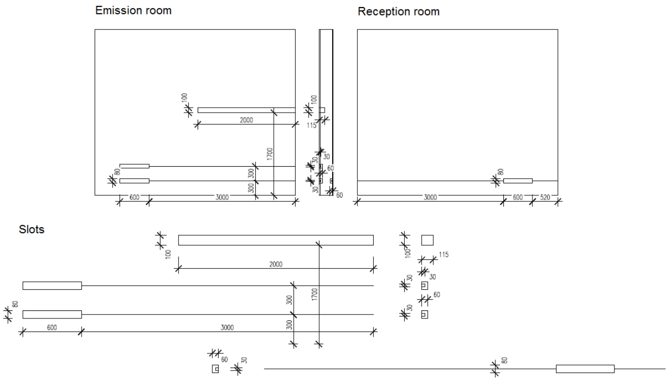


Figure 4: The geometry of slots

Figure 5: A view of the wall in the source room test laboratory, post-test completion

Figure 6: A detail of the electrical box socket area

Akustické listy: ročník 20, číslo 1 březen 2014 ISSN: 1212-4702
Vydavatel: Česká akustická společnost, Technická 2, 166 27 Praha 6 Vytisklo: Nakladatelství ČVUT, výroba
Počet stran: 20 Počet výtisků: 200
Redakční rada: M. Brothánek, O. Jiříček, J. Kozák, R. Čmejla, J. Volín
Jazyková úprava: R. Svobodová

Uzávěrka příštího čísla Akustických listů je 30. května 2014.