Spectral Measurements of Vowels for Speaker Identification in Czech

Lenka Weingartová — Jan Volín

ABSTRACT:
The expansion of telecommunication increased the availability of speech recordings which can be used in criminal investigations. Forensic science is a multidisciplinary approach that provides scientific grounds for assessing the evidence in such investigations. Forensic phonetics explores segmental (vocalic, consonantal) and suprasegmental (prosodic) speech parameters that are discriminant among speakers. There is, however, a gap between technical data-driven and linguistically informed approaches, which we attempt to bridge in this study by examining Czech vowels through rigorous computational means. Seven different methods of quantifying vocalic spectral slope were compared for the purposes of speaker identification. In forensics, the use of spectral slope is mainly limited to the long-term average spectra, which are easy to obtain, but have some serious drawbacks. Therefore, in this study, short-term spectra of Czech vowels were used: although their extraction is more laborious, they provide more speaker-specific information. Of the seven methods tested, two software predefined functions performed unsatisfactorily, while a combination of modified band density difference and band density ratio was able to differentiate among all of our speakers. The effect of vowel quality on these measures was also investigated.

KEY WORDS:
forensic phonetics, speaker identification, speaker recognition, spectral slope, spectral til

1. INTRODUCTION

When a familiar person starts talking on the phone without introducing him or herself, the recipients are usually able to determine the identity of the caller after a few seconds — possibly just after a single word of greeting. Our perceptual mechanisms instantly and automatically match the acoustics of the incoming signal to a stored pattern of the caller's voice. Speech characteristics that are idiosyncratic or unique to a speaker are explored in the field of speaker identification. One of the important practical applications lies in the area of forensic phonetics, a dynamically developing discipline, which among other things makes use of speaker identification in solving criminal cases.

Forensic phoneticians are often asked to provide expert judgements in court regarding the identity of a suspect on the basis of his or her speech sample. As the number of communication possibilities increases, audio recordings as forensic evidence are becoming more and more frequent and the need for qualified experts is growing.

Even lay listeners are usually able to determine the gender or approximate age of an unfamiliar speaker from his or her voice and notice some distinct peculiarities. In addition to that, linguistically trained experts can extract information about the

speaker's social background, education level, geographical origin, etc., as well as assess some idiosyncrasies of the voice.

The fundamental questions lying at the core of the area of speaker identification are the following:

1) What information are listeners using to pass their judgement concerning the identity of the speaker?
2) Which extractable acoustic measures of speech can be employed to perform a successful computer-aided speaker identification?

These two questions might be related — if we determine the information in the speech crucial for identification by listeners, we can employ it for an automatic identification. And vice versa — finding acoustic features that perform well in computer-aided identification experiments may enhance our knowledge about human identification abilities.

From this point of view, speech descriptors can be roughly divided into two categories — high-level and low-level features (Doddington, 1985, p. 1653). The former consist of linguistic information such as speaking style, word choice, syntactical or morphological peculiarities, etc., not to mention the content of the utterance. While these features can be successfully used by a forensic linguist, they are very difficult or nearly impossible to be extracted and analyzed by a computer. However, forensic specialists currently agree that purely auditory analysis (or "aural-perceptual" in Hollien's (2002, p. 11) terminology) is insufficient to form an expert opinion to present before court, and a joint auditory-acoustic approach is strongly recommended (see e.g. Nolan, 1990, p. 461, or French, 1994, p. 173).

Therefore, considerable attention is paid to the low-level acoustic features, e.g., the fundamental frequency, segmental parameters or descriptors of coarticulation, temporal features such as pauses, articulation rate, timing metrics, or a number of spectral features. All these characteristics are assumed to contain speaker-specific information — however, the main challenge lies in their mechanical extraction and quantification. Our study would like to contribute in this respect by testing seven general methods of spectral analysis with the aim to establish their sensitivity to speaker-specific differences.

Spectral characteristics are tied to the notion of voice quality, commonly known as timbre or colour — the combination of the source signal and its resonances in the speaker's vocal tract. Voice quality has been evaluated purely auditorily for a long time, using descriptive frameworks such as those of Laver (1980) or Hammarberg et al. (1980). Nevertheless, emergent techniques of spectral analysis made acoustic assessments possible as well.

It is reasonable to expect that spectral characteristics — since they reflect speaker's individual vocal tract physiology — are able to convey speaker-idiosyncratic information. Indeed, spectral features are used in forensics, in particular in the form of long-term average spectrum (LTAS) parameters. This approach averages the spectra over longer stretches of utterances (see, e.g., Nolan, 2009 [1983], pp. 130ff., or Doddington, 1985) and is computationally relatively straightforward.

Several other acoustic features are also used to describe spectral properties of speech — either with the goal of characterizing abnormal voices or phonation types (as in Hammarberg et al., 1980), or for the purposes of speaker identification (see for example Nolan, 2009 [1983]). The parameters used were at first mainly formant means or formant ratios. Newer approaches employ the so-called spectral slope. This term (also spectral tilt or spectral balance) refers to the gradual decay of energy towards higher frequencies in the frequency spectrum of a voice.

A change in spectral slope is associated with a change in voice quality (Hammarberg et al., 1980), in vocal effort (Doddington, 1985; Sluijter — van Heuven, 1996) or with linguistic prominence (Sluijter — van Heuven, 1996; Sluijter — van Heuven — Pacilly, 1997; Campbell — Beckman, 1997, or Heldner, 2001). In many of these studies the authors need to compensate for inter-speaker differences and normalize the data; for example, in the paper concerned with spectral slope as a correlate of emotions in speech, Tamarit, Goudbeek and Scherer (2008) have proposed computing a speaker--specific pivot, otherwise the differences between speakers might obscure the results. Our goal, on the other hand, is to exploit this information to differentiate one speaker from another.

It should be noted that in order to characterize persons' overall vocal quality, the spectral slope of a long-term spectrum needs to be measured. However, when describing syllabic prominences, short-term spectra of individual vowels are exploited.

The former approach, although widely used, has some disadvantages — first, the experimental method has to deal somehow with unvoiced regions of speech or pauses (usually eliminating them from the LTAS computations, as in Hammarberg et al., 1980; Kitzing, 1986; Löfqvist, 1986, or Tamarit — Goudbeek — Scherer, 2008). Moreover, the results can be affected by content and length of the speaker's utterance — in forensic cases, the speech samples can be completely incomparable in this regard, specifically their length is often insufficient to be reliable. Rodman et al. (2002) or Master et al. (2006), explicitly state that long-term averages of spectral characteristics require a great amount of speech data.

It could be therefore useful for practical purposes to measure comparable "chunks" of the material and to compute the spectral slope from short-term spectra of the same linguistic elements — these comparable elements are often nasals, fricatives or vowels (see, e.g., French, 1994, p. 176; Nolan, 1983, pp. 75–77, or Jackson et al., 1985).

In our study, we can benefit from the fact that Czech has relatively few vowel qualities and these are not systematically reduced. Therefore, we can compare the same category of sounds for every speaker, whether or not the utterances are the same. Even shorter speech samples can be used provided that the number of usable tokens is sufficient.

In this study, we plan to compare several methods for measuring spectral slope of Czech vowels to see which of them are sensitive enough to capture differences among speakers. At this point, we use laboratory, high-quality speech data with the aim to prevent introduction of any artifacts caused by signal distortions. If some of the outcomes are promising, the course for future research is to try them under more natural conditions (more spontaneous speech, noise, degraded signal, etc.).

2. METHODS OF MEASURING SPECTRAL SLOPE

The spectral envelope is an inherently two-dimensional curve. To quantify its slope with a single value the computational procedure has to be determined.

Perhaps the most influential method was used by Hammarberg and her colleagues (1980) and was consequently termed the Hammarberg index. The original principle was to measure the difference between sound pressure level peaks in given frequency bands.

This basis (measuring the difference between energies above and below a given frequency value) was developed further by Sluijter and van Heuven (1996), Sluijter, van Heuven and Pacilly (1997), Sundberg and Nordenberg (2006), or Boersma and Kovacic (2006).

A different approach is to fit a regression line to the spectrum. It was explored for example in the work of Kochanski et al. (2005).

A third group of methods quantifies ratios of certain peaks in the spectrum, e.g., the amplitude of the first harmonic to the amplitude of the second harmonic, the first harmonic and the third formant (as in Hanson, 1997; Hanson — Chuang, 1999), or the amplitudes of the first two formants (Fulop — Kari — Ladefoged, 1998; Mills, 2009). This approach was with moderate success used for speaker discrimination in Czech by Skarnitzl, Vaňková and Weingartová (2012), although the manual extraction of these parameters was with current technological background found to be difficult and time-consuming. Volín and Zimmermann (2011) used a method similar to the Hammarberg index measurements on Czech vowels. Although their goal was to quantify linguistic stress and not to identify speakers, they reported a significant inter-speaker variability in the spectral slope behaviour.

One of our objectives is to compare different methods with respect to their practical applicability: they should be easy to employ without extensive technical training and should be dependent only on commonly available technology and software.

3. EXPERIMENT

METHOD

The material used in this study consisted of recordings of four male native Czech speakers at the age of 20–30 years, with university education, coming from Central Bohemia. They were recorded individually in a sound-treated studio of the Institute of Phonetics in Prague. The men were selected to represent four distinctive voice types (i.e., they were deliberately chosen not to sound alike).

The speakers were asked to read a list of pseudowords in the form CVCVCV (consonants and vowels being the same in one pseudoword), where the vowel was one of the Czech five short vowels /ɪ ɛ a o u/ and the onset consonant either /m/, /t/ or /h/. These consonants were selected to represent three frequent classes: nasals, voiceless stops and fricatives. We also considered their ease of articulation and distinctiveness in spectrograms (for pre-analysis signal processing). Moreover, /h/ was also chosen

as a consonant with no transient effects. Each pseudoword was recorded three times with different placement of stress — either on the first, second or third syllable. We decided to use pseudowords in order to obtain the results as clear as possible. The effects we were looking for should not be distorted by any extraneous variables.

The recordings were automatically labelled (using the Prague Labeller software, see Pollák — Volín — Skarnitzl, 2007) and the boundaries of the phones were manually corrected in the open-source software Praat (Boersma — Weenink, 2012).

A total number of 180 pseudowords and 540 tokens of individual vowels was obtained, 135 vowels per speaker. Six of the tokens (all from the same speaker) had to be discarded due to background noise or non-standard quality of the vowel. Thus, for further analyses, 534 vowels were used.

A spectral slice was extracted from the middle third of each vowel in the Praat program. For measurements of the spectral slope, we used functions available in this software. The investigated methods are listed in Table 1.

No.	Name	Interpretation	Formula	Praat function
1	Band energy difference 1	difference between the sums of energies in the given frequency bands	$10 \times \log_{10}$ *hbenergy* — $10 \times \log_{10}$ *lbenergy*	Spectrum > Get energy difference
2	Band density difference 1	difference between the averages of energies in the given frequency bands	$10 \times \log_{10}$ *hbdensity* — $10 \times \log_{10}$ *lbdensity*	Spectrum > Get density difference
3	Band energy difference 2	difference between the sums of energies in the given frequency bands	$10 \times \log_{10}$ (*lbenergy* — *hbenergy*)	see text
4	Band density difference 2	difference between the averages of energies in the given frequency bands	$10 \times \log_{10}$ (*lbdensity* — *hbdensity*)	see text
5	Band energy ratio	ratio of the sums of energies in the given frequency bands	$10 \times \log_{10}$ *lbenergy* / $10 \times \log_{10}$ *hbenergy*	see text
6	Band density ratio	ratio of the averages of energies in the given frequency bands	$10 \times \log_{10}$ *lbdensity* / $10 \times \log_{10}$ *hbdensity*	see text
7	Skewness	the shape of the spectrum above and below its centre of gravity	see text	Spectrum > Get skewness

TABLE 1: Overview of the methods used to calculate spectral slope. *Hbenergy* and *lbenergy* refers to high band energy and low band energy, respectively. The same applies for *hbdensity* and *lbdensity*.

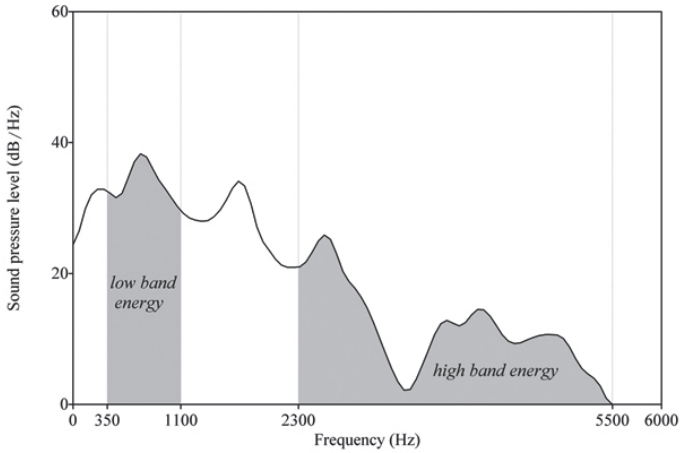

FIGURE 1: Spectrum of a vowel [ɛ] with highlighted energy in measured high and low frequency bands.

The first six methods are based on measuring energies in given frequency bands — the bands being the same across all methods: the low band spans the frequencies from 350 to 1100 Hz and the high band from 2300 to 5500 Hz. It is the best frequency band setting from those examined by Volín and Zimmermann (2011). Note that F0 and, more importantly, F2 range (or at least a great part of it) are excluded in this setting — we hypothesize that F0, being very variable within one speaker, would with its energy obscure the spectral measurements; F2 on the other hand marks the linguistic category of the vowel and therefore might contribute less to speaker identity. For more information about average formant ranges of Czech speakers the reader should refer to Skarnitzl and Volín (2012).

Figure 1 provides an illustration of a vowel spectrum with the measured frequency bands highlighted.

The first six methods are arranged in pairs. The first of the pair takes as input the sum of energy in the given band, the second uses the average energy (called density in this context) in the same band. Even though energy and density are dependent on each other, their interpretation is not the same and they might behave differently when using different formulae. One of our goals is to see whether it is more useful to employ the sum rather than average of spectral energies, or vice versa.

The seventh method computes the skewness of the spectrum, which is a measure of asymmetricity of the spectral shape. It is quantified as the difference between the shape of the spectrum below and above its centre of gravity (a frequency which divides the spectrum into two parts so that the energy in the lower half is equal to the energy in the upper half). An illustration is provided in Figure 2.

Methods 3 to 6 were not computed directly with a predefined function from the spectrum, but with the help of a Praat script — the formulae are given in the fourth column of Table 1. Band energies and densities are calculated with the function *Get band energy/density*.

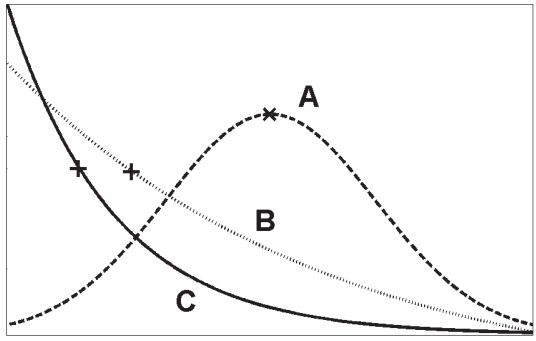

FIGURE 2: Three functions with different skewness. The points mark the centre of gravity of each function. Function A is symmetric, therefore its skewness equals 0. Functions B and C are skewed to the right, which results in positive values of skewness. Function C is more asymmetric than B, therefore its skewness will be higher.

