

## FUNDAMENTAL FREQUENCY STATISTICS FOR MALE SPEAKERS OF COMMON CZECH

RADEK SKARNITZL AND JITKA VAŇKOVÁ

### ABSTRACT

In speaker identification, a forensic phonetician's task often involves comparing voices of two or more speakers and assessing their similarity, but also their typicality. For the latter, it is necessary to have background information about the relevant speaker population. This paper introduces the database of Common Czech which was compiled as a reference database, and presents the first set of compiled statistics pertaining to fundamental frequency (F0). The population statistics are computed from a reading task and spontaneous speech. The results confirm the superiority of F0 baseline over mean or median values when assessing typicality and demonstrate, in many speakers, a narrower intonation range in spontaneous speech than in reading. The role of F0 in speaker comparison is also discussed.

**Key words:** fundamental frequency, forensic phonetics, speaker identification, Czech

### 1. Introduction

Voice comparison is probably the most frequent task within the large domain of speaker identification (Jessen, 2012; Skarnitzl, 2014). This task consists in comparing, by means of detailed acoustic and auditory analyses (Nolan, 1999), the voice in the unknown recording (originating, for instance, from an anonymous telephone call) with that in the known recording (obtained typically during a police interrogation). The determination of identity or non-identity of the voices depends on the extent to which various acoustic features and phenomena determined by listening (see, e.g., Hollien & Hollien, 1995; Hollien, 2002: 78ff. for analytic approaches to auditory speaker identification) are similar or different.

However, apart from looking for *similarities* and *differences* between the voices under comparison, it is necessary to take into consideration the *typicality* of the observed values (Jessen, 2012: 40ff.). It is clear, for example, that a difference between two voices of 10 Hz in the mean F2 frequency of the Czech vowel [u:] has a different explanatory power when

the means are 780 and 790 Hz, and when they are 1280 and 1290 Hz. While the similarity is practically meaningless in the former case, as this is very close to the mean F2 value of the Czech [u:] (Skarnitzl & Volín, 2012 report 770 Hz for young Czech male speakers), the latter situation would be considerably more interesting in forensic phonetic casework. Such a high frequency of F2 in the [u:] vowel indicates substantial – and untypical – fronting and/or de-labialization. If combined with other acoustic and auditory features, a forensic phonetician’s conclusion may be that it is much more likely for such similarities to arise when the voices under comparison originate from the same speaker than if they were to originate from different speakers.

In order to be able to make judgements about the typicality of observed similarities and differences in the measured values, these values must be related to the statistics of the relevant population. In other words, we need to know the patterning of each feature in the population: its mean or median value as a representative of the central tendency (i.e., the 770-Hz value of F2 for the [u:] vowel; see Volín, 2007a: section 3.3.1), standard deviation (SD) or other measures of the feature’s variability (*ibid*: section 3.3.2), and possibly a more detailed expression of the shape of its distribution. Only then can we determine whether the observed similarities may speak for the hypothesis of identity of the voices.

Given the importance of population data for forensic practice, it should not be surprising that recording databases of target populations and compiling population statistics of various acoustic features rank among the most important tasks of forensic phonetic research. While historically, to the best of our knowledge, some of the first reference data were produced for German (Jessen, Köster & Gfroerer, 2005 for fundamental frequency; Jessen, 2007 for articulation rate), it is currently Standard Southern British English (SSBE) where the greatest advances are being made. The DyViS database (Dynamic Variability in Speech; Nolan, McDougall, de Jong & Hudson, 2009) features recordings of 100 young male speakers in various forensically relevant speaking tasks. Population statistics have so far been presented for fundamental frequency (Hudson, de Jong, McDougall, Harrison & Nolan, 2007), and less comprehensive data also exist for disfluency features (McDougall, Duckworth & Hudson, 2015). The comparative lack of more available population data serves to show how demanding the task of producing population data is – the requirements in terms of personnel and time are tremendous, and they are far from negligible when it comes to technical and phonetic expertise.

The term “relevant” or “target” population has been used above, without going into much detail. However, it is clear that the decision as to what is a relevant population in a specific case is by no means trivial (Morrison & Ochoa, 2012). In fact, there is an inherent paradox involved in this decision (Hughes & Foulkes, 2015): we are trying to establish the population (speech community) relevant for the given speaker without knowing the speaker’s identity. Forensically applicable databases typically feature male speakers of one variety (determined regionally or socially), especially in languages with high regional variability. The question is far less straightforward when it comes to the age of the speakers. The above-mentioned DyViS database only includes speakers aged 18 to 25 years, presumably in part due to the relative accessibility of this age group at universities, but also due to the phonetic changes which SSBE is undergoing (de Jong, McDougall & Nolan, 2007). On the other hand, the German database included speakers of a rather wide age range, between 21 and 63 years.

It should be pointed out that a database of the relevant population is necessary not only in phonetic speaker identification, but also in automatic speaker identification, a factor which often seems to be overlooked in casework (Svobodová, 2014, personal communication). Automatic approaches typically exploit GMM-UBM, where the UBM stands for a universal background model; this represents the distributions of feature vectors of a general population of speakers (Reynolds, Quatieri & Dunn, 2000; Reynolds & Campbell, 2008), and it is clear that this population should correspond to some extent to the speakers under investigation.

The objective of this paper is twofold: we will report on the compilation of a new forensically applicable database of the Czech language, and we will present the first population statistics; as in most other languages, this will involve population data of fundamental frequency (F0). F0 distribution is a suitable parameter when beginning to create population statistics, because it belongs to the most commonly used parameters in forensic speaker comparison (Jessen, 2012: 67) and also thanks to the ease of its extraction.

## 2. Database of Common Czech

Until recently, there have been no forensically applicable population statistics for the Czech language. Only a modest reference set of manually measured formant values had been compiled for a small number of 27 male speakers by Skarnitzl & Volín (2012); however, this database, besides being limited in terms of the number of speakers, is also based only on a read text.

The choice of the variety for our database – Common Czech – was relatively straightforward. Common Czech is a supraregional, non-standard variety of Czech commonly used in everyday communications. It is often referred to as an interdialect which, in addition, has been affecting the standard variety (Krčmová, 2005; Chromý, 2014). Since Common Czech is used habitually for casual oral communication with family, friends or acquaintances, but it can also be encountered in more formal situations, it is the ideal variety on which population statistics of Czech should be based.

The recordings for the database were acquired during 2015. The database includes 100 male speakers aged between 19 and 50 (mean age: 25.6 years, SD: 6.7 years); this is the population which is most relevant in forensic phonetic contexts. Each person was recorded by a friend or an acquaintance to facilitate natural speech production. The recordings were obtained in quiet environments via a professional portable recorder Edirol HR-09 in a WAV format with 48-kHz sampling frequency. In order to gain a representative sample of the target speakers' speech performance, the recorded material covered several speaking styles (the stylistic differentiation follows the structure of the VASST corpus, which features predominantly older speakers from various regions of the Czech Republic). Every speaker completed six speaking tasks which lasted between 45 and 60 minutes in total: 1) a structured interview which comprised a pre-defined set of general questions (e.g., "What kind of music do you like?", "At which occasions do you listen to it?", "Do you like sports?"), as well as demographically oriented questions (e.g., "Where did you spend your childhood?", "Have you moved a lot?", "What foreign languages do you speak?")

- 2) a free, spontaneous interview taking approximately 25 minutes, in which the target speaker was encouraged to talk freely about any topic; the experimenters had a set of questions to get the speaker talking, but the speakers were free to choose any topic they felt like talking about, so that their speech was as spontaneous as possible
- 3) reading a phonetically rich text of 150 words which includes all phonemes and their context-dependent variants; the speakers were given time to familiarize themselves with the text
- 4) a picture description task lasting approximately 6 minutes
- 5) reading specific phrases and sentences
- 6) a disguise task in which the speakers were asked to disguise their voice so that it is not recognizable; they were given time to decide what kind of disguise they would use, and then they read a text which contained similar word sequences as task 3 (see also Růžičková & Skarnitzl, 2017)

As of March 2017, the recordings have been cut into the six parts and downsampled to 32 kHz. Recordings of the read tasks (numbers 3 and 6 above) have been segmented using the Prague Labeller, an HMM-based forced alignment tool (Pollák, Volín & Skarnitzl, 2007). In addition, the text of approximately 20 of the free interviews (task 2 above) has been transcribed, and it is hoped that this work will continue. While none of the tasks in the database were transmitted over a mobile phone, we plan to simulate mobile transmission by compressing the signal using the GSM AMR codec (3GPP, 2012), the most widespread codec in mobile telephones (Vaňková & Bořil, 2014).

### 3. Population statistics of fundamental frequency

Fundamental frequency (F0) is the acoustic correlate of the frequency of vocal fold vibration, and as we have mentioned above, F0 population statistics belong to those which are available most frequently for a given language. There are several ways of capturing the speaker-specific behaviour of F0 (see Rose, 2002: Chapter 8, Jessen, 2012: Chapter 3, or Skarnitzl & Hývlová, 2014 for more details and also for methodological issues related to using F0 in forensic phonetic contexts). It is beneficial to be able to determine a central value of F0 for a given speaker, one which characterizes his or her modal phonation behaviour; such a value has been called the *speaking fundamental frequency* (SFF). Researchers have most frequently used the arithmetic mean or median value to refer to SFF. More recently, the so-called F0 baseline has been proposed by Lindh & Eriksson (2007) as a speaker's neutral, carrier F0. This value, which is supposed to be most robust vis-à-vis various technical and behavioural distortions, is computed as the 7.64<sup>th</sup> percentile of a speaker's F0 dataset. All these three expressions of SFF will be provided for Common Czech, along with measures of variability.

#### 3.1 Method

Since both reading and spontaneous speech are typically used in police interrogations to obtain speech samples for comparison, F0 was examined in two of the speech styles mentioned above: in the spontaneous interview (task 2; ca. one minute was selected from

the central portion of the recording) and in ordinary reading (task 3, which typically lasted between 60 and 75 seconds). One minute of speech has been shown to be sufficient to determine a speaker's SFF (e.g., Nolan, 1983: 123; Volín, 2007b).

F0 values were extracted automatically using autocorrelation in Praat (Boersma & Weenink, 2016), with the interval of 10 msec and extraction range being 60–350 Hz; all other settings were kept at their default values. The extracted values were not corrected manually, partly due to the vast amount of data, and also, more importantly, because manual corrections are out of the question in actual forensic phonetic casework, which tends to be severely time-constrained.

The raw F0 data were processed in R (R Core Team, 2015) and visualized using the package *ggplot2* (Wickham, 2009).

### 3.2 Results and discussion

Table 1 shows the overall mean values of three SFF indicators in spontaneous male Czech speech and in reading. As we can see, the mean and median values are lower in spontaneous speech than in reading, and this difference is highly significant (*t*-test for repeated measures for mean values:  $t(99) = 11.4, p < 0.0001$ ; for median values:  $t(99) = 12.9, p < 0.0001$ ). This result is in agreement with studies of F0 in English (Hirson, French & Howard, 1995; Hollien, Hollien & de Jong, 1997); interestingly, an opposite but insignificant tendency has been observed in German (Jessen et al., 2005).

Perhaps a more interesting result, however, is the fact that the mean value of F0 baseline changes the least between the spontaneous interview and the reading task – the difference is not significant:  $t(99) = 1.5, p > 0.1$ . This lends support to Lindh and Eriksson's (2007) claim that this empirically derived indicator of F0 central tendency is more robust to changes in speaking style than the mean or median values.

**Table 1.** Indicators of speaking fundamental frequency and of F0 variability among 100 male Czech speakers in spontaneous and read speech.

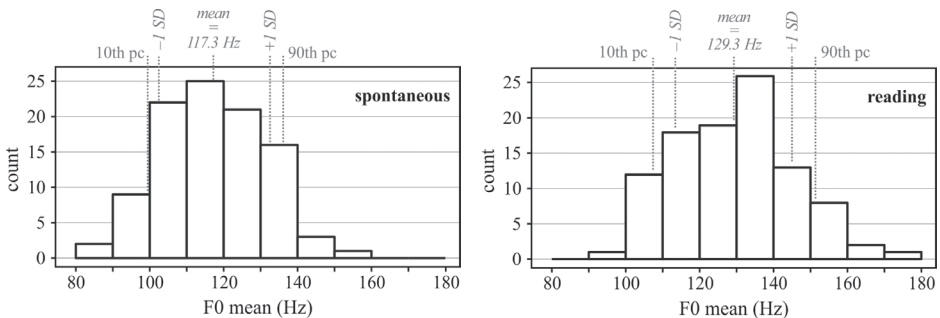
	spontaneous	reading
mean (Hz)	117.3	129.3
median (Hz)	113.1	125.8
baseline (Hz)	95.4	96.8
SD (Hz)	15.0	15.9
varco (%)	12.8	12.3

The table also displays two variability measures, standard deviation (the most frequently given measure) and coefficient of variation, or *varco*, which relates SD to the mean and is thus more straightforward (although the benefit is marginal with the mean F0 values lying close to 100 Hz). These figures for all the speakers in our corpus indicate that the overall variability does not change between spontaneous and read speech.

Next, we will consider the distribution of F0 data in more detail. In order to be able to compare our results with studies conducted on other languages, we will use the mean value, but let us repeat our conviction that the baseline has a lot of potential in F0 analy-

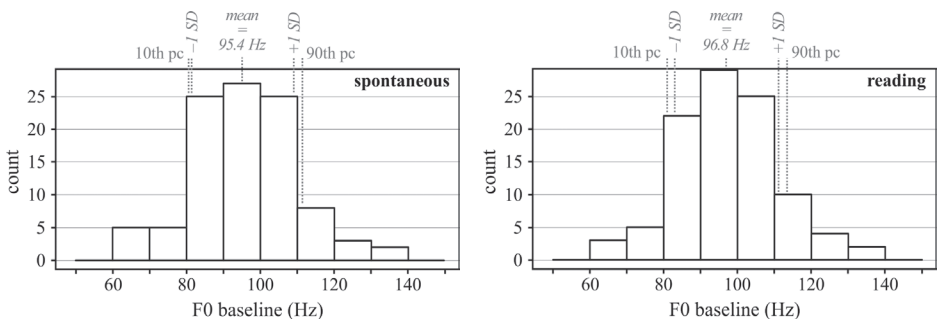
sis; that is why baseline distributions will be described as well. Figure 1 shows the distribution of the mean values of the 100 speakers in 10-Hz intervals, for both spontaneous and read speech. Figure 2 shows the histograms for baseline values. The overall means (*cf.* Table 1), as well as the  $\pm 1$  SD and 10<sup>th</sup>–90<sup>th</sup> percentile ranges from the mean are indicated in the figures.

**Figure 1.** Histograms of the F0 mean values among 100 male Czech speakers in spontaneous speech (left) and in reading (right); the overall mean and the range  $\pm 1$  SD from the mean and the 10<sup>th</sup>–90<sup>th</sup> percentile (pc) range are indicated in grey.



We can infer from the histogram of mean F0 values in spontaneous speech (top part of Fig. 1) that mean F0 is situated between 110 and 120 Hz in 25 speakers (i.e., 25 percent of all speakers) and that more than two thirds (68%) of mean values are located in the 100–130 Hz range. Similarly, we can see from Fig. 2 that F0 baseline is located within the 80–110 Hz range in more than three quarters of the speakers in spontaneous speech, as well as in reading. It is these distributions which will allow forensic practitioners to judge the relevance of the similarities or differences between analyzed voices: if SFF values of two voices fall within the same range as most speakers in the population, the finding cannot contribute to the hypothesis of identity or non-identity in any way; if, on the other hand, the SFF of one or more of the voices under comparison falls below or above the most typical 30Hz interval, the information may be relevant.

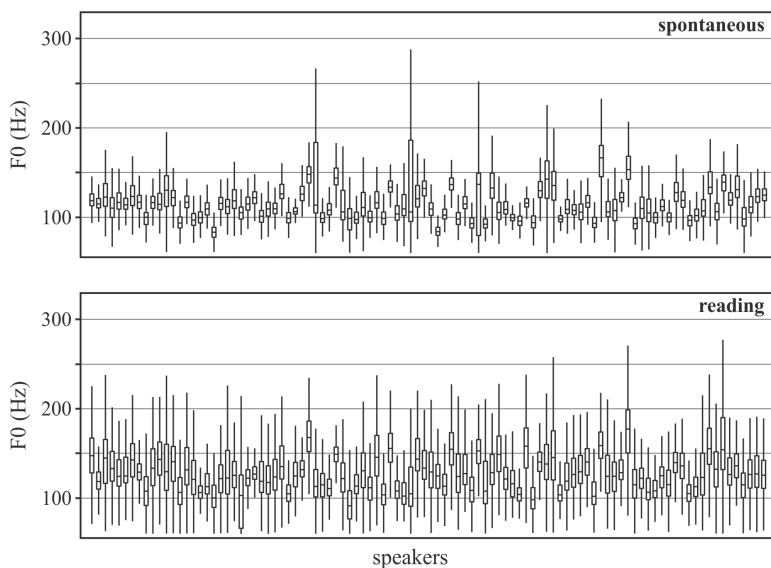
**Figure 2.** Histograms of F0 baseline values among 100 male Czech speakers in spontaneous speech (left) and in reading (right); the overall mean and the range  $\pm 1$  SD from the mean and the 10<sup>th</sup>–90<sup>th</sup> percentile (pc) range are indicated in grey.



The objective of the following analysis is to provide at least a small glimpse into the variability of individual speakers. This was partly motivated by our preliminary results based on 26 speakers from the current corpus which showed markedly lower F0 variability in spontaneous speech than in reading (Skarnitzl & Vaňková, 2015). The overall difference in the entire dataset of 100 speakers, however, is negligible, with the mean value of the coefficient of variation being 21.1% in spontaneous and 21.5% in read speech ( $t$ -test for repeated measures:  $t(99) = 0.6, p > 0.5$ ).

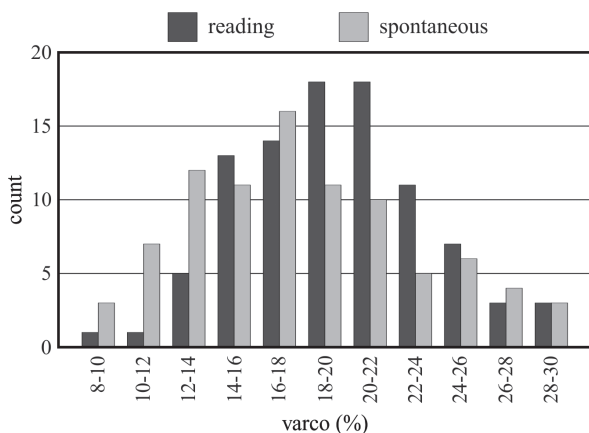
Figure 3 provides a more detailed look at the behaviour of individual speakers, and we can see that in many speakers the variability does appear to be lower in spontaneous speech. Some speakers, on the other hand, produced an extremely wide range of F0 values, so that the differences therefore seem to average themselves out. It must be noted that, first, outlier values are not plotted in Figure 3 for the sake of clarity and second, that the F0 tracking has not undergone manual correction, as noted above, and the F0 extraction error rate may be greater for some speakers than others. A cursory inspection of the recordings and F0 tracks indicates that this does seem to be the case with the highest values depicted in Figure 3.

**Figure 3.** Variability of F0 in the individual speakers in spontaneous speech (top) and in reading (bottom). The boxes indicate quartile ranges, whiskers denote  $\pm 1.5$  IQR from the quartiles; outlier values are not plotted.



We can also analyze the melodic variability by comparing the coefficients of variation in reading and in spontaneous speech, as shown in Figure 4. We are interested especially in the lower end of the scale, where we can see that, indeed, a greater number of speakers manifests low *varco* values in spontaneous speech than in reading. We may thus tentatively conclude that quite a lot of speakers in our database manifest a narrower intonation range in spontaneous speech than in reading. Flatter intonation in many speakers' spontaneous interviews is also the impression we have had when informally listening to them.

**Figure 4.** Histogram of the coefficient of variation (*varco*) values among 100 male Czech speakers in reading (dark grey) and in spontaneous speech (light grey).



#### 4. Role of F0 in voice comparison

Fundamental frequency is one of the most frequently used parameters in forensic voice comparison: according to Gold & French (2011), who surveyed 36 forensic phonetic practitioners regarding their analyses, “all respondents routinely measure fundamental frequency” (p. 301), with various indicators of SFF being extracted most commonly. There are definitive positive aspects of F0 analysis: it is not affected by the lexical content of the material so that lexical identity is not required (unlike in the analysis of vowel formants). Although F0 extraction algorithms are certainly not faultless, the median value should not be affected too much by outliers, and especially the baseline value should be most robust in scenarios when manual corrections are not viable.

On the other hand, it is well known that fundamental frequency belongs to the most variable features of a person’s voice. Our SFF is affected by a large number of factors, which can be divided into three main groups (Braun, 1995; see also Skarnitzl & Hývlová, 2014 for more details). *Technical factors* include voice disguise; indeed, our voices are very plastic when it comes to the range of F0 we are able to produce. Changing one’s SFF has been repeatedly shown to be the most popular disguise strategy in actual casework, and the same has been found in the database of Common Czech (see Růžičková & Skarnitzl, 2017 for more detail). Fortunately, it is generally easy to determine – using auditory analysis – whether a speaker is producing an utterance with his or her natural pitch or whether they are attempting to disguise their voice by shifting the SFF. Braun (1995) mentions the effects of age, smoking or alcohol intoxication as *physiological factors*, while *psychological factors* include especially various affective states or stress. Stress may obviously play a crucial role in forensic settings; in most speakers, stress has been shown to increase SFF (Kirchhübel, Howard & Stedmon, 2011; Giddens, Barron, Byrd-Craven, Clark & Winter, 2013). In addition, Braun (1995) classifies under psychological factors those which we may refer to as *situational*: vocal fatigue, time of the day, level of ambient noise (where increased F0 is caused by the so-called Lombard effect), or speech style.



In light of the many influences on a speaker's F0, it is clear that one must be very cautious when comparing SFF in different voices for forensic purposes. Researchers usually agree that F0 is useful in forensic settings, as long as the physiological and psychological factors of the recordings under investigation are comparable (Hirson et al., 1995; Boss, 1996; Jessen et al., 2005). It is believed that since SFF tends to be higher in unknown recordings, during which speakers usually experience some level of stress, the finding of an SFF value lower than in the known recording obtained during interrogation should speak for non-identity of the speakers (French, 2012, personal communication).

## 5. Conclusion

The first objective of this paper was to introduce the newly recorded database of Common Czech for forensic purposes, its current level of processing and future plans for its development. The second, and more important objective was to present the first population statistics based on this database; fundamental frequency was chosen as the first parameter.

It is interesting to note once again that, despite all the above-mentioned drawbacks of F0 stemming mostly from the tremendous plasticity of our speech production, F0 still tends to be the most frequently available statistic for a given language. In other words, if there are some population data available for a language, F0 is most probably included in them. Undoubtedly, compiling population statistics of Common Czech for other acoustic parameters typically employed in forensic casework is work which will continue in the future.

---

## ACKNOWLEDGEMENT

This research was supported by the project "Compilation of representative database of Common Czech for forensic purposes" (no. 239515) solved by the second author in 2015 at the Faculty of Arts, Charles University in Prague, financed by Charles University Grant Agency. The first author was also supported by the Charles University project Progres 4, *Language in the shiftings of time, space, and culture*.

---

## REFERENCES

- 3GPP. (2012). TS 26.071 AMR speech CODEC; General description. Retrieved from [http://www.3gpp.org/ftp/specs/archive/26\\_series/26.071/](http://www.3gpp.org/ftp/specs/archive/26_series/26.071/)
- Boersma, P. & Weenink, D. (2016). *Praat: Doing Phonetics by Computer (Version 6.0.20)*. Retrieved on September 12, 2016 from [www.praat.org](http://www.praat.org)
- Boss, D. (1996). The problem of F0 and real-life speaker identification: a case study. *Forensic Linguistics*, 3, 155–159.
- Braun, A. (1995). Fundamental frequency – How speaker-specific is it? *BEIPHOL 64, Studies in Forensic Phonetics*, 9–23.
- Chromý, J. (2014). Demokratizace spisovné češtiny a ideologie jazykové kultury po roce 1948. *Acta Universitatis Carolinae, Philologica*, 3, 71–81. [in Czech]

- De Jong, G., McDougall, K. & Nolan, F. (2007). Sound change and speaker identity: An acoustic study. In: Müller, C. (Ed.), *Speaker Classification II*, 130–141. Berlin: Springer.
- Hirson, A., French, P. & Howard, D. (1995). Speech fundamental frequency over the telephone and face-to-face: some implications for forensic phonetics. In: Windsor Lewis, J. (Ed.), *Studies in General and English Phonetics: Essays in Honour of Professor J. D. O'Connor*, 230–240. London: Routledge.
- Hollien, H. (2002). *Forensic Voice Identification*. San Diego: Academic Press.
- Hollien, H. & Hollien, P. A. (1995). Improving aural-perceptual speaker identification techniques. *BEIPHOL 64, Studies in Forensic Phonetics*, 87–97.
- Hollien, H., Hollien, P. A. & de Jong, G. (1997). Effects of three parameters on speaking fundamental frequency. *Journal of the Acoustical Society of America*, 102, 2984–2992.
- Hudson, T., de Jong, G., McDougall, K., Harrison, P. & Nolan, F. (2007). F0 statistics for 100 young male speakers of Standard Southern British English. *Proceedings of the 16<sup>th</sup> ICPHS*, 1809–1812.
- Hughes, V. & Foulkes, P. (2015). The relevant population in forensic voice comparison: Effects of varying delimitations of social class and age. *Speech Communication*, 66, 218–230.
- Jessen, M. (2007). Forensic reference data on articulation rate in German. *Science & Justice*, 47, 50–67.
- Jessen, M. (2012). *Phonetische und linguistische Prinzipien des forensischen Stimmenvergleichs*. München: Lincom.
- Jessen, M., Köster, O. & Gfroerer, S. (2005). Influence of vocal effort on average and variability of fundamental frequency. *International Journal of Speech, Language and the Law*, 12, 174–213.
- Krčmová, M. (2005). Stratifikace současné češtiny. *Linguistica Online*. Retrieved from <http://www.phil.muni.cz/linguistica/art/krcmova/krc-012.pdf> [in Czech]
- Lindh, J. & Eriksson, A. (2007). Robustness of Long Time Measures of Fundamental Frequency. *Proceedings of Interspeech 2007*, 2025–2028.
- McDougall, K., Duckworth, M. & Hudson, T. (2015). Individual and group variation in disfluency features: A cross-accent investigation. *Proceedings of the 18<sup>th</sup> ICPHS*.
- Morrison, G. S. & Ochoa, F. (2012). Database selection for forensic voice comparison. *Proceedings of Odyssey 2012: The Language and Speaker Recognition Workshop*. Singapore: ISCA.
- Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.
- Nolan, F. (1999). Speaker recognition and forensic phonetics. In: Hardcastle, W. J. & Laver, J. (Eds.), *The Handbook of Phonetic Sciences*, 744–767. Oxford: Blackwell Publishers.
- Nolan, F., McDougall, K., De Jong, G. & Hudson, T. (2009). The DyViS database: style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law*, 16, 31–57.
- Pollák, P., Volín, J. & Skarnitzl, R. (2007). HMM-Based Phonetic Segmentation in Praat Environment. *Proceedings of SPECOM 2007*, 537–541. Moscow: MSLU.
- R Core Team (2015). *R: A language and environment for statistical computing (version 3.2.2)*. R Foundation for Statistical Computing, Vienna. Retrieved from <http://www.R-project.org/>.
- Reynolds, D. A. & Campbell, W. M. (2008). Text-independent speaker recognition. In: Benesty, J., Sondhi, M. M. & Huang, Y. (Eds.), *Springer Handbook of Speech Processing*, 763–781. Berlin: Springer-Verlag.
- Reynolds, D. A., Quatieri, T. F. & Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing*, 10, 19–41.
- Rose, P. (2002). *Forensic Speaker Identification*. London: Taylor & Francis.
- Růžičková, A. & Skarnitzl, R. (2017). Voice disguise strategies in Czech male speakers. *AUC Philologica 3, Phonetica Pragensia*, pp. 19–34.
- Skarnitzl, R. (2014). Forezní fonetika. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvího*, pp. 11–20. Praha: Faculty of Arts, Charles University in Prague. [in Czech]
- Skarnitzl, R. & Hývlová, D. (2014). Statistický popis hodnot základní frekvence. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvího*, pp. 49–64. Praha: Faculty of Arts, Charles University in Prague. [in Czech]
- Skarnitzl, R. & Vaňková, J. (2015). Presenting the population statistics of Common Czech: preliminary F0 results. Presented at IAFPA 2015, Leiden.
- Skarnitzl, R. & Volín, J. (2012). Referenční hodnoty vokálních formantů pro mladé dospělé mluvčí standardní češtiny. *Akustické listy*, 18, 7–11. [in Czech]

- Vaňková, J. & Bořil, T. (2014). Telefonní přenos. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvího*, pp. 104–115. Praha: Faculty of Arts, Charles University in Prague. [in Czech]
- Volín, J. (2007a). *Statistické metody ve fonetickém výzkumu*. Praha: Epocha. [in Czech]
- Volín, J. (2007b). Data volume requirements for reliable F0 normalization. In: Vích, R. (Ed.), *17<sup>th</sup> Czech-German Workshop – Speech Processing*, 62–67. Praha: Czech Academy of Sciences.
- Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.

---

## RESUMÉ

Identifikace mluvího je jedním z nejčastějších úloh forenzního fonetika. Při identifikaci mluvího často porovnáme nahrávky hlasů dvou nebo více mluvích a cílem srovnání je posoudit jejich podobnost, ale také typičnost. Je zřejmé, že posuzování typičnosti hlasů a rozdílů mezi nimi je třeba mít k dispozici informace o populaci mluvích. Pořizování databází a kompilace statistik pro použití ve forenzně fonetické praxi proto patří k důležitým aspektům současného výzkumu. Článek popisuje databázi obecné češtiny, která byla pořizena právě s cílem, aby sloužila jako referenční databáze pro forenzní účely, a zároveň představuje první sadu populačních statistik týkajících se základní frekvence (F0). Prezentované údaje jsou založené na čtené i spontánní řeči. Výsledky potvrzují, že tzv. základní hladina F0 navržená Lindhem a Erikssonem (2007) vystihuje idiosynkratičnost F0 lépe než aritmetický průměr či medián. Naznačují také, že velká část mluvích ve spontánním projevu hovoří monotónněji, tedy vykazuje nižší intonační rozpětí. Článek také rozebírá roli základní frekvence ve srovnávání mluvího.