2014 ACTA UNIVERSITATIS CAROLINAE PAG. 43–54
PHILOLOGICA 1 / PHONETICA PRAGENSIA XIII

THE EFFICIENCY OF DIFFERENT FORMANT PARAMETERS FOR SPEAKER DISCRIMINATION

JITKA VAŇKOVÁ

ABSTRACT

This study examines the efficiency of different ways of capturing vowel formants for speaker discrimination. It compares the speaker-discriminating potential of static (F1–F4) and dynamic formant values (F1–F3), and assesses the usefulness of long-term formant distribution (LTF of F1–F4) for discriminating between 16 female speakers of Czech. The results show that dynamic parameters overall perform slightly better than static ones; the most useful parameter of all is static F4. The study found no systematic differences in discriminability of speakers with regards to the position of word stress, i.e. speaker-specific information can be present in stressed as well as unstressed syllables. LTF seems to be a promising complement to the segment-based methods as it provides an overall picture of the behaviour of each formant. The distribution of all four formants (especially F4) has been shown to have some speaker-discriminating potential, which has been assessed both visually and statistically.

Key words: vowel formants, static and dynamic values, long-term formant distribution, speaker identification, forensic phonetics

1. Introduction

Voice undoubtedly conveys some information about a speaker, and speaker identification is thus a process we all do on an everyday basis. In terms of professional approaches towards speaker identification, current best practice consists in a combination of auditory and acoustic analysis (see e.g. Nolan and Grigoras, 2005; Jessen, 2008). Auditory analysis is essential for assessing the linguistic phonetic material – it enables identification of linguistically relevant data and indicates what is comparable. Acoustic analysis then serves to refine and quantify the linguistic phonetic analysis and, more importantly, to uncover details to which the ear is insensitive or for which there are no adequate auditory analysis frameworks (Nolan, 1994; Nolan and Grigoras, 2005; Jessen, 2008).

A lot of research has been dedicated to searching for acoustic parameters which provide some speaker-specific information. It has led to four main areas: segmental information (both vocalic and consonantal), melodic parameters, temporal structuring of speech and phonatory modifications. Views differ on the relative importance of various parameters in real forensic cases, but it appears that supralaryngeal cues to speaker identity are more stable than laryngeal ones. Vowel formants – the resonant frequencies of vowels – have been recognized as crucial for speaker identification by all comprehensive accounts of this area (e.g. Hollien, 2002; Rose, 2002; Jessen, 2008), and their usefulness in a specific forensic case was demonstrated by Nolan and Grigoras (2005). Formants are relatively easy to extract from the material using freely available software. They provide information about the speaker resulting from the interaction of an individual vocal tract, idiosyncratic articulatory gestures which are needed to achieve the linguistically determined targets in that vocal tract, as well as the speaker's acquired sociophonetic behaviour – a combination which is highly regarded in forensic phonetics.

The history of speaker identification research has seen three ways of exploiting vowel formants. In the traditional approach, a static value of each formant was used, represented by the mean value in the central, stable part of a vowel (Nolan and Grigoras, 2005; de Jong et al., 2007; Marrero et al., 2008; Duckworth, 2011). Typically, values of the first three formants have been considered; F4, though known to be the most dependent on speaker identity, is rarely present or detectable in recordings obtained in forensic contexts due to lowpass filtering or strong background noise.

While static formant values provide some information about speaker's anatomy of the vocal tract (Stevens, 1971), there is increasing evidence that formant trajectories – the dynamic, time-varying properties of formants within a vowel or also across several sonorant sounds – could be even more useful for discriminating between speakers (Goldstein, 1976; Ingram et al., 1996; McDougall, 2004) as they reflect, in addition, the movement of the individual's speech organs. If we think of speech as a series of linguistically determined targets (the centres of segments) linked by transitions, it can be argued that while those targets are highly constrained by the language system, the transitions offer greater scope for individual variation, and reflect an individual's articulatory solutions to achieve these linguistically agreed targets. Dynamic formant values are thus a product of the interaction of an individual's vocal tract with idiosyncratic articulatory gestures (McDougall and Nolan, 2007). The usefulness of dynamic formant values for speaker discrimination remains, however, to a large extent unexplored as previous studies focused mainly on trajectories of a single long vowel or a diphthong (McDougall, 2004, 2006; McDougall and Nolan, 2007).

