2017 ACTA UNIVERSITATIS CAROLINAE PAG. 19–34
PHILOLOGICA 3 / PHONETICA PRAGENSIA

VOICE DISGUISE STRATEGIES IN CZECH MALE SPEAKERS

ALŽBĚTA RŮŽIČKOVÁ AND RADEK SKARNITZL

ABSTRACT

Voice comparison in forensic phonetic casework requires the assessment of the similarity of the target voices. This task may be impeded by the speakers' attempt to disguise their voice. The objective of this study is to map the strategies of voice disguise as employed by 100 native speakers of Common Czech. In agreement with findings in other languages, changes in speaking fundamental frequency appeared in most speakers, and phonatory modifications were the second most frequent type of disguise. Other strategies included modifications of the resonance characteristics of the voice, melodic and temporal changes. Recognition difficulty and naturalness of the disguise were also assessed. In addition, several acoustic parameters in the natural and disguised condition were compared in 15 speakers whose disguise rendered recognition more difficult.

Key words: voice, voice disguise, forensic phonetics, speaker identification, Czech

1. Introduction

The most typical task of a forensic phonetician concerns the recording of a perpetrator of a criminal act – often obtained from mobile telephone interception – which has to be compared with a recording of a detained suspect. In this procedure, which is called voice comparison, the aim is to arrive at a statement of probability with which the two recordings originate from the same speaker. If the perpetrator is aware of the option of their communication being monitored, he or she may attempt to disguise their voice, which can render voice comparison difficult or even impossible.

Current statistics of the occurrence of voice disguise are, unfortunately, not available. Data from Germany indicate that, overall, such cases are generally not very frequent: in the years 1988–1995, attempts at voice disguise appeared in less than 5 per cent of all cases investigated at Trier University (Masthoff, 1996). However, the frequency of occurrence was reported as considerably higher in cases of abductions, extortions, sexual harassment, or hoax calls (Künzel, 2000), with disguise attempts identified as frequently as

https://doi.org/10.14712/24646830.2017.30

in 69% of blackmail cases (Masthoff, 1996). It may be expected that, with the increase in mobile phone communication, the proportion of voice disguise will continue to increase. Indeed, in a more recent study, Braun (2006) examined 175 forensic cases and reported the occurrence of disguise in 22.9% of them, with the number highest in harassment and extortion cases. The possibility of voice disguise is therefore something which must be kept in mind at all times when comparing voices in forensic phonetic practice.

This paper is interested in the strategies which speakers employ when attempting to disguise their voice. One type of voice disguise which will not be addressed here in more detail is electronic disguise, in which the speaker makes use of a specialized device; this type of disguise used to be quite infrequent (less than 1% of all cases according to Masthoff, 1996) but seems to be on the rise (Clark & Foulkes, 2007; see also Perrot et al., 2007). The majority of voice disguise cases thus still involves speakers exploiting the natural potential for variability in their speech production (Skarnitzl, 2016) and changing some characteristics of their speech.

Before discussing the individual strategies which speakers use to disguise their voice, let us point out the results of Masthoff's study, in which speakers were instructed to disguise their voice in any way they could think of. Masthoff shows that speakers typically chose to change only one or two parameters in their speech. A single parameter was altered by 55% of the subjects. This corresponds to the experience of forensic phoneticians in the Czech Republic (Svobodová & Voříšek, 2014), who also report relatively primitive voice disguise attempts. As Masthoff (1996) states, the fact that the speakers do not use more complex strategies may be caused by the speaker's need to formulate utterances simultaneously with attempting voice disguise, which requires a relatively high cognitive effort.

In the following sections, the most frequent disguise strategies reported in literature will be described; where available, we will also discuss the effect of the particular strategies on recognition of the speaker's voice by listeners.

1.1 Fundamental frequency changes

The most frequent voice disguise strategy has been repeatedly shown to be changes of fundamental frequency (F0). When discussing the overall level of a speaker's fundamental frequency, we talk about speaking fundamental frequency (SFF). Modifications of SFF are advantageous for the speaker because they are easy to perform, they do not cause incomprehensibility of the utterance (and therefore do not restrain the transfer of information), and they are sufficiently effective for disguising the speaker's identity (Künzel, 2000).

According to Künzel's (2000) experiment, speakers with a higher natural SFF tend to raise it even more, while speakers with a lower natural SFF tend to lower it, sometimes shifting to creaky voice. In his study, in which speakers were asked to speak with a higher and lower SFF, females made less significant changes whereas males in general modified SFF more prominently and they shifted to falsetto more frequently (see section 1.2 for more information on creaky voice and falsetto).

Extreme cases of F0 modification also lead to changes of vowel formants; although F0 and formants are often said to be independent of each other, extreme F0 values are

associated with changes in the vertical position of the larynx. A raised larynx due to high F0 results in a shorter vocal tract and hence higher resonance frequencies (formants), and vice versa.

Künzel (2000) also notes that F0 modifications lead to the decrease of articulation rate and to the increase in the occurrence of pauses and their duration, probably as a result of the speaker's concentration on maintaining the disguise strategy.

The effect of SFF changes on the listeners' ability to identify speakers was studied by one of the first researchers in this field of study, the psychologist Frances McGehee (1937; cited in Eriksson, 2010). She experimentally tested the accuracy of voice line-ups when F0 had been modified and found that recognition dropped from 80% (when listeners heard the speakers' natural voice) to 63% (when they first heard the speaker's disguised voice). Other studies focused on voice identification when the change in SFF was a by-product of a different type of phonation; these will be discussed in the following section.