The formula for computing skewness will not be stated here for the sake of simplicity, but it is the standard formula for calculating the third central moment of a spectrum.

For assessing the statistical significance of the results, one-way ANOVA was used with the values of spectral slope as a dependent variable and SPEAKER as factor. For evaluating individual inter-speaker differences, Tukey HSD post-hoc test was used. To see whether identity of the vowel influenced the results, we also used two-way ANOVA with SPEAKER and VOWEL as factors.

RESULTS

All of the methods resulted in statistically significant ANOVA differences (at the level of $\alpha = 0.05$), in both of the tests for SPEAKER and SPEAKER*VOWEL. This means that each of the method was able to differentiate at least one speaker from the others. The magnitude of the test criterion F was considerable for all main effects ($30 < F < 85$), thus the effect size was not calculated.

We found out that the results for energy vs. density measures were very similar, they differed only marginally and yielded very similar p values. The advantage we found was that when using methods 3 and 4, it is theoretically possible for the difference between high band and low band energies to be negative — it is then impossible to calculate the logarithm. In our data, this happened only in two cases and only when measuring energy; density seems to be more robust. For the sake of simplicity, only graphs with density measures will be presented further.

Note that throughout the paper we keep the convention for steeper spectral slope values to be represented lower in the graphs, which in some cases required reversion of the y-scale.

METHODS 1 + 2: BAND ENERGY/DENSITY DIFFERENCE 1 (BED1/BDD1)

Figure 3 shows values of band density difference 1 (the standard function from Praat) for each of the four speakers (one-way ANOVA results: $F(3, 530) = 30.85$, $p < 0.001$). Higher negative values of BDD1 mean bigger difference between low band and high band energies and therefore a steeper spectral slope. M1 has, using these methods, therefore the steepest spectral slope in his vowels.

BED1/BDD1 was able to discriminate only speaker M1, which was confirmed by Tukey HSD post-hoc test ($p < 0.001$ for all differences between M1 and the others). Speakers M2, M3 and M4 were not statistically different from each other ($p > 0.05$).

METHODS 3 + 4: BAND ENERGY/DENSITY DIFFERENCE 2 (BED2/BDD2)

Figure 4 shows values of the fourth method, the modified band density difference (one-way ANOVA results: $F(3, 530) = 84.04$, $p < 0.001$). Note that the scale of the graph is reversed due to a difference in calculation (high band and low band values are interchanged). As in Figure 3, the values lower in the graph (but in this case higher values) mean a steeper spectral slope.

The drawback of methods 3 and 4, as mentioned above, is the possibility for the difference between bands to be negative (the case of a flat or reversed spectral slope) — the calculation then fails because of the position of the logarithm in the formula. However, in our corpus of 534 vowels this happened only twice and only for method 3, i.e., energy. Nevertheless, this should be remembered for further research when measuring spectral slope of degraded or noisy signals.

Contrary to methods 1 and 2, we now have two speakers (M2 and M3) highly statistically different from each other as well as from the two remaining individuals (confirmed by Tukey HSD post-hoc test: $p < 0.001$). The values of M1 and M4 are very nearly the same. In conclusion, these methods clustered the four speakers into three groups, which are clearly distinguishable

METHODS 5 + 6: BAND ENERGY/DENSITY RATIO (BER/BDR)

The results of method 5 are shown in Figure 5 (one-way ANOVA results: $F(3, 530) = 31.8$, $p < 0.001$). Again, a steeper spectral slope is represented lower in the graph. Although the picture is quite similar to Figure 3, the speakers are more spread out from each other — M1 is highly significantly different from all the others (Tukey HSD post-hoc test: $p < 0.001$), M2 and M3 are also significantly distinct (Tukey HSD post-hoc test: $p = 0.008$) and M3 and M4 are marginally different (Tukey HSD post-hoc test: $p = 0.059$). The only difference this method failed to recognize was between speakers M2 and M4. This makes BDR and BDD2 successful to the same extent. Moreover, calculating the ratio eliminates the mathematical problem that was identified as the disadvantage of BDD2.

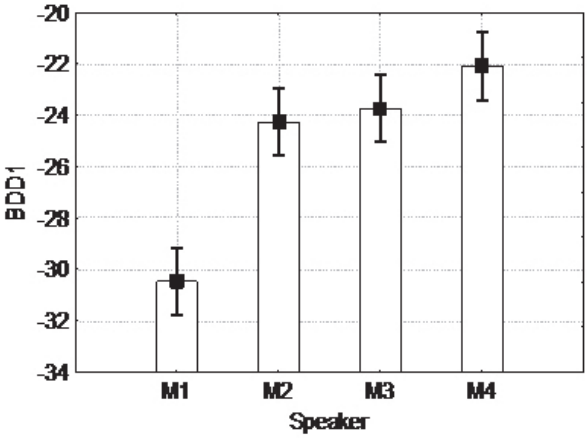

FIGURE 3: Mean values of the band density difference 1 for four speakers. Whiskers denote 0.95 confidence interval.

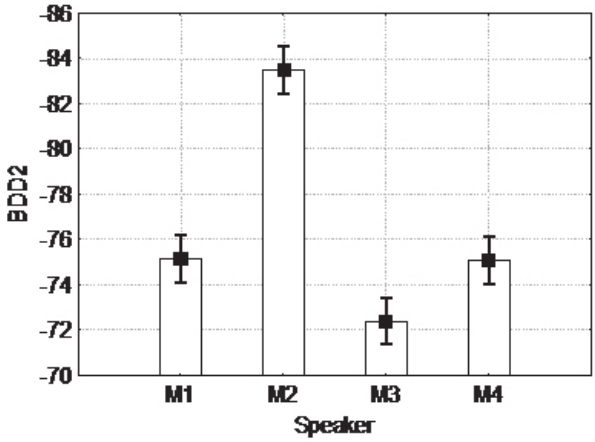

FIGURE 4: Mean values of the band density difference 2 for four speakers. Whiskers denote 0.95 confidence interval.

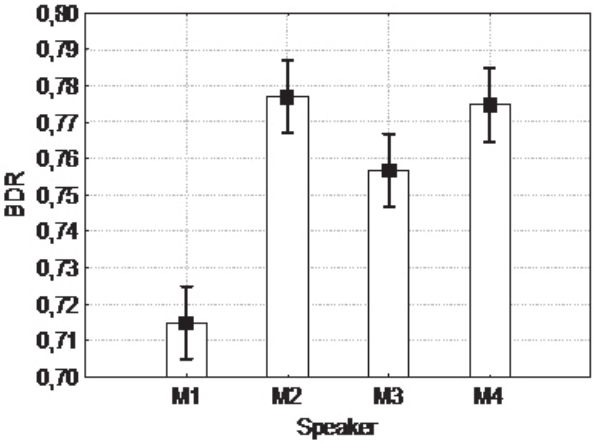

FIGURE 5: Mean values of the band density ratio for four speakers. Whiskers denote 0.95 confidence interval.

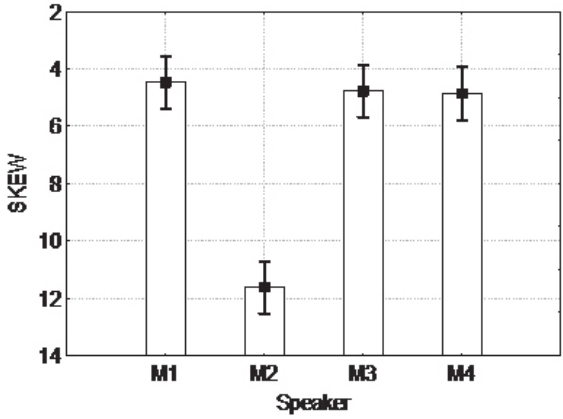

FIGURE 6: Mean values of spectrum skewness for four speakers. Whiskers denote 0.95 confidence interval.

METHOD 7: SKEWNESS (SKEW)

Figure 6 shows values of spectrum skewness for the four speakers (one-way ANOVA results: $F(3, 530) = 56.23$, $p < 0.001$). Positive skewness of a spectrum, in this case, means decline of the values towards higher frequencies — and higher values mean more asymmetry and therefore a higher spectral slope (see Figure 2).

It is obvious from the figure that this method differentiated only speaker M2 from the others (Tukey HSD post-hoc test confirmed this: $p < 0.001$), while M1, M3 and M4 were undistinguishable. This makes skewness measurement the least successful method, comparable to the predefined band energy or density differences (BED1/BDD1).

For speaker identification purposes, it is reasonable to take the two most successful methods (i.e., the two methods that were able to distinguish three groups out of four speakers) which do not produce highly correlated results and combine them, since they both seem to convey different information. Figure 7 shows a scatterplot of band density difference 2 and band density ratio. Their combination improves the identification results and it is apparent that the four speakers have been differentiated clearly from one another, which is the desired forensic outcome.

VOWEL PROFILES

If we look at the contribution of individual vowels to the overall results (Figures 8–11), it is the behaviour of [ɪ] and [u] that is most conspicuous.

In the case of BDD1 (two-way ANOVA results for SPEAKER*VOWEL interaction: $F(12, 514) = 3.72$, $p < 0.001$), all vowels behave more or less identically, but the differences in [ɪ] and [u] are most distinct — in those two vowels, all the speakers are most distinguishable from one another. An evident trend is a decline from front vowels to the back — meaning that front vowels display a flatter spectral slope value than back vowels.

Two-way ANOVA for BDD2 yielded also a significant interaction: $F(12, 514) = 2.92$, $p < 0.001$, but the profiles in Figure 9 are visibly different. Speakers M2 and M4 dis-

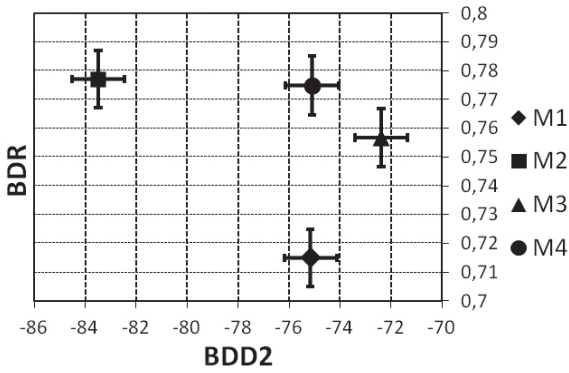

FIGURE 7: Scatterplot of BDD2 and BDR values for all speakers. Whiskers denote 0.95 confidence interval.

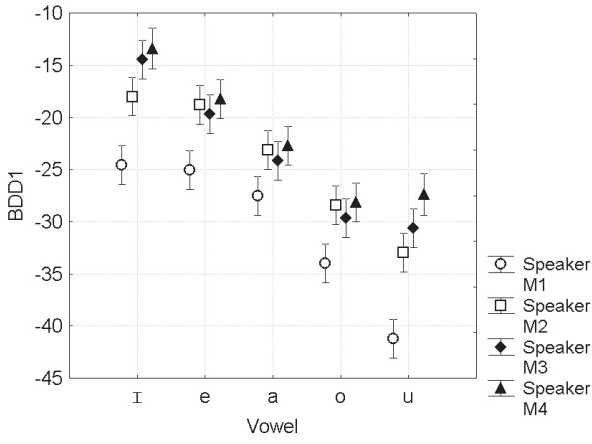

FIGURE 8: Mean values of the band density difference 1 for four speakers and five Czech vowels. Whiskers denote 0.95 confidence interval.

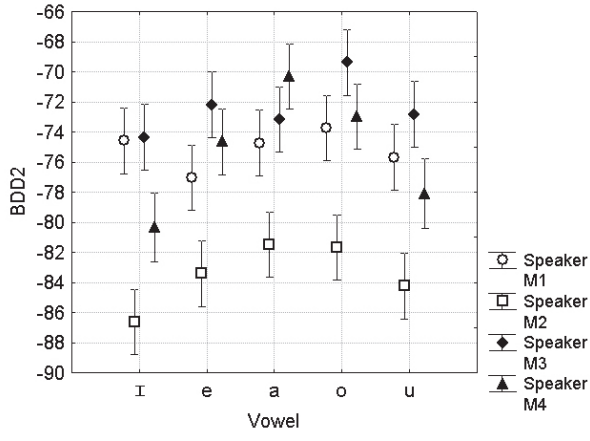

FIGURE 9: Mean values of the band density difference 2 for four speakers and five Czech vowels. Whiskers denote 0.95 confidence interval.

play a tendency for close vowels to have a flatter slope and open [a] to have steeper. The values of speakers M1 and M3 display a more complicated pattern.

Two-way ANOVA for BDR showed also a significant interaction: $F(12, 514) = 4.96$, $p < 0.001$. Figure 10 shows a situation similar to BDD1, speaker M1 has again the steepest spectral slope value than the others, which is again most prominently shown in [ɪ] and [u].

As in Figure 10, front vowels show higher values of spectral slope than back vowels — the values decrease from front [ɪ] to back [u], suggesting that even though F2 should be excluded from the measurements, some of its influence on the spectra still remains.

The values of skewness for individual vowels of the speakers are shown in Figure 11 (two-way ANOVA results: $F(12, 514) = 11.03$, $p < 0.001$). In this case, M2 is the most distinct speaker, with [ɪ] reflecting this fact most evidently. [u] again appears to show more inter-speaker differences than the remaining non-high vowels. Similarly to Figure 9, there is a tendency for the skewness to correlate with openness of the vowel, which is most prominently seen in speaker M2.

4. DISCUSSION

In this study, we have examined seven different methods of measuring spectral slope of vowels for the purposes of speaker discrimination from short-term spectra. It appears that the four methods with modified computational procedure (namely 3, 4, 5 and 6, all based on measuring energies in limited frequency bands) perform significantly better than the default methods available in the speech analysis software Praat. Measuring skewness — which employs information about the whole range of the spectrum — was also not very successful. This seems to support the notion that not all information contained in the spectra is contributive to speaker identification or differentiation. The predefined functions in Praat that compute band energy/density difference and skewness were not able to distinguish more than one speaker from the others. The methods that performed best were modified band energy/density difference (BED2/BDD2) and band energy/density ratio (BER/BDR). Since both were able to discriminate among different speakers, their combination improved the individual results.

However, it should be noted that BED2/BDD2 with this frequency setting depends much more on the energy of the low band than on the high band. The energy of the high band does not contribute much to the resulting value of the spectral slope. Although it may be useful for the purposes of speaker identification, we expect BER/BDR to perform better in other tasks, such as differentiating emotions or stressed and unstressed vowels, where higher frequencies are hypothesized to play a more significant role (Tamarit — Goudbeek — Scherer, 2008).

We also determined that when using the same frequency band setting for the first six methods, it does not matter whether we measure spectral energies or densities, the results are very similar. This question, though, needs to be addressed in the future — if it is desirable for example to broaden or narrow down the frequency bands, then densities (i.e., averages of energies in the frequency band) ensure comparability across different measurements, whereas sums of energies do not.

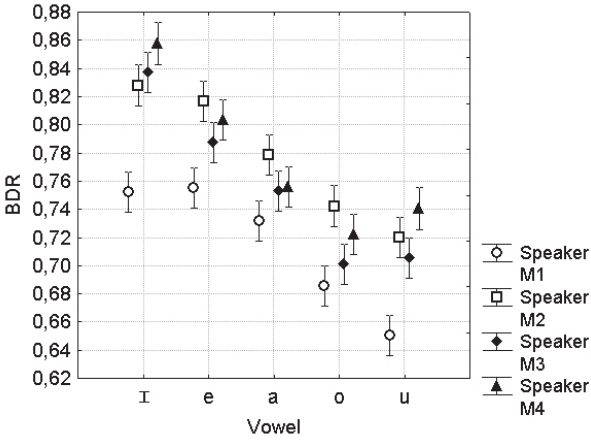

FIGURE 10: Mean values of the band density ratio for four speakers and five Czech vowels. Whiskers denote 0.95 confidence interval.

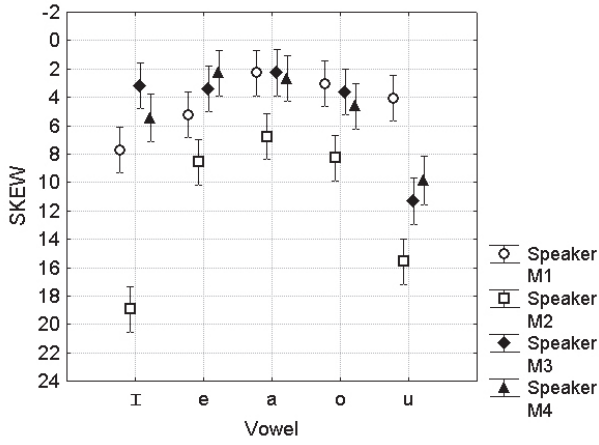

FIGURE 11: Mean values of skewness for four speakers and five Czech vowels. Whiskers denote 0.95 confidence interval.

The contribution of individual vowels to speaker recognition was also assessed. From this perspective, the methods can be divided into two groups — in BED1/BDD1, BER and BDR all vowels behaved similarly, while the close [ɪ] and [u] showed most inter-speaker differences and, most importantly, the values decrease from front vowels to back ones, which points to a significant role of the second formant. On the other hand, BED2/BDD2 and SKEW did show different vowel profiles for each speaker, not that much dependent on vowel identity, but there still was a visible trend (at least in some of the speakers) for values to correlate with the open-close distinction, which implies an effect of the first formant.

This could mean that the first group of methods attaches higher importance to the high frequency band, while the second to the low band. And even though the range of F2 is excluded from our measurements, its influence on the overall spectra is still reflected in BED1/BDD1 and BER/BDR.

The evident similarity of BDD1 and BDR is caused by similarity of the formula — BDD1 is computed as a difference of logarithms, which is in fact logarithm of a ratio. The only mathematical difference between BDD1 and BDR is the order of the operations applied. Nevertheless, both variables are able to represent different phonetic information.

For speaker identification purposes it might be beneficial to eliminate the effect of the vowel (perhaps by changing the frequency band settings for some of the vowels) or to choose only some of the vowels for analysis ([u] and [ɪ] offer themselves in this regard). The trouble is that in Czech, [u] is not particularly frequent. Considering only a part of the available vowel set could lead to substantial reduction of the material and the results might lose their significance. This question will be addressed in our further research.

Our future course of investigation will, above all, include the topic of prominence and its influence on spectral slope. We also plan to verify our findings on larger corpora with more speakers and more natural speech samples.

ACKNOWLEDGEMENTS:

The first author was supported by an internal grant of the Faculty of Arts, Charles University in Prague (VG124) 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 acknowledged. The authors would also like to thank Tomáš Bořil and František Vlasák for their kind assistance with technical and mathematical issues.

REFERENCES:

BOERSMA, Paul — WEENINK, David (2012): *Praat: Doing Phonetics by Computer*. *Version 5.2.19*. [computer program; online]. Retrieved from WWW: <http://www.praat.org/>.

BOERSMA, Paul — KOVACIC, Gordana (2006): Spectral characteristics of three styles of Croatian folk singing. *Journal of the Acoustical Society of America*, 119(3), pp. 1805–1816.

CAMPBELL, Nick — BECKMAN, Mary (1997): Stress, prominence, and spectral tilt. *ESCA Workshop on Intonation: Theory, Models and Applications*. Athens: University of Athens, pp. 67–70.

DODDINGTON, George R. (1985): Speaker recognition — identifying people by their voices. *Proc. IEEE*, 73, pp. 1651–1664.

FRENCH, Peter (1994): An overview of forensic phonetics with particular reference to speaker identification. *Forensic Linguistics*, 1(1), pp. 169–181.

FULOP, Sean A. — KARI, Ethelbert — LADEFOGED, Peter (1998): An acoustic study of the tongue root contrast in Degema vowels. *Phonetica*, 55, pp. 80–98.

HAMMARBERG, Britta — FRITZELL, Björn — GAUFFIN, Jan — SUNDBERG, Johan — WEDIN, Lage (1980): Perceptual and acoustic correlates of abnormal voice quality. *Acta Otolaryngologica*, 90, pp. 441–451.

HANSON, Helen M. (1997): Glottal characteristics of female speakers: acoustic

correlates. *Journal of the Acoustical Society of America*, 101(1), pp. 466–481.
HANSON, Helen M. — CHUANG, Erika s. (1999): Glottal characteristics of male speakers: acoustic correlates and comparison with female data. *Journal of the Acoustical Society of America*, 106(2), pp. 1064–1077.
HELDNER, Mattias (2001): Spectral emphasis as a perceptual cue to prominence. *TMH-QPSR*, 42, pp. 51–57.
HOLLIEN, Harry (2002): *Forensic Voice Identification*. London: Academic Press.
JACKSON, Michel — LADEFOGED, Peter — HUFFMAN, Marie K. — ANTOÑANZAS-BARROSO, Norma (1985): Measures of spectral tilt. *UCLA Working Papers in Phonetics*, 61, pp. 72–78.
KITZING, Peter (1986): LTAS criteria pertinent to the measurement of voice quality. *Journal of Phonetics*, 14, pp. 477–482.
KOCHANSKI, Greg — GRABE, Esther — COLEMAN, John — ROSNER, Burton (2005): Loudness predicts prominence: fundamental frequency lends little. *Journal of the Acoustical Society of America*, 118(2), pp. 1038–1054.
LAVER, John (1980): *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.
LÖFQVIST, Anders (1986): The long-time-average spectrum as a tool in voice research. *Journal of Phonetics*, 14, pp. 471–475.
MASTER, Suely — DE BIASE, Noemi — PEDROSA, Vanessa — CHIARI, Brasília Maria (2006): The long-term average spectrum in research and in the clinical practice of speech therapists. *Pro Fono*, 18(1), pp. 111–120.
MILLS, Timothy (2009): *Speech Motor Control Variables in the Production of Voicing Contrasts and Emphatic Accent* [PhD thesis]. Edinburgh: University of Edinburgh.
NOLAN, Francis (1983; reissued in 2009): *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.
NOLAN, Francis (1990): The limitations of auditory phonetic speaker recognition. In: Hannes Kniffka (ed.), *Texte zu Theorie und Praxis forensischer Linguistik*. Tübingen: Niemeyer, pp. 457–471.
POLLÁK, Petr — VOLÍN, Jan — SKARNITZL, Radek (2007): HMM-based phonetic segmentation in Praat environment. *Proceedings of the XIIth International Conference "Speech and computer — SPECOM 2007"*, pp. 537–541.
RODMAN, Robert D. — MCALLISTER, David F. — BITZER, Donald L. — CEPEDA, Luis — ABBITT, Pam (2002): Forensic speaker identification based on spectral moments. *International Journal of Speech Language and the Law*, 9(1), pp. 22–43.
SKARNITZL, Radek — VAŇKOVÁ, Jitka — WEINGARTOVÁ, Lenka (2012): Speaker discrimination using short- and long-term segmental information in vowels. *Proceedings of IAFPA 2012 (International Association for Forensic Phonetics and Acoustics 2012 Annual Conference)*. Santander: IAFPA.
SKARNITZL, Radek — VOLÍN, Jan (2012): Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické listy*, 18, pp. 7–11.
SLUIJTER, Agaath M. C. — VAN HEUVEN, Vincent J. (1996): Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, 100(4), pp. 2471–2485.
SLUIJTER, Agaath M. C. — VAN HEUVEN, Vincent J. — PACILLY, Jos J. A. (1997): Spectral balance as a cue in the perception of linguistic stress. *Journal of the Acoustical Society of America*, 101(1), pp. 503–513.
SUNDBERG, Johan — NORDENBERG, Maria (2006): 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), pp. 453–457.
TAMARIT, Lucas — GOUDBEEK, Martijn — SCHERER, Klaus R. (2008): Spectral slope measurements in emotionally expressive speech. *ISCA Tutorials and Research Workshop*. Aalborg, Denmark.
VOLÍN, Jan — ZIMMERMANN, Július (2011): Spectral slope parameters and detection of word stress. *Technical Computing Prague — Proceedings*. Praha: Humusoft, p. 125.

ABSTRAKT:
Vzestup telekomunikačních technologií v současné době umožňuje častější využití řečových nahrávek při vyšetřování trestných činů. Forenzní věda je multidisciplinární obor, který poskytuje vědeckou bázi pro posuzování důkazního materiálu během těchto vyšetřování. Forenzní fonetika se zabývá segmentálními (vokalickými a konsonantickými) a suprasegmentálními (prozodickými) řečovými rysy, které mohou odlišovat jednotlivé mluvčí. V tomto ohledu se nicméně rozšiřuje propast mezi technicky a lingvisticky orientovanými přístupy — tato studie je pokusem o její překlenutí zkoumáním českých vokálů rigorózními komputačními přístupy: pro účely rozpoznávání mluvčího ve forenzní praxi je zde porovnáno sedm metod stanovení vokalického spektrálního sklonu. Ve forenzní fonetice byl dosud spektrální sklon používán zejména při měření dlouhodobých průměrných spekter. Tato spektra se snadno získávají, avšak vykazují několik podstatných omezení. Zde jsou tedy využita krátkodobá spektra českých krátkých vokálů, jež přinášejí větší množství charakteristik specifických pro mluvčího, ale jejich extrakce je pracnější. Ze sedmi testovaných metod se softwarem předdefinované funkce ukázaly jako nevyhovující, zatímco kombinace modifikovaného rozdílu hustot pásem a poměru hustot pásem od sebe dokázala odlišit všechny mluvčí. Dále byl také prozkoumán vliv kvality vokálů na výsledky jednotlivých měření.

Lenka Weingartová | Institute of Phonetics, Faculty of Arts, Charles University in Prague
lenka.weingartova@ff.cuni.cz

Jan Volín | Institute of Phonetics, Faculty of Arts, Charles University in Prague
jan.volin@ff.cuni.cz