In contrast to these segment-based methods, Nolan and Grigoras (2005) introduced a more global approach to representing formant frequencies, namely the long-term formant distribution (LTF). LTF reflects the long-term disposition of formants by providing an overview of all values for each formant, thus providing a clear picture of its behaviour. Compared to the segment-based approaches, it has several advantages. First, it is less time-consuming as it does not require vowel categorization. Instead, all vowels are used for analysis. In fact, other sounds with formant structure – sonorants, but also hesitation sounds – may also be used (see Moos, 2012), which may be of crucial importance in forensic casework in which speech material is often sparse. Second, the distribution of a formant reflects not only the dimensions of the speaker's vocal tract but also articulatory habits like palatalization or lip rounding. Lastly, the shape of the distribution can provide some information about the speaker's vowel space (Nolan and Grigoras, 2005), apart from the positioning of the distribution along the frequency axis. The LTF fails, however, to show between-speaker differences in individual vowels and to reflect dynamic aspects of speech; that is why it is recommended to combine the global LTF with a more "local" analysis of vowel formants, as described above. Yet, LTF seems a powerful tool for forensic-phonetic purposes and its effectiveness in a forensic case has been shown by Nolan and Grigoras (2005). Recent studies (see Jessen and Becker, 2010; Moos, 2012) lend further support to the claim that LTF provides speaker-specific cues.

The aim of this study is to compare the three methods of capturing formant values – i.e., the static values (F1–F4), the dynamic changes in formant trajectories (F1–F3) and the long-term formant distribution (F1–F4) – for discriminating between Czech speakers. Formants have not been analysed from the speaker-specific perspective in Czech (see Skarnitzl, 2012 for only a preliminary analysis). More importantly, however, we want to examine the speaker-discriminating potential of formant trajectories in short monophthongs; formant dynamics have been investigated only in inherently changing sounds like diphthongs or vowel–sonorant sequences. Another objective is also to assess the usefulness of statistical methods when comparing long-term formant distributions.

2. Method

2.1 Material

The speech material for this study was taken from a subset of the Prague Phonetic Corpus (Skarnitzl, 2010), in which students of linguistic programs, aged 20 to 25, were instructed to "act out" a series of short read dialogues after sufficient preparation. The motivation for acting the dialogues out was to add some degree of spontaneity to the performance while preserving textual identity at the same time – it should be pointed out that vowel formants, especially their dynamic characteristics, have been investigated on considerably controlled speech material. The recordings were obtained in the sound-treated recording studio of the Institute of Phonetics in Prague at 32-kHz sampling frequency and 16-bit resolution. For the purpose of this study, we analysed recordings of 16 female students. The recordings were automatically segmented using the Prague Labeller (Pollák et al., 2007) and the boundaries of the target vowels were then adjusted manually (Machač and Skarnitzl, 2009).

Vowel formants were extracted from 75 vowels for each speaker, i.e. 15 items (the same for all subjects) of each of the five Czech short vowels /ɪ ɛ a o u/. Several criteria were observed when choosing the final set. First, only vowels in autosemantic words were taken into consideration because synsemantic words are more likely to undergo reductions (Johnson, 2004). Second, the segmental context was examined: vowels followed by a palatal or liquid consonant were disregarded since these consonants are known to exert a considerable influence on vowel formant frequencies. Lastly, the 15 items of each vowel quality were balanced for the position of word stress, i.e. five items of every vowel quality were selected each from stressed, post-stressed and unstressed syllables. The reason for distinguishing post-stressed syllables was that they were reported to exhibit specific behaviour (Palková and Volín, 2003) in terms of their prosodic characteristics:

they seem to have higher intensity, F0 and sometimes to be of a more peripheral vowel quality than stressed syllables.

2.2 Analyses

The static values of F1–F4 were measured in seven equidistant points in the middle third of each vowel using the Burg method implemented in Praat (Boersma and Weenink, 2010). Each token was then represented by the mean value from these seven measurements. The default settings for female speakers were used for the extraction of F1–F4. In cases where no F4 value was detected in the default range of 0–4.4 kHz, the upper frequency was raised up to 4.8 kHz. 20 per cent of the highest and lowest values of the automatically extracted values were checked manually and, if necessary, corrected by means of direct estimation from the spectrogram. Most errors involved a nasal formant being erroneously identified as F2 or a formant being "skipped" and a higher formant being identified instead. The set of formant values obtained in this way also served as the basis for the comparison of long-term formant distributions (LTF).

Formant trajectories – the dynamic values – of F1–F3 were captured by measuring formant frequencies in four equidistant points within the whole duration of a vowel. The automatically extracted values were again checked and manually corrected if necessary.

Out of the total number of 1,200 items (15 representations of each of the 5 Czech short vowels for 16 speakers), 23 items were discarded because F4 and/or F3 could not be identified automatically or visually from the spectrogram. In total, our analyses are therefore based on formant values from 1,177 vowels.

To assess the usefulness of the static and dynamic formant values for speaker discrimination, we used linear discriminant analysis (LDA). As enough data – following the recommendations in Volín (2007) – was used for the results of LDA to be reliable, the tokens were not partitioned into training and test sets. The discrimination task was a closed-set one, i.e. the identity of a speaker had to be assigned to one of the fixed set of known speakers. In the case of LTF, the distributions were compared visually and the median signalled to capture the central tendency (cf. McDougall, 2012). Though previous studies rely solely on visual comparisons (Nolan and Grigoras, 2005; McDougall, 2012; Moos, 2012), the aim of this study was also to check the significance of the differences in formant distributions statistically. This was done by means of a series of two-sample Kolmogorov-Smirnov tests which assess the hypothesis that two samples were drawn from different distributions. It is sensitive not only to differences in the location of two samples (their central tendency) but also to the differences in the shapes of the distributions, i.e. differences in skewness and dispersion.

3. Results

The overall classification rate and the classification rates for individual speakers are presented in Table 1. The columns show classification based on static formant values (mean values of F1–F4), dynamic formant values (F1–F3) and both types of parameters combined.

In general, we can see that dynamic formant values are slightly better predictors of speaker identity than static ones. The highest classification rate, approximately 25%, is achieved when both types of formant values are combined. Though the score is not very high, it can be considered a promising result since chance classification would yield a rate of 6%. These parameters thus do provide some speaker-specific cues. Wilks' lambda for static and dynamic values combined is 0.426, the discrimination being highly significant: $F\,(240, 12641) = 4.25$; $p < 0.001$.

If we have a more detailed look at the classification rates of individual speakers, we can see that dynamic values do not score higher than static ones for every speaker. Five out of the 16 speakers (most notably KODA) are better discriminated on the basis of their static values. The table also shows that there is a considerably wide range of classification rates among speakers even when all parameters are used (the highest score is achieved by SOBA, 52.8%, and the lowest by POKA, 6.7%), i.e. while some speakers are identified with relatively high accuracy, others are more difficult to recognize.

Table 1. Classification rate (in %) for individual speakers and total (overall classification) for static and dynamic formant values and for both types of parameters combined.

Speaker	**F1–F4 static**	**F1–F3 dynamic**	**Both**
BURA	1.4	4.1	13.7
DAMA	23.0	16.2	31.1
FISA	15.7	17.1	25.7
KADA	8.2	15.1	15.1
KRIA	28.4	28.4	32.4
PRIA	24.3	23.0	41.9
SMLA	25.7	28.6	22.9
SOBA	32.0	44.0	32.0
STUA	52.8	55.6	52.8
TOMA	29.7	32.4	36.5
KRUA	5.4	20.3	28.4
KODA	13.3	4.0	20.0
KUDA	20.3	18.9	18.9
MIKA	1.3	5.3	14.7
POKA	1.3	4.0	6.7
VRNA	8.0	5.3	16.0
Total	**18.1**	**20.1**	**25.5**

To compare the contribution of individual parameters to speaker discrimination, i.e. to see whether some parameters are markedly better predictors of speaker identity than others, the values of Wilks' lambda for each parameter were examined. The analysis showed that the values are very similar – Wilks' lambda of all parameters falls within a narrow range of 0.429–0.446, with the exception of mean F4 (λ = 0.569). The values of the other static parameters are 0.437 for both F1 and F2, and 0.444 for F3, which is in agreement with general phonetic theory claiming that while lower formants code mainly phonological vowel quality (F1, F2 and to some extent also F3), higher formants – especially F4 – reflect speaker-specific physiological characteristics.

As for dynamic formant parameters, the first value of each of the three formant trajectories tends to have the highest Wilks' lambda. In other words, it appears to be the movement of formants from the preceding consonant which is the most useful point in the trajectory, though the difference is rather small.

A partial objective of this study was to examine the possible effect on classification rate of syllable status with respect to word stress; that is, whether speaker discrimination is more successful in stressed, post-stressed or other unstressed syllables. Czech is a fixed-stress language so the position of stress can be predicted from word structure. Importantly, Czech has no phonological reduction – all five vowels may, unlike in for example English (Johnson, 2004), appear in both stressed and unstressed syllables. The classification rates for individual speakers in stressed, post-stressed and other unstressed syllables are presented in Figure 1.

The most general results (marked as "Total" on the very right of the figure) show almost no difference between classification rates in the three conditions. The score is the highest for stressed syllables (31.2%), only slightly lower for unstressed ones (30.3%) and still a bit lower for post-stressed syllables (28.5%). The figure also reveals considerable differences in within-speaker variability of the scores – while they exhibit a large span for some speakers (notably BURA and PRIA), they are closely comparable for others (SOBA, MIKA, POKA). The presented results also confirm the findings of Palková and Volín (2003) mentioned above, namely that post-stressed syllables behave differently from other unstressed syllables.

As the results of static and dynamic formant parameters have been discussed, long-term formant distribution (LTF) analyses will be now presented. Figures 2–5 show the LTF of F1, F2, F3 and F4, respectively, for individual speakers. The median is captured

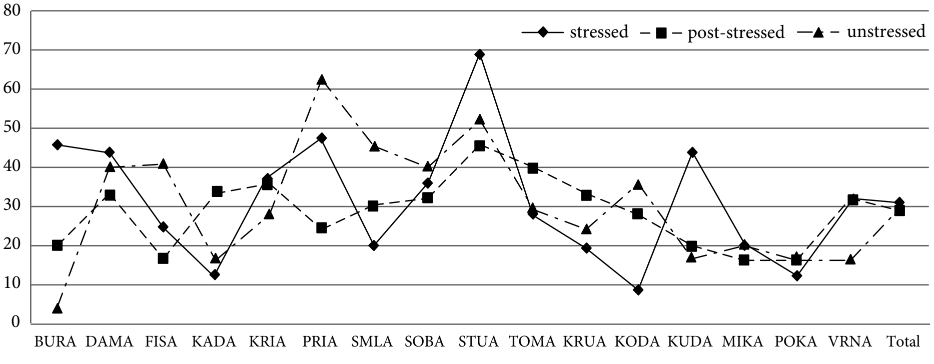

Figure 1. Classification rate (in %) for individual speakers and total in stressed, post-stressed and other unstressed syllables.

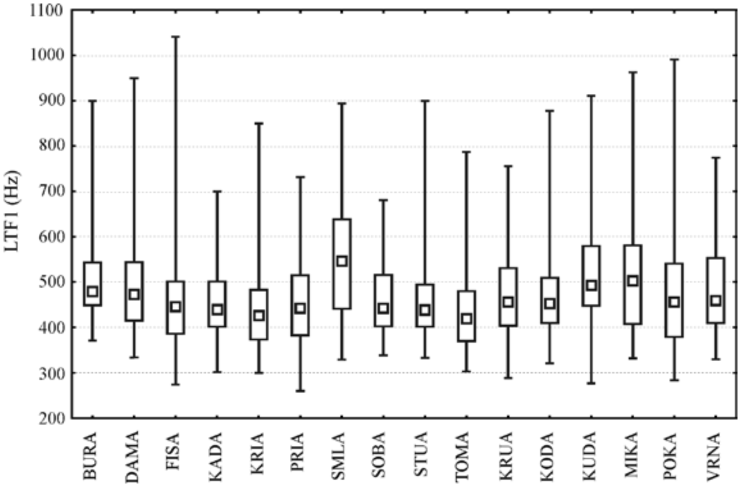

Figure 2. LTF of F1 for each speaker, with the median, middle half (25–75%), and the complete range indicated.

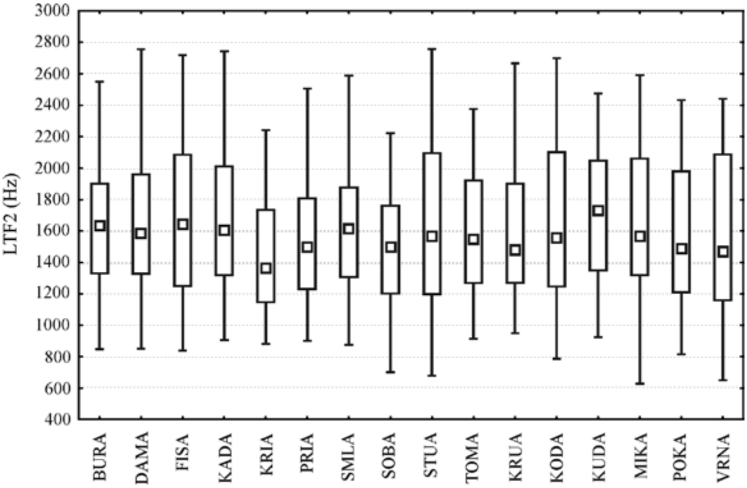

Figure 3. LTF of F2 for each speaker, with the median, middle half (25–75%), and the complete range indicated.

to mark the central tendency of the distribution; the middle half (25–75%) of all values is also signalled.

If we compare the four figures, we can see that the speakers overlap in different degrees. The median and the middle half of LTF values seem to discriminate the speakers best in the case of F3 and F4; for F1 and even more for F2 the degree of overlap is relatively high.

Apart from the LTF mean, speakers can also differ in the distribution of values. It can thus happen that two speakers with similar means exhibit significant differences in the

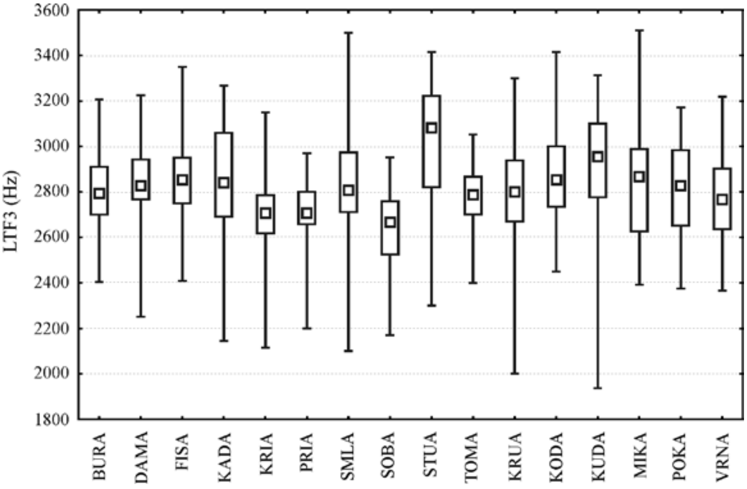

Figure 4. LTF of F3 for each speaker, with the median, middle half (25–75%), and the complete range indicated.

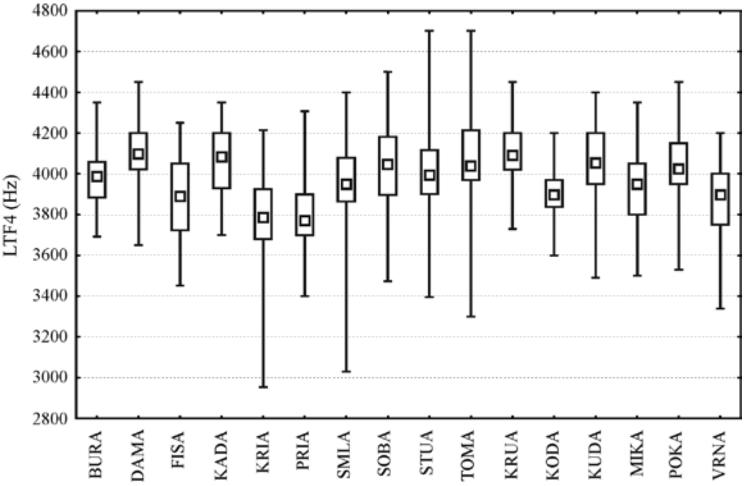

Figure 5. LTF of F4 for each speaker, with the median, middle half (25–75%), and the complete range indicated.

distribution. Such a case is presented in Figure 6 which shows that speakers STUA and TOMA, whose mean LTF value of F2 is almost the same (see Figure 3), differ in formant distribution – the distribution of STUA is more platykurtic (broad peak) and that of TOMA more leptokurtic (narrow peak). The importance of the shape of the distribution for speaker discrimination was highlighted by Moos (2012) who argues that the shape can vary significantly between speakers, but it appears stable within a speaker.

In contrast with previous studies on LTF, its usefulness for speaker discrimination has been assessed also statistically, by means of a series of two-sample Kolmogorov-Smirnov

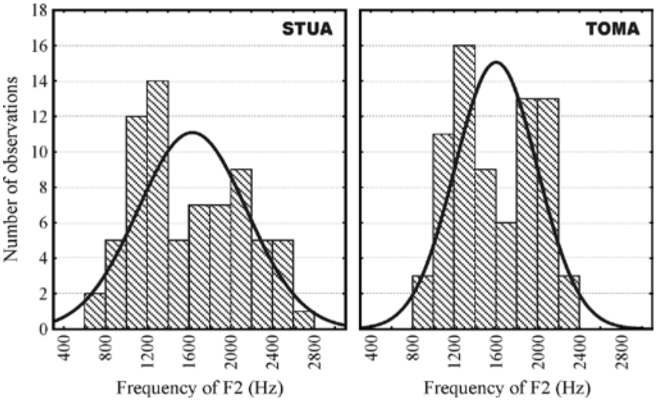

Figure 6. The distribution of F2 values of two speakers. The distribution of speaker STUA is more platykurtic (broad peak), the distribution of speaker TOMA is more leptokurtic (narrow peak).

tests. The outcomes are summarized in Table 2. The distributions of every speaker were compared with those of all other speakers. With 16 speakers, the resulting number of comparisons per each formant was 120; the sum in each row therefore equals 120.

Table 2. The usefulness of LTF (F1–F4) for speaker discrimination assessed by two-sample Kolmogorov-Smirnov test. The numbers refer to the number of pairs which are differentiated by the respective LTF on a given level of significance.

LTF	$p > 0.05$	$p < 0.05$	$p < 0.001$
F1	64	35	21
F2	98	22	0
F3	46	25	49
F4	31	25	64

If we have a look at the comparisons which are statistically highly significant (column $p < 0.001$), we can see that the LTF of F3 and F4 discriminate the most pairs of speakers (49 and 64, respectively), confirming the hypothesis that their distributions are the most speaker-specific. The long-term values of F1 appear to be a considerably worse predictor of identity (discriminating between 21 pairs of speakers) and LTF of F2 the worst, as no pair of speakers shows statistically highly significant differences in the distribution of this formant. Collapsing the two levels regarded as significant together confirms these tendencies, with F2 yielding the lowest number of significant pairwise comparisons. Similarly, the first column shows that the highest number of statistically insignificant results was obtained for F2 and the lowest for F4.

These results are very promising: they suggest that, in contrast with previous examinations of LTFs, it should be possible to perform pairwise statistical comparisons of formant distributions and not only to rely on visual comparisons. This global approach to vowel formants reveals some speaker-specific information for all four formants.

4. Discussion and conclusion

Vowel formants are considered to play a crucial role in speaker identification as they convey speaker-specific cues. The aim of this study was to compare the efficiency of the traditionally used static values with dynamic, time-varying formant values, as well as to examine the usefulness of long-term formant distributions (LTFs) for speaker discrimination.

In general, dynamic values led to a slightly better discrimination between our 16 speakers (20.1%) than the static ones (18.1%). As could have been expected, the highest classification rate was achieved when both types of parameters were combined (25.5%). Though the discrimination is in no way impressive, it is well above the 6% chance classification rate, which indicates that these parameters do convey some speaker-specific information.

The comparatively low classification rates were, at least to some extent, caused by the nature of the speech material. The target vowels appeared in various consonantal contexts, positions in the utterance, and in both stressed and unstressed syllables. Although the recordings were of laboratory quality, the speakers were asked to act the dialogues out so as to achieve some degree of spontaneity. Moreover, we analysed only short vowels, in which target undershoot (Lindblom, 1963) is likely to have affected the formant values.

Overall, the most useful parameter for speaker discrimination in the present study was F4 which, however, tends to be unreliable or absent in forensic casework.

Dynamic formant values might therefore be a useful complement to static ones as they capture not only the phonetic target but also the transitions between the targets, and thus have a potential to reflect idiosyncratic solutions to achieve these targets. A higher classification rate might also be hindered by the automatic extraction used. Fitting the LPC to the speech material is known to be rather complex as the optimum order and the resulting performance vary not only across speakers but also across individual tokens (Vallabha and Tuller, 2002; Harrison and Clermont, 2012). Optimizing vowel formant extraction, i.e. fitting it more closely to the actual material, may lead to a more accurate representation of formant values.

The present study found no major differences in discriminability of speakers in stressed, post-stressed and unstressed syllables. A number of previous studies on speaker-specific cues concentrated on stressed syllables only (e.g. Nolan and Grigoras, 2005). Testing the possible usability of not only stressed but also unstressed (including post-stressed) syllables was motivated by the fact that no clear correlates of the stressed syllables have been found in Czech and that there is no phonological reduction of vowels in Czech in unstressed syllables. It thus appears that speaker-specific information can be contained in stressed as well as unstressed syllables. This is highly beneficial for speaker identification, considering the limited material one usually is bound to work with.

As for the LTF, all four formants seem to provide some speaker-specific cues. The most information about speaker identity is again encoded in F4, the least in the LTF of F2. One obvious source of the wide range of F2 values for a speaker is vowel quality. Moos (2012) in addition showed that the LTF of F2 exhibits the largest differences between read and spontaneous speaking styles. As our material combined the two (read dialogues which were acted out), it could have caused additional within-speaker variability and especially so of the LTF of F2. However, as F4 (and F1) tend to be unreliable or even invisible in telephone speech (Künzel, 2001), the most useful parameter in forensic conditions might be the LTF of F3. Also, the LTF of F3 (as well as F2) does not seem to change significantly for varying vocal efforts (Jessen and Becker, 2010). This is an important finding as Lombard speech is a commonly encountered problem in forensic material.

LTF thus seems to be a powerful tool for speaker identification as speakers can differ in their means and, more importantly, their distributions. However, it should be seen rather as a complementary method to short-term analyses as it fails to reveal sound-by-sound variation which we can expect between speakers. Moreover, some vowel qualities might be more speaker-specific than others. For this, individual sounds need to be analysed. Finally, both LTF and static formant values fail to capture speech dynamics. Dynamic values should therefore be also examined as it might be here where the most idiosyncrasies reside.

Our future research will focus on optimizing automatic formant extraction both in recordings of laboratory quality and telephone speech where formant detection becomes more erroneous (Künzel, 2001). Considering the limited time and material that forensic cases tend to involve, improving automatic detection is crucial and can lead to significantly better results in speaker discrimination.

ACKNOWLEDGEMENTS

This research was supported by the project GAČR 406/12/0298 and by the Programme of Scientific Areas Development at Charles University in Prague (PRVOUK), subsection 10 – Linguistics: Social Group Variation.

REFERENCES

Boersma, P. & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.1.31, retrieved on April 10, 2010 from <http://www.praat.org>.

de Jong, G., McDougall, K., Hudson, T. & Nolan, F. (2007). The speaker discriminating power of sounds undergoing historical change: A formant-based study. Proceedings of 16th ICPhS. Saarbrücken: ISPhS, pp. 1813–1816.

Duckworth, M., McDougall, K., de Jong, G. & Shockey, L. (2011). Improving the consistency of formant measurement. International Journal of Speech, Language and the Law, 18, pp. 35–51.

Goldstein, U. (1976). Speaker-identifying features based on formant tracks. Journal of the Acoustical Society of America, 59, pp. 176–182.

Harrison, P. & Clermont, F. (2012). The Influence of LPC Order on the Accuracy of Formant Measurements across Speakers. Proceedings of IAFPA 2012, Santander, Spain.

Hollien, H. (2002). Forensic Voice Identification. San Diego: Academic Press.

Ingram, J., Prandolini, R. & Ong, S. (1996). Formant trajectories as indices of phonetic variation for speaker identification. Forensic Linguistics, 3, pp. 129–145.

Jessen, M. (2008). Forensic Phonetics. Language and Linguistics Compass, 2/4, pp. 671–711.

Jessen, M. & Becker, T. (2010). Long-term Formant Distribution as forensic-phonetic feature. ASA 2nd Pan-American/Iberian Meeting on Acoustics, Cancún, México.

Johnson, K. (2004). Massive reduction in conversational American English. In: K. Yoneyama & K. Maekawa (Eds.), Spontaneous Speech: Data and Analysis. Tokyo: The National Institute for Japanese Language, pp. 29–54.

Künzel, H. (2001). Beware of the "telephone effect": The influence of telephone transmission on the measurement of formant frequencies. Forensic Linguistics, 8, pp. 1350–1371.

Lindblom, B. (1963). Spectroraphic Study of Vowel Reduction. Journal of the Acoustical Society of America, 35, pp. 1773–1781.

Machač, P. & Skarnitzl, R. (2009). Principles of Phonetic Segmentation. Praha: Epocha.

Marrero, V., et al. (2008). Identifying speaker-dependent acoustic parameters in Spanish vowels. Proceedings of Acoustics '08, Paris, pp. 5673–5677.

McDougall, K. (2004). Speaker-specific formant dynamics: An experiment on Australian English /ai/. International Journal of Speech, Language and the Law, 11, pp. 103–130.

McDougall, K. (2006). Dynamic Features of Speech and the Characterisation of Speakers: Towards a New Approach Using Formant Frequencies. International Journal of Speech, Language and the Law, 13, pp. 89–126.

McDougall, K., Nolan, F., Harrison, P. & Kirchhübel, C. (2012). Characterising Speakers Using Formant Frequency Information: A Comparison of Vowel Formant Measurements and Long-Term Formant Analysis. Proceedings of IAFPA 2012, Santander, Spain.

McDougall, K. & Nolan, F. (2007). Discrimination of speakers using the formant dynamics of /u:/ in British English. In: Proceedings of 16th ICPhS. Saarbrücken: ISPhS, pp. 1825–1828.

Moos, A. (2012). Long-term formant distribution as a measure of speaker characteristics in read and spontaneous speech. The Phonetician, 101/102, pp. 7–25.

Nolan, F. (1994). Auditory and acoustic analysis in speaker recognition. In: J. Gibbons (Ed.), Language and the Law. London: Longman, pp. 326–345.

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.

Palková, Z. & Volín, J. (2003). The role of F0 contours in determining foot boundaries in Czech. Proceedings of 15th ICPhS. Barcelona: ISPhS, pp. 1783–1786.

Pollák, P., Volín, J. & Skarnitzl, R. (2007). HMM-Based Phonetic Segmentation in Praat Environment. Proceedings of the XIIth International Conference "Speech and computer – SPECOM 2007". Moscow: MSLU, pp. 537–541.

Rose, P. (2002). Forensic Speaker Identification. London: Taylor & Francis.

Skarnitzl, R. (2010). Prague Phonetic Corpus: status report. AUC Philologica 1/2009, Phonetica Pragensia, XII, pp. 65–67.

Skarnitzl, R. (2012). Dvojí *i* v české výslovnosti. Naše řeč, 95/3, pp. 141–153.

Stevens, K. N. (1971). Sources of inter- and intra-speaker variability in the acoustic properties of speech sounds. Proceedings of 7th ICPhS. Montreal: ISPhS, pp. 206–232.

Vallabha, G. K. & Tuller, B. (2002). Systematic errors in the formant analysis of steady-state vowels. Speech Communication, 38, pp. 141–160.

Volín, J. (2007). Statistické metody ve fonetickém výzkumu. Praha: Nakladatelství Epocha.

ÚSPĚŠNOST RŮZNÝCH FORMANTOVÝCH PARAMETRŮ PŘI ROZLIŠENÍ MLUVČÍCH

Resumé

Tato studie srovnává užitečnost různých způsobů zachycení vokalických formantů. Ukazuje, že dynamické hodnoty (F1–F3) vedou k celkově úspěšnější diskriminaci mluvčích, než hodnoty statické (F1–F4). Jelikož nebyl odhalen žádný vliv přízvučnosti slabiky na klasifikační úspěšnost, zdá se, že nepřízvučné slabiky mohou být pro účely identifikace mluvčího stejně přínosné, jako slabiky přízvučné. Dlouhodobá distribuce formantů (LTF) se jeví jako vhodné doplnění těchto segmentálních metod, neboť poskytuje přehled všech hodnot pro daný formant. Distribuce hodnot jednotlivých mluvčích se mohou lišit v počtu modů, sešikmení, apod. Zatímco předešlé studie porovnávaly LTF pouze vizuálně, naše studie vyhodnotila výstupy LTF také statisticky, kde se jeho užitečnost pro diskriminaci mluvčích potvrdila. Jelikož ale LTF abstrahuje od rozdílů mezi vokály, měly by být statické a dynamické metody, stejně jako segmentální a dlouhodobé komplementárními přístupy.