1.2 Phonatory modifications

The most frequent modifications to modal phonation which have been reported in forensic casework, creaky voice and falsetto, also include a change in F0. Creaky voice is a result of a complex laryngeal adjustment which involves tighter adduction of the arytenoid cartilages and low-frequency and low-amplitude vocal fold vibration only in the front portion of the glottis. The effectiveness of creaky voice disguise was examined by Hirson and Duckworth (1995) who compared the performance of expert and lay listeners. They showed that while speaker identification was high when the speakers used modal phonation, 99% for phoneticians and 93% for non-phoneticians, performance dropped significantly when the speakers used creaky voice: to 73% for phoneticians and to 51% for non-phoneticians. The results therefore suggest that phonetically trained listeners perform better in voice line-ups, even when the speakers disguise their voice.

The second phonatory modification, falsetto, involves a non-modal pattern of vocal fold vibration which is characterized by high longitudinal tension and high-frequency vibration only in the front portion of the glottis. The effectiveness of falsetto as a voice disguise strategy was studied by Wagner and Köster (1999), who presented recordings of familiar and unfamiliar speakers to listeners; the speakers used both modal phonation and falsetto. The results show how detrimental falsetto was to recognition: while the listeners were able to identify 97% of the familiar speakers in the undisguised condition, only 4% of the same speakers were correctly recognized when they used falsetto. The authors highlight the importance of using phonetically trained experts for forensic phonetic analyses.

Another type of non-modal phonation, called pressed or tense voice, involves a high compression of the vocal folds in the medial portions of the vocal folds; this voice is sometimes accompanied by the vibration of the false vocal folds, by the so-called ventricular phonation. To our knowledge, pressed voice has not been examined from the perspective of its effectiveness as a voice disguise strategy; its use in forensic settings has, however, been reported (Künzel, 2000).

Finally, speakers disguising their voice may decide to eliminate phonation entirely and whisper. It is to be expected that whispered speech lacks some crucial speaker-specific

information, and speaker identification will thus be compromised. The effect of whisper on voice line-ups was investigated by Orchard and Yarmey (1995) who found that line-ups were most successful when the listeners had heard the speaker's natural speech in both speech samples, less accurate when the listeners had heard the speaker's whisper on both samples, and least accurate when the listeners first heard the speaker's whisper and then were asked to identify the speaker based on his normal speech. Speaker identification in normal and whispered speech has also been examined by Bartle and Dellwo (2015) who, similarly, report a drop in correct identifications in whispered speech, and more so for naïve listeners than for forensic phoneticians.

1.3 Resonance modifications

Apart from voice disguise strategies which exploit changes in the frequency or manner of vocal fold vibration, speakers can modify the resonance characteristics of the speech signal in several ways. First, we can mention changes to the supralaryngeal long-term settings (Laver, 1980; Skarnitzl, 2016). These include palatalization, labialization or pharyngealization (in other words, quasi-permanent shifts away from the neutral position of the tongue), as well as hyper- and hyponasality (habitual opening of the velopharyngeal port and speaking "through the nose" in the former case, and pinching one's nose so as to prevent nasal airflow in the latter). To the best of our knowledge, the effectiveness of adopting these long-term setting modifications for voice disguise purposes has not been studied.

The second type of resonance changes involves the speaker inserting a foreign object into his or her vocal tract or holding it in front of their mouth (Künzel, 2000). Figueiredo and Britto (1996) investigated the effect on the voice of holding a pen between the front teeth: the articulation of speech sounds is restricted due to the fixed position of the jaw, spreading of the lips and retraction of the tongue. Various speech segments are, understandably, affected in different degrees; it is therefore not possible to make a general prediction as to the acoustic effect of a foreign object in the oral cavity. The authors point out that any disguise strategy which modifies the speaker's supralaryngeal characteristics impedes identification more than phonatory modifications, because the latter changes preserve most dialectal and segmental features in speech. The reason for the low predictability of supralaryngeal changes consists mainly in the complex interaction between the restriction caused by the object in the speaker's mouth on the one hand and the speaker's natural articulation manners on the other hand (Figueiredo & Britto, 1996). In more primitive ways of voice disguise, whose impact on the resonance characteristics is relatively low, speakers often use a handkerchief placed between their mouth and the microphone (Svobodová & Voříšek, 2014).

1.4 Dialect or foreign accent imitation

A dialect can be so strong a component of identity that it can obscure, or override other features of the speaker's voice (Eriksson, 2010). It is therefore not surprising that imitating another dialect belongs to frequently encountered strategies in forensic phonetic casework. The key question is, naturally, the authenticity of the imitation. Markham (1999) studied eight speakers of Southern Swedish, each imitating three regional accents

of Swedish, and found that some of the speakers were able to achieve a consistently authentic impression and conceal their own identity.

Another frequent strategy of changing one's pronunciation concerns the imitation of a foreign accent (Masthoff, 1996), even in the Czech environment (Svobodová & Voříšek, 2014). It is obvious that as with dialect imitation, authenticity must be maintained for disguise to be successful, and speakers are not always able to achieve that (Neuhauser, 2008). Neuhauser and Simpson (2007) investigated the ability of native German listeners to identify authentic and imitated French and American English accents in German. Their results show, surprisingly, that German listeners were better able to identify imitated accents by German speakers than authentic non-native accents (uttered by native speakers of the respective languages); the authors hypothesize that the German imitators made use of stereotypical pronunciation patterns which were, in turn, picked up by the listeners.

1.5 Voice disguise strategies and this study

When we compare the overview studies which have been quoted in the previous sections (Masthoff, 1996; Künzel, 2000; Braun, 2006), the emerging picture of voice disguise strategies tends to be quite similar. The majority of speakers who attempt at disguise modify one, maximum two parameters in their speech. Voice disguise mostly involves a change in the phonatory behaviour, i.e., in some characteristics of the voice: according to Braun (2006), this concerned 65% of all disguise attempts. Most frequently, speaking fundamental frequency is raised (possibly switching into falsetto), sometimes also lowered (possibly switching into creaky phonation); the former seems to be preferred by male speakers, the latter by females. Pressed voice and whisper are also relatively frequent. Other strategies include denasalization (achieved by pinching one's nose), the imitation of a foreign accent, or speaking with a handkerchief in front of the mouth. The same strategies have been witnessed in the Czech forensic environment (Svobodová & Voříšek, 2014).

Investigations of voice disguise also clearly indicate that speakers differ in their ability to effectively and consistently (without "gaps" in the disguise) conceal their identity (Masthoff, 1996; Neuhauser, 2008; Vyhnálková, 2013), just as listeners differ in their ability to identify speakers who manipulate their voices (Vyhnálková, 2013).

This paper follows up on Vyhnálková's pilot study and examines voice disguise strategies in a large corpus of 100 male speakers of Common Czech (Krčmová, 2005; Chromý, 2014), a nonstandard but supraregional dialect of the Czech language used in everyday communication by many speakers not only in informal but even in slightly formal situations.

2. Auditory mapping of voice disguise strategies

2.1 Material

The mapping of voice disguise strategies was conducted on the Database of Common Czech, which contains recordings of 100 male speakers aged between 19 and 50 (mean age: 25.6 years) and represents a reference database used for forensic purposes; the database only features male speakers precisely because most forensic material is pro-

duced by males. The recordings were acquired in a quiet environment, using a portable recorder Edirol R09, with 48-kHz sampling frequency. The speakers performed several speaking tasks which involved several speech styles (see Skarnitzl & Vaňková, 2017 for more information on the corpus); this study is based on the analysis of two reading tasks.

In the first task, the subjects were asked to read in their natural voice a phonetically rich text of 150 words, lasting approximately one minute. For the second, voice disguise task, the speakers were instructed to imagine that they were criminals who had to report to their boss, while suspecting the call might be monitored. They were therefore told to change their voice so they would not be identifiable based on the recording. They were given sufficient time to devise a strategy to disguise their voice. The two texts differed but contained some identical phrases.

2.2 Procedure

The first objective was to perform auditory analyses of both types of recordings – natural and disguised – in order to map the range of voice disguise strategies used by the speakers; this step was performed by the first author. The careful, repeated auditory comparison of the natural and disguised recordings yielded a list of disguise strategies corresponding to the following categories:

- speaking fundamental frequency (pitch level) modifications (higher, lower, no audible change)
- phonatory modifications (modal, creaky, pressed, breathy phonation, whisper)
- speech rate (faster, slower, no audible change)
- any other notable strategies (nasalization and other resonance changes, etc.)

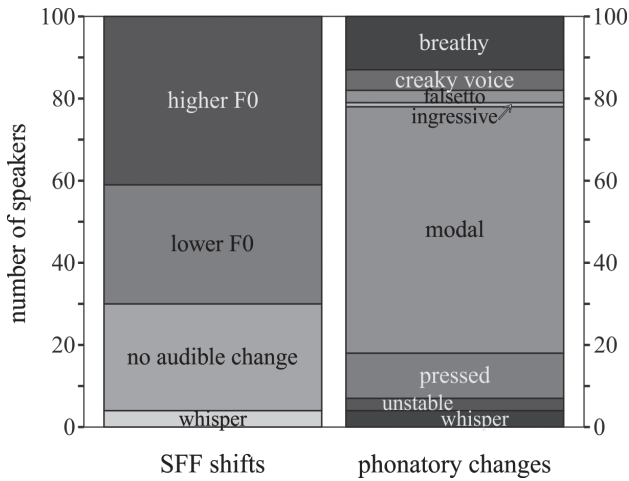
Apart from noting the disguise strategy, we subjectively evaluated another two criteria: the difficulty of recognition of the speaker based on his natural and disguised voice (easy, medium, difficult), and the naturalness of the speaker's disguised voice (natural, unnatural). The second criterion was thus concerned whether it was easily noticeable that the speaker was disguising their voice or not. For instance, changes in articulation sounded natural in most cases, and a hearer might therefore not identify such speech as intentionally modified, whereas prominent SFF changes would be regarded as an obvious disguise attempt. The natural and disguised production were compared within the same speaker, because the objective of this step was to separate speakers who chose rather simple and ineffective ways of disguise from those who used more complex strategies and whose voice disguise strategies were suitable for a more thorough analysis.

2.3 Results

As expected, based on the literature reviewed above, the majority of speakers tended to shift their speaking fundamental frequency (see the left panel of Figure 1). In most cases, SFF was audibly higher (41 of the 100 speakers), with 3 of them even shifting to falsetto. 29 speakers lowered their SFF, and 5 of them shifted into creaky voice. 26 speakers' pitch

level remained without any audible change. The remaining 4 speakers disguised their voice using whisper, so that F0 was eliminated altogether. Of the speakers who used SFF shift as voice disguise, 17 only changed this parameter (6 speakers raised and 11 lowered their SFF).

Figure 1. Disguise strategies in the form of speaking fundamental frequency (SFF) shifts (on the left) and phonatory modifications (right).



As can be seen in the right panel of Figure 1, the manner of phonation was modified only by a minority of the speakers: 58 of them adhered to their modal phonation while disguising their voice. Of the remaining 42 speakers, 13 employed breathy phonation and 13 pressed phonation, 5 spoke in a creaky voice, 3 shifted into falsetto, and 4 speakers used whisper. 1 speaker employed ingressive phonation. In 3 speakers, the manner of phonation was unstable: they used more types within their disguised recording. Interestingly, a phonatory modification was the single changed parameter only in 4 speakers: 2 of them employed pressed phonation, 1 speaker used ingressive phonation, and 1 speaker's phonation was unstable (falsetto, creaky voice, as well as modal phonation appeared in his disguised speech). Some of the speakers' disguise strategies involved more phonatory modifications, for example a combination of pressed and creaky voice; these more sophisticated techniques will be mentioned separately in section 4.

Speech rate was audibly modified even less frequently than phonation: it remained unchanged in the disguised condition of 86 of the 100 speakers. Lower speech rate appeared in 11 speakers, only 3 speakers decided to speak faster. None of the speakers employed a change in speech rate as the only modification.

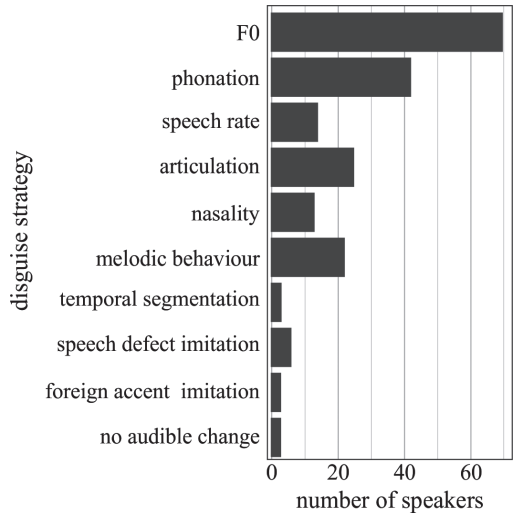
39 speakers used some kind of modification of the resonance characteristics of their speech. Quite popular among these were changes in nasality: 11 speakers used hypernasality as a long-term setting, while 2 speakers pinched their nose and sounded de-nasalized. 4 speakers modified their long-term lingual setting to pharyngealization. In 10 cases, we observed changes in the articulation of vowels: anteriorization in 3 cases, posteriorization in 2 cases, higher articulation in 3 cases and lower articulation in 2 cases.

6 speakers imitated one or more speech defects, and 3 speakers imitated a foreign accent (dialect imitation did not occur in our sample, though).

25 speakers modified their prosodic behaviour: 13 of these changed the realization of conclusive falling tones, while 9 spoke in a monotonous voice. Another 3 speakers used a lot of pauses and hesitation markers in their speech.

Finally, it remains to be pointed out that only 3 of the 100 speakers failed to perform any audible changes in their voice disguise. The strategies selected by the speakers in our study are summarized in Figure 2; note that as some speakers disguised their voice in more ways at the same time, the total number of disguise strategies exceeds the number of speakers.

Figure 2. Summary of disguise strategies, as identified by means of careful listening.



Next, we will turn to the two additional criteria that we rated, recognition difficulty and naturalness (see above in section 2.2). The ratings are summarized in Table 1. In terms of the difficulty of recognition, it was usually obvious that the given speaker's disguised voice was produced by the same individual as his undisguised voice: 53 speakers' disguise was thus rated as easy to recognize. In 30 of these speakers, the disguise did not sound natural, while the other 23 ones achieved naturally sounding disguised speech.

For 32 speakers, we rated the difficulty of recognition as medium; the match between their disguised and undisguised voices was not entirely straightforward at first "sight", but careful listening would probably lead to a successful recognition without any serious troubles. Within this group of speakers, 22 of them sounded unnatural in their disguised condition, and only 10 speakers' disguised voices were rated as sounding natural.

High difficulty of matching the disguised speech to the natural speech of the same speaker was assigned in 15 cases; 7 of the speakers' disguise sounded unnatural and 8 natural.

As is summarized in the last column of Table 1, 59 speakers' disguised voices sounded intentionally modified (i.e., the speakers did not sound natural), while 41 speakers' disguise gave the impression of a more or less natural (not intentionally modified) speech.

Table 1. Summary of the naturalness and recognition difficulty of disguise strategies, as identified by means of careful listening.

	recognition difficulty			
	low	medium	high	total
natural disguise	23	10	8	41
unnatural disguise	30	22	7	59
total	53	32	15	100

3. Acoustic analyses

3.1 Method

Based on the outcomes of the auditory analyses presented in section 2.3, we selected 15 speakers whose recognition was judged as difficult. Subsequently, these 15 speakers' natural and disguised voices were analyzed acoustically.

Segment boundaries of the 30 target recordings (natural and disguised for 15 speakers) were aligned automatically using Prague Labeller (Pollák et al., 2007) and corrected manually. Acoustic analyses were then performed on the entire recordings, as well as on individual sounds, as described below.

As for speaking fundamental frequency, we extracted raw F0 values automatically every 10 msec using autocorrelation in Praat (Boersma & Weenink, 2015); as some of the speakers used falsetto or a notably high SFF, the extraction range was set to 60–450 Hz. Speaking fundamental frequency was then quantified in several ways (*cf.* Skarnitzl & Vaňková, 2017): we calculated the mean and median value of F0, as well as the more robust F0 baseline (Lindh & Eriksson, 2007; see also Skarnitzl & Hývlová, 2014).

The following analyses regarding phonatory modifications were conducted on sounds resampled to 16 kHz. First of all, harmonicity (harmonics-to-noise ratio, HNR), which is related to the degree of noise in the spectrum (Yumoto, Gould & Baer, 1982), was extracted for each vowel using the default settings in Praat (Boersma, 1993). It was necessary to deal with those tokens where Praat yielded an "undefined" value; it did not seem advantageous to exclude such tokens since the absence of periodicity reflected the disguise strategy employed by some speakers. With HNR, the lowest detected value was –3.6 dB; undefined values were thus replaced with –10 dB (*cf.* Skarnitzl, 2011: 223). The other two parameters expressing some aspects of voice quality were jitter and shimmer as measures of voicing irregularity; these were also extracted from Praat, using the default values.

The extent of (at least some) articulatory modifications was assessed by means of comparing long term formant distributions (LTFs; see Nolan & Grigoras, 2005). Formant values (F1–F3) were extracted in the 0–5 kHz range every 10 msec from vowel and non-nasal sonorant segments, using the Burg algorithm implemented in Praat.

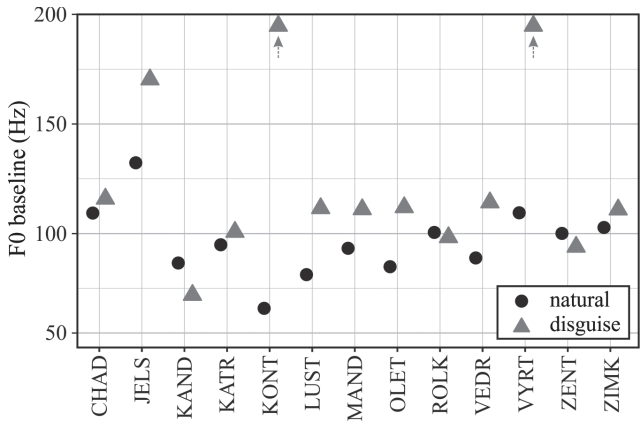
The last parameter we wanted to objectify was speech rate. We therefore measured for each speaker's masked natural and disguised reading, in syllables per second, speech rate and articulation rate (the latter being speech rate stripped of pauses, hesitations etc.).

Statistical analyses were performed in R (R Core Team, 2015) using the package *effects* (Fox, 2003) and visualized using the package *ggplot2* (Wickham, 2009).

3.2 Results

It is obvious from the results of the auditory assessment presented in section 2.3 that speaking fundamental frequency (SFF) belonged to the most frequently modified parameters. Figure 3 confirms the perceptual data: in many of the 15 speakers we can see a clear change in SFF, although SFF is expressed as F0 baseline, which is relatively more robust to behavioural changes. It is clear, however, that extreme pitch shifts will translate into baseline shifts as well. The differences in SFF between the natural and disguised condition would have been considerably more pronounced if SFF were depicted using the median value. Note that the comparison for speakers KUCK and MAKN is missing, as they employed whisper to disguise their voice, and any F0 values detected in the disguise condition were spurious.

Figure 3. Comparison of speaking fundamental frequency (expressed as the baseline) in natural and disguised speech.



Next, let us turn to the parameters pertaining to voice quality. Figure 4 shows a comparison of mean harmonicity in the natural and disguise condition for each speaker. Harmonicity seems to be sensitive at least to some audible changes in phonation: the speakers who changed their voice are indicated in italics in the figure, and greater differences in HNR can be observed in most of these speakers. It is worth commenting on speaker KATR, whose mean harmonicity is considerably higher in the disguise mode but was not marked as having modified his voice quality. What he did change was the resonance characteristics of his voice by means of pharyngealization and lowering his larynx, with the result of his voice sounding more resonant; hence higher harmonicity. Some of the other speakers whose disguised speech yielded higher HNR values, KONT and VYRT, spoke in falsetto (*cf.* Fig. 3), which also makes the speech wave more regular and thus represents a relative increase in tonal components. On the other hand, the lowest HNR values in the disguise condition (KUCK, MAKN, ROLK, ZENT) are the result of breathy voice, whisper, or a voice which combines creaky and breathy quality.

Figure 4. Comparison of mean harmonicity in natural and disguised speech; speakers who were identified as having modified their phonation behaviour are indicated in italics.

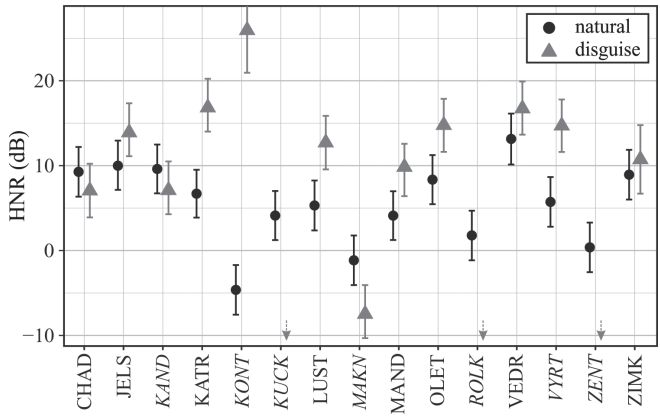
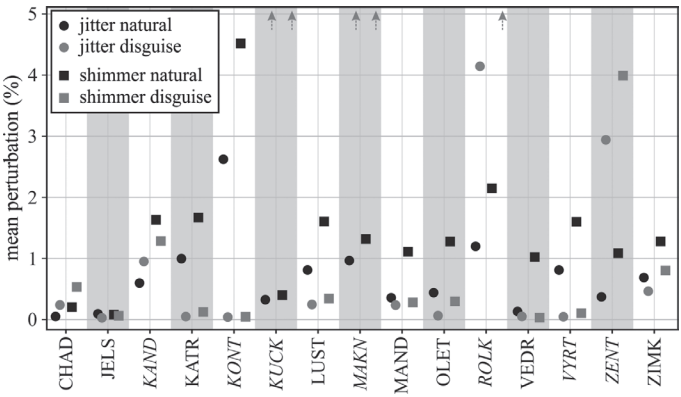


Figure 5 shows results for the F0 perturbation measures, jitter and shimmer. It must be kept in mind that the values are quite low since they were extracted from vowels in natural speech, which are quite short. As with HNR above, the speakers indicated in italics have audibly modified their voice quality. Again, in most of these speakers, the disguise conditions (in gray) are associated with considerably higher values than natural conditions (in black).

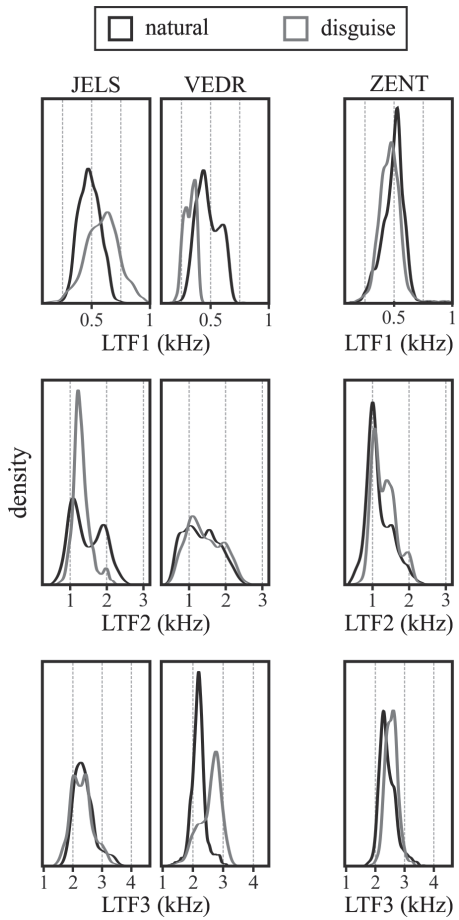
Figure 5. Comparison of mean jitter and shimmer in natural and disguised speech; speakers who were identified as having modified their phonation behaviour are indicated in italics. The arrows at the top of some bars mean a higher mean value due to the replacement of undefined values (see section 3.1).



Let us now move from characteristics pertaining to the voice source to changes in supraglottal resonances. The auditory mapping presented in section 2.3 revealed that several speakers did make changes to their articulation. We therefore compared long-term formant distributions in these speakers whose articulatory characteristics sounded different in the disguise condition with LTFs of those where such a difference was not

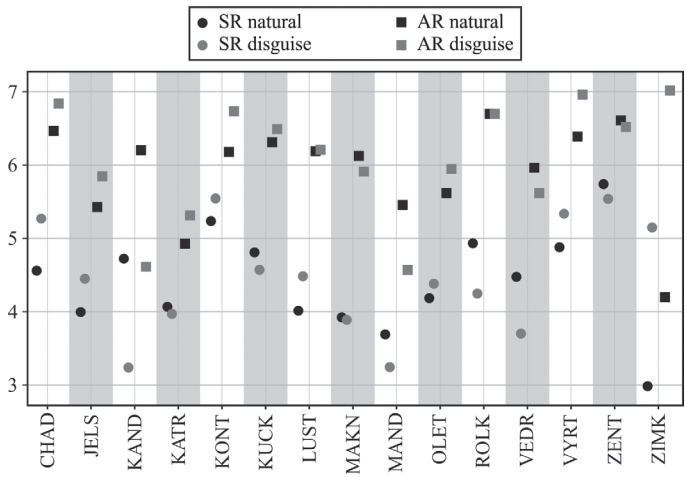
identified. Partial results of this comparison are shown in Figure 6, with distributions of F1–F3 for two speakers from the former and one speaker from the latter group. It can be seen that, indeed, the formant distributions in natural and disguised speech are much more similar – especially in terms of the location of the primary peak – in the speaker on the right, where no articulatory modifications were detected, than in the two speakers on the left, whose disguise strategy included articulatory shifts.

Figure 6. Comparison of LTF1–3 in natural and disguised speech for two speakers whose articulation was audibly different (on the left) and for one speaker whose articulation did not sound changed (on the right).



The final analysis concerns speech rate (SR) and articulation rate (AR). In the impressionistic assessment of speech rate changes in section 2.3, no salient changes have been noted. Figure 7 confirms that changes exceeding 1 syllable per second are quite rare (speakers KAND and ZIMK in both SR and AR, with the former speaking slower in the disguise condition while the latter speaking faster).

Figure 7. Speech rate (SR) and articulation rate (AR) in natural and disguised speech, in syll/s.



4. Interesting cases of voice disguise

As was mentioned in section 3.2, there were 3 out of the 100 speakers who did not perform any modifications under the disguise condition. Another 31 speakers performed only one change to their speech. In this section, we are going to survey several speakers from the other end of the scale, who demonstrated a greater degree of imagination when choosing the strategies for voice disguise. All of them were quite difficult to recognize when compared with their natural speech production. Sound examples are provided on the accompanying webpage, http://fonetika.ff.cuni.cz/vyzkum/materialy/maskovani/.

First, let us mention speaker ROLK, who combined a strongly breathy voice quality (*cf.* Fig. 4) with imitating a foreign accent; we can assume that he was aiming at a Russian accent. This was manifested by a closer (higher) articulation of vowels, by quite subtle shifts of consonantal articulation (especially the postalveolar fricatives and the fricative trill ř), and by changes in word stress and in rhythmical patterning in general. Along with high consistency, this all led to a naturally sounding disguise which would render speaker identification rather difficult.

Another successful imitation strategy was employed by speaker ZIMK, who decided to imitate the current Czech president Zeman. The strategies included melodic and rhythmic changes, with individual words or short groups of words being chopped off, sometimes even using preglottalization (e.g., [ʔʒɛ ˈʔr̥iːɦovo͡u ˈʔsɛm]; see Skarnitzl & Machač, 2010), as well as segmental changes such as shortening of vowels or more close articulation.

Speaker OLET managed to disguise his voice very effectively by modifying a single parameter: the vertical position of the larynx. His laryngeal lowering still sounds relatively natural, but its degree is suggested by changes in vowel formants: for instance, in the word *vlastně*, the F1 in [ɛ] dropped by 0.67 ERB and the F2 by over 2 ERB.

Modifications of articulation were the dominant disguise strategy of speaker VEDR who, apart from raising SFF, imitated several speech defects. The lateral alveolar approximant [l] was pronounced with no alveolar contact, only with slight raising of the tongue tip, and therefore vocalized. The alveolar trill [r] was in most cases realized in a very similar way, but sometimes also as a uvular fricative. The fricative trill was usually pronounced as a fricative, without the initial trill. Apart from these articulatory changes, the speaker also raised his SFF and spoke with very prominent lengthening of phrase-final syllables.

The last speaker to be mentioned here is LUST, who also exploited a range of changes to disguise his voice. Apart from a slightly higher SFF (*cf.* Fig. 3), he also imitated a speech defect, specifically more anterior (dentalized) articulation of alveolar fricatives and a shift in the Czech fricative trill ř. On top of that, he also employed melodic and rhythmic changes.

5. Discussion

The aim of this study was to examine the strategies of voice disguise employed by 100 native speakers of Czech. Not surprisingly, changes in speaking fundamental frequency appeared in 70% of the speakers. Phonatory modifications as a group were the second most frequent type of disguise, with creaky and pressed voice each appearing in 13 speakers. For each speaker, we also noted the naturalness of his speech in the disguise mode, as well as how difficult it was to recognize the speaker in his disguise. Apart from mapping disguise strategies in the entire database, we performed acoustic analyses in 15 speakers who managed to change their voice in such a way that recognition would be difficult (section 3) and introduced the most interesting disguise strategies of six selected speakers.

Based on the results of this study, it does not seem that we could easily identify one strategy or a combination of more strategies which are most effective when one attempts to disguise their voice. In section 2.3, we mentioned that there were 8 speakers whose disguise sounded natural and was difficult to assign to their natural speech. No clear pattern emerges even if we examine only these speakers: 6 of them did change their SFF but 2 did not; 7 of them performed some kind of articulatory modification (nasalization, closed jaw setting, speech defects); phonation was changed in only 2 of these speakers.

Voice disguise strategies represent a very interesting research topic, not only due to its practical implications in forensic phonetic casework, but also due to the way in which it illustrates the astounding plasticity of the human speech production mechanism. It is clear that investigations of voice disguise will continue in the future, as will further surveys of the Common Czech database for other forensically relevant properties of speech.

ACKNOWLEDGEMENTS

The main author was supported by the Internal Grants of the Faculty of Arts programme, registration number FF/VG/2017/83. The second author was supported by the Charles University project Progres 4, *Language in the shiftings of time, space, and culture*.

REFERENCES

Bartle, A. & Dellwo, V. (2015). Auditory speaker discrimination by forensic phoneticians and naive listeners in voiced and whispered speech. *The International Journal of Speech, Language and the Law*, 22, pp. 229–248.

Boersma, P. (1993). Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound. *IFA Proceedings*, 17, pp. 97–110.

Boersma, P. & Weenink, D. (2015). *Praat: doing phonetics by computer (Version 6.0)*. Retrieved from http://www.praat.org

Braun, A. (2006). Stimmverstellung und Stimmenimitation in der forensischen Sprechererkennung. In: Kopfermann, T. (Ed.), *Das Phänomen Stimme: Imitation und Identität*, pp. 177–181. Röhrig: St. Ingbert.

Chromý, J. (2014). Demokratizace spisovné češtiny a ideologie jazykové kultury po roce 1948 [The democratization of Standard Czech and the ideology of language culture after 1948]. *Acta Universitatis Carolinae – Philologica*, 3, pp. 71–81.

Clark, J. & Foulkes, P. (2007). Identification of voices in electronically disguised speech. *The International Journal of Speech, Language and the Law*, 14, pp. 195–221.

Eriksson, A. (2010). The Disguised Voice: Imitating Accents or Speech Styles and Impersonating Individuals. In: Llamas, C. & Watt, D. (Eds.), *Language and Identities*, pp. 86–96. Edinburgh: Edinburgh University Press.

Figueiredo, R. M. & Britto, H. S. (1996). A report on the acoustic effects of one type of disguise. *Forensic Linguistics*, 3, pp. 168–175.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8(15), pp. 1–27.

Hirson, A. & Duckworth, M. (1995). Forensic implications of vocal creak as voice disguise. In: *BEIPHOL 64*, *Studies in Forensic Phonetics*, pp. 67–76.

Krčmová, M. (2005). Stratifikace současné češtiny [Stratification of contemporary Czech]. *Linguistica Online*. Retrieved from http://www.phil.muni.cz/linguistica/art/krcmova/krc-012.pdf

Künzel, H. J. (2000). Effects of voice disguise on speaking fundamental frequency. *Forensic Linguistics*, 7, pp. 149–179.

Laver, J. (1980). *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Lindh, J. & Eriksson, A. (2007). Robustness of long time measures of fundamental frequency. In: *Proceedings of Interspeech 2007*, pp. 2025–2028.

Masthoff, H. (1996). A report on a voice disguise experiment. *Forensic Linguistics*, 3, pp. 160–167.

Neuhauser, S. (2008). Voice disguise using a foreign accent: phonetic and linguistic variation. *The International Journal of Speech, Language and the Law*, 15, pp. 131–159.

Neuhauser, S. & Simpson, A. P. (2007). Imitated or authentic? Listeners' judgements of foreign accents. In: *Proceedings of 16th ICPhS*, pp. 1805–1808.

Nolan, F. & Grigoras, C. (2005). A case for formant analysis in forensic speaker identification. *International Journal of Speech, Language and the Law*, 12, pp. 143–173.

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, pp. 31–57.

Orchard, T. L. & Yarmey, A. D. (1995). The Effects of Whispers, Voice-Sample Duration, and Voice Distinctiveness on Criminal Speaker Identification. *Applied Cognitive Psychology*, 9, pp. 249–260.

Perrot, P., Aversano, G. & Chollet, G. (2007). Voice disguise and automatic detection: review and perspectives. In: Stylianou, Y., Faundez-Zanny, M. & Esposito, A. (Eds.), *Workshop on Nonlinear Speech Processing 2005, LNCS 4391*, pp. 101–117. Berlin: Springer Verlag.

Pollák, P., Volín, J. & Skarnitzl, R. (2007). HMM-based phonetic segmentation in Praat environment. In: *Proceedings of the XIIth International Conference "Speech and computer – SPECOM 2007"*, pp. 537–541.

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

Skarnitzl, R. (2011). *Znělostní kontrast nejen v češtině* [*Voicing Contrast Not Only in Czech*]. Praha: Epocha.
Skarnitzl, R. (2016). Co dokáže náš hlas? Fonetický pohled na variabilitu řečové produkce [What is our voice capable of? Phonetic perspective on the variability of speech production]. *Slovo a smysl*, 26, pp. 95–113.
Skarnitzl, R. & Hývlová, D. (2014). Statistický popis hodnot základní frekvence [Statistical description of fundamental frequency values]. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvčího* [*Phonetic Speaker Identification*], pp. 49–64. Praha: Faculty of Arts, Charles University in Prague.
Skarnitzl, R. & Machač, P. (2010). Domain-initial coordination of phonation and articulation in Czech radio speech. *Acta Universitatis Carolinae – Philologica*, 1/2009, pp. 21–35.
Skarnitzl, R. & Vaňková, J. (2017). Fundamental frequency statistics for male speakers of Common Czech. *Acta Universitatis Carolinae – Philologica*, 3, pp. 7–17.
Svobodová, M. & Voříšek, L. (2014). Identifikace mluvčích z pohledu autentické kriminalistické praxe v České republice [Speaker identification from the perspective authentic criminological practice in the Czech Republic]. In: Skarnitzl, R. (Ed.), *Fonetická identifikace mluvčího* [*Phonetic Speaker Identification*], pp. 136–144. Praha: Faculty of Arts, Charles University.
Vyhnálková, L. (2013). *Vliv vzdělání na schopnost maskovat svůj hlas* [*The Effect of Education on the Ability to Disguise One's Voice*]. Unpublished diploma thesis. Prague: Institute of Phonetics, Faculty of Arts, Charles University.
Wagner, I. & Köster, O. (1999). Perceptual recognition of familiar voices using falsetto as a type of voice disguise. In: *Proceedings of 14th ICPhS*, pp. 1381–1384.
Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag.
Yumoto, E., Gould, W. J. & Baer, T. (1982). Harmonics-to-Noise Ratio as an index of the degree of hoarseness. *Journal of the Acoustical Society of America*, 71, pp. 1544–1549.

RESUMÉ

Srovnávání mluvčích ve forenzně fonetické praxi spočívá ve vyhodnocování podobnosti hlasů v cílových nahrávkách. Tato úloha může být ztížena – a někdy dokonce i znemožněna – snahou mluvčích měnit svůj hlas a změnit tak svou identitu. Cílem této studie je za prvé zmapovat strategie, které mluvčí při maskování hlasu používají; vycházíme přitom z databáze obecné češtiny, která obsahuje 100 mužských mluvčích a která byla sestavena jako referenční databáze pro forenzní účely. V souladu se zjištěními z jiných jazyků byly nejčastěji využívány změny střední hlasové frekvence, tedy celkové polohy hlasu, druhým nejčastějším způsobem maskování byly fonační modifikace, tedy změny kvality hlasu. K dalším strategiím patří změny rezonančních charakteristik hlasu, melodické nebo temporální změny. Součástí výzkumu bylo i posouzení přirozenosti maskování a náročnosti rozpoznání mluvčího při srovnání původní a maskované nahrávky. U 15 mluvčích, u nichž byla náročnost rozpoznání vyhodnocena jako poměrně vysoká, byla provedena i akustická analýza několika parametrů.