Comparative Legilinguistics
vol. 32/2017
DOI : http://dx.doi.org/10.14746/cl.2017.32.5

TEMPORAL PARAMETERS OF SPONTANEOUS SPEECH IN FORENSIC SPEAKER IDENTIFICATION IN CASE OF LANGUAGE MISMATCH: SERBIAN AS L1 AND ENGLISH AS L2

Kristina TOMIĆ
Non-affiliated researcher
Niš, Serbia
kristinatomic89@hotmail.com

Abstract: The purpose of the research is to examine the possibility of forensic speaker identification if question and suspect sample are in different languages using temporal parameters (articulation rate, speaking rate, degree of hesitancy, percentage of pauses, average pause duration). The corpus includes 10 female native speakers of Serbian who are proficient in English. The parameters are tested using Bayesian likelihood ratio formula in 40 same-speaker and 360 different-speaker pairs, including estimation of error rates, equal error rates and Overall Likelihood Ratio. One-way ANOVA is performed to determine whether inter-speaker variability is higher than intra-speaker variability across languages. The most successful discriminant is degree of hesitancy with ER of 42.5%/28%, (EER: 33%), followed by average pause duration with ER 35%/45.56%, (EER: 40%). Although the research features a closed-set comparison, which is not very common in forensic reality, the results are still relevant for forensic phoneticians working on criminal cases or as expert witnesses. This study pioneers in forensically comparing Serbian and English as well as in forensically testing temporal parameters on bilingual speakers. Further research should focus on comparing two stress-timed or two syllable-timed languages to test whether they will be more comparable in terms of temporal aspects of speech.

Key words: forensic speaker identification, cross-lingual comparison, tempo, articulation rate, speech rate, pauses

TEMPORALNI PARAMETRI SPONTANOG GOVORA U FORENZIČKOJ IDENTIFIKACIJI GOVORNIKA U SLUČAJEVIMA KADA SU UZORCI NA RAZLIČITIM JEZICIMA: SRPSKI KAO MATERNJI I ENGLESKI KAO STRANI JEZIK

Apstrakt: Cilj ovog istraživanja je da uporedi mogućnost forenzičke identifikacije govornika, kada su sporni i nesporni uzorak na različitim jezicima, uzimajući u obzir temporalne parametre (tempo artikulacije, tempo govora, stepen oklevanja, tj. stepen zastupljenosti hezitacionih pauza, procenat pauza u govoru, prosečnu dužina pauza). Korpus obuhvata 10 izvornih govornica srpskog jezika koje imaju visok nivo poznavanja engleskog. Svaki parametar je testiran putem Bajesove formule verovatnoće u 40 uzoraka gde sporni i nesporni uzorak dolaze od istog govornika i 360 uzoraka gde oni dolaze od različitih govornika sa datom procenom pouzdanosti. Potom je izvedena jednofaktorska analiza varijanse kako bi se odredilo za svaki parametar da li je njegovo variranje u okviru jednog govornika kada govori dva jezika manje nego variranje među govornicima. U istraživanju je otkriveno da je parametar sa najviše uspeha stepen oklevanja, sa stepenom pouzdanosti 42.5%/28%, a zatim prosečna dužina pauze (35%/45.56%). Iako se istraživanje bavi zatvorenim skupom govornika koji daju sporni uzorak, što nije veoma često u forenzičkoj realnosti, ono nosi veliki značaj za forenzičke fonetičare koji rade na kriminalističkim slučajevima ili kao sudski veštaci. U ovoj studiji po prvi put susrećemo forenzičku komparaciju srpskog i engleskog jezika i po prvi put su temporalni parametri govora forenzički testirani u dva jezika.

Ključne reči: forenzička identifikacija govornika, tempo, tempo artikulacije, tempo govora, pauze

Streszczenie: Celem badania jest analiza możliwości identyfikacji mówcy kryminalistycznego i sądowego podczas zadawania pytań w różnych językach, z wykorzystaniem parametrów temporalnych. (wskaźnik artykulcji, wskaźnik mowy, stopień niezdecydowania, odsetek pauz, średnia czas trwania pauzy). Korpus obejmuje 10 mówców kobiet z Serbii, które znają język angielksi na poziomie zaawwansowanym. Patrametry są badane z wykorzystaniem beayesowskiego wzoru wskaźnika prawdopodobieństwa w 40 parach tcyh samych mówców i w 230 parach różnych mówców, z uwzględnieniem szacunku wskaźnika błędu, równiego wskaźnika błędu i Całościowego Wskaźnika Prawdopodobieństwa. badanie ma charakter pionierski w zakresie językoznawstwa sądowego i kryminalistycznego por1) ónawczego w parze jezyka serbskiego i angielskiego, podobnie, jak analiza parametrów temporalnych mówców bilingwalnych. Dalsze badania inny skoncentrować się na porównaniu języków z rytmem akcentowym i z rytmem sylabicznym.

Słowa kluczowe: identyfikacja mówcy sądowego i kryminalistycznego, porównanie języków, tempo, wskaźnik artykulacji, wskaźnik mowy, pauzy

Introduction

The goal of the current research is to examine the possibility of forensic speaker identification if the question and suspect sample are in different languages (on the example of native Serbian and English as a foreign language) using temporal parameters: articulation rate (AR), speaking rate (SR), degree of hesitancy (DOH), percentage of pauses in speech (PPS) and average pause duration (APD).

In modern times, multilingual communities have become an inevitable part of our reality. There have been numerous occurrences of criminal cases involving multilingual speakers. One of the examples is presented by Künzel (Künzel, 2010: 20):

> An African defendant D was indicted for trading with illicit drugs. Relevant telephone conversations in Igbo intercepted by the police revealed details of the deals. D claimed the GSM phone to which the conversations had been tracked may have been his but that it had been stolen prior to the criminal act. He also claimed the only languages he was able to speak were (Nigerian-) English and some German. Since he was unwilling to deliver a speech sample in either language the court decided the expert should use as reference material several calls made undoubtedly by the defendant in German, using his real name when asking the social welfare department for financial aid.

Another, much older case, presented by Hollien, occurred in America. Namely, in a telephone bomb threat, the defendant had been identified by his voice. However, the forensic phoneticians proved that the voice of the offender was the voice of a non-native speaker of English while the defendant was obviously a native English speaker (Hollien, 1990, as cited in Rose, 2002). Thus, it was proven that the offender and the defendant were not one and the same person. In cases like this, similarly as in the one given by Künzel, it may occur that the offenders are not willing to deliver a speech sample in the language they used while committing the crime, which leaves the forensic experts with speech samples in different languages.

The problem with the forensic phonetic cases of language mismatch is reflected in the fact that until very recently, the comparison of such speech samples has been deemed “counter-indicated” (Rose, 2002: 342). Namely, the implementation of traditional forensic phonetic methods has been considered impossible due to the language-specific phenomena (Rose, 2002).

The aim of this project is to explore and explain under what conditions and to what extent forensic speaker identification can be performed when the speech samples are in different languages.

Temporal parameters in cross lingual FSI

Until very recently, there has been little or no research on cross-lingual forensic speaker identification (FSI) in traditional forensic phonetics. However, with the appearance of software for automatic speaker recognition, the interest into speaker identification in multilingual environment has increased. Recently, many scientists have examined the effectiveness of automatic speaker recognition software in the environments where speaker models are trained with recordings in one language and tested in another (Luengo, et al. 2008). Although it is undeniable that automatic speaker recognition has largely contributed to speaker identification with language mismatch, one must not be misled to believe that the problem is solved in its entirety. Namely, as Jessen put it “automatic speaker recognition is an important development” in forensic sciences; however, it should be observed only “as an addition to phonetic/linguistic methods, not as a replacement” (Jessen 2010: 393).

As Rose (2002) put it, forensic phoneticians discouraged cross-linguistic forensic comparison due to the fact that preferred linguistic parameters (pitch range and vowel quality) might be incomparable in different languages (Rose, 2002). In order to compare speech samples that are in different languages, first it is important to choose parameters which are not language-specific, i.e., the parameters that do not depend on the language spoken but rather reflect the habitual behavior of the speaker himself/herself. To clarify, parameters such as fundamental frequency or formant frequencies tend to reflect the features of the language spoken. Namely, formant frequencies depend on the quality of the speech sound and it is well known that different languages have different speech sounds. For instance, the quality of vowels and their formant values in English are quite different from those in Serbian (Paunović, 2011). Also, some cross-linguistic research has proven that different languages select different default values for F0. For example, Gfroerer and Wagner (1995) and Braun (1995) demonstrate higher mean long-term F0 values for Turkish than for German speakers (Rose, 2002). Therefore, the forensic research in the cases of language mismatch should concentrate on exploring not linguistic but rather extralinguistic and paralinguistic parameters (Köster et al., 2007).

Lehiste (1970) argued that whether there are any significant divergences in the average articulation and speech rates across languages "is not well known" (Lehiste, 1970: 52). More than forty years later, there is still no unanimous agreement on the issue. Namely, there are reports that the values of these parameters may differ across languages (see Laver, 1994: 541); however, many researchers have not confirmed the existence of such differences, at least when the number of sound segments per second is used as the unit of measurement (see Roach, 1998: 153; Trouvain & Möbius, 2014: 277).

Köster et al. suggest that "[t]he only aspects in foreign languages that the expert can still process" are the paralinguistic and extralinguistic ones (Köster et al., 2007: 1845). Temporal parameters of speech are not a genuine linguistic property (Laver, 1994; Trouvain, 2003; Jessen, 2010). Namely, tempo is extralinguistic in nature because it does not bear any inherent linguistic meaning nor does it differentiate any meaning (Trouvain, 2003). Any sentence spoken slower means exactly the same when spoken faster (Trouvain, 2003). Also, they may be paralinguistic to the extent that they "are subject to conventional interpretation" (Laver, 1994: 534) reflecting attitudinal and emotional states of the speaker. Also, these parameters are extralinguistic in the sense that they may be "characteristic of the speaker as an individual, without culturally specific conventional paralinguistic interpretability" (Laver, 1994: 534). Jessen (2010) classifies temporal parameters as habitual and defines this category negatively. Namely, according to him, "speaker characteristics that are subsumed under this category do not have any obvious organic foundation nor are they related to the linguistic conventions that are required or expected by the language system or the social community" (Jessen, 2010: 392). A slightly different view with regard to the social function of tempo is expressed by Roach (1998), who gives the example of Zulu society where slower tempo of speech may be interpreted as "a sign of respect and sincerity" (Roach, 1998: 155). Bearing in mind the arguments expressed above, there is reason to believe that temporal aspects of speech are retained even if a person speaks in a foreign language.

One may argue that it is futile to consider speaking rate and pauses as viable parameters for cross-lingual forensic comparison because silent pauses occur more often when speaking in the foreign language, as non-native speakers tend to leave their hesitancy markers unfilled, produce false starts and make more repetitions and errors (Rieger, 2003). It has been hypothesized that speakers of a foreign language, at least at lower levels, need to employ larger efforts to access the exact repository of vocabulary (Rieger, 2003). Researchers have proven that L2 speech tends to be more hesitant with longer and more frequent pauses (Wu, 2008), it consists of shorter utterances, and contains many more slips of the tongue than L1 speech, but only until the higher level of proficiency has been achieved (Weise, 1984; Lennon, 1990; Poulisse, 1999; Hu, 2007). As second language speakers become more fluent, speaking rate and length of run increase, and the number of filled and unfilled pauses decrease (Lennon, 1990; Gut, 2003, as cited in Gut, Trouvain & Barry, 2007). Thus, temporal parameters of speech may indeed be speaker specific across languages, provided that the speaker is proficient in both languages.

Choice of the unit of measurement

There are a number of measuring units used to express the values of AR and SR. Various researchers used different units such as words per minute (Miller, Grosjean, & Lomanto, 1984; Grosjean & Deschamps, 1975), syllables per minute or per second (Goldman-Eisler, 1968, Miller, Grosjean, & Lomanto, 1984; Grosjean & Deschamps, 1975; Laver, 1994, Künzel, 1997, Trouvain, 2003) and sound segments per second (Fónagy & Magdics, 1960; Osser & Peng, 1964; Walker et al., 1992; Trouvain et al., 2001; Trouvain & Möbius, 2014). The question that is often asked is whether there is a proper unit of tempo measurement and how we should choose it.

The sound segment is rarely used as the unit of tempo measurement; however, certain studies show that it may be the most appropriate one, especially when comparing different languages (Künzel, 1997; Trouvain, 2003). For instance, when comparing English and Italian using syllables per second as a unit, Italian has higher articulation rate, but when they are compared using phones per second as a unit, the difference disappears (Roach, 1998). Similarly, Osser and Peng (1964) found no significant differences for American English and Japanese. In addition, Trouvain and Möbius showed that AR of German and French differ significantly when expressed in syllables per second, while it is not the case when the phones per second are used (Trouvain & Möbius, 2014). On the other hand, den Os (1988) compared Dutch and Italian in terms of sounds per second and concluded that Italian was slightly "slower" (Roach, 1998).

One of the dilemmas with regard to segment as a unit is whether we should use intended segments (phonemes) or actually realized segments (phones). According to Roach (Roach, 1998: 152).

> […] it could happen that in speaking quickly I might produce no more sounds per second than when speaking slowly. In order to get a meaningful measure, it would be necessary to count not the sounds actually observable in the physical signal, but the "underlying phonemes" that I would have produced in careful speech.

On the other hand, Trouvain et al. (2001) showed that the number of realized phones per second correlates with the articulation time best. It is followed by intended phones and realized syllables

(Trouvain et al., 2001). As they conclude, “realized phone best expresses the articulation rate” (Trouvain et al., 2001: 156). However, they agree that the number of realized phones per second as a unit is not unproblematic and that it is not always the best choice. Namely, the number of realized phones depends on reliable phonetic transcription of the acoustic signal. In addition, the definition of the phone as a unit is rather controversial. (Trouvain et al., 2001; Trouvain, 2003) More precisely, there are different views on whether glottal stop should be regarded as a phone, whether affricates and diphthongs are comprised of one or two phones or how to treat degemination of homorganic consonants such as in “cannot” (Trouvain et al., 2001; Trouvain, 2003). According to Trouvain, the unit of tempo measurement should be chosen with regard to the purpose and the methodology of the research in question (Trouvain et al., 2001; Trouvain, 2003).

As the current research deals with cross-linguistic tempo comparison, the appropriate unit for measurement is considered to be the number of realized phones per second. Affricates are observed in both languages, thus, they are considered to comprise of one phone. Contrary to this, diphthongs are present only in English and cannot be observed in Serbian, therefore, for the sake of comparability, they are treated as consisting of two phones. Glottal stop is also regarded as a phone, similarly as in the research performed by Trouvain et al. (2001.

Methodology and procedure

Participants

For the current research, 10 native speakers of Serbian were recorded while speaking spontaneously. To produce maximum possible voice similarity, all of the speakers were female, between 24 and 27 years old, with similar cultural and geographic background and similar accent. Namely, they were all born and have lived in South-Eastern Serbia. Two of the speakers (S8-EB and S9-AB) are sisters. None of the participants have any hearing or speaking difficulties. All of them have a bachelor’s or master’s degree in English Language and Literature. Furthermore, they have all graduated from the Faculty of Philosophy in Niš, Department of English Language and Literature and they were not required to complete any prior tests of language proficiency as it is understood that they are proficient speakers of English. Below, for the convenience of comparison and discussion, the participants will be referred to as S1-MZ, S2-BT, S3-TP, S4-MD, S5-JJ, S6-MM, S7-AK, S8-EB, S9-AB, S10-NM.

Equipment

The recordings were performed in a quiet room using Hama CS-188 headset with microphone. The software used for speech recording and speech analysis was Speech Filing System (SFS), as it is free and very useful for annotating and exporting annotations to XML or Praat TextGreed Formats. The recordings were performed at the sampling rate of 16000 Hz. Different transmission channels were not used in this experiment as it has already been proven that temporal aspects of speech such as articulation rate, speaking rate and pauses are mostly independent of recording conditions (Künzel, 1997).

Procedure

Each participant was interviewed separately in two non-contemporaneous sessions (1 and 2), with a two-week distance between the recordings, to account for intra-speaker variability and to approximate forensic reality as much as possible. Each session consisted of two parts, participants speaking in their mother tongue, Serbian, (1A and 2A) and participants speaking in the foreign language, English, (1B and 2B). The interviews lasted approximately 5 minutes per speaker per language per session. The author aimed to obtain approximately 200 minutes of acoustic material.

During the first session (1A and 1B), the speakers were presented with 10 different slides for each language and were asked to talk about each slide for approximately 30 seconds (modelled on Nakasone & Beck, 2001). Each slide was comprised of a single picture and different sets of pictures were used for different languages. The participants were instructed to speak as long as the image was displayed. The timer was not visible so that the speakers' speed of talking would not be affected by the feeling that they are running out of time (Künzel, 1997).

During the second session (2A and 2B), the speakers were given a map of the city of Niš and were asked to explain to the interlocutor (the interviewer) how to arrive from point A to point B. This setup was designed to elicit more speech from the participants. The same procedure was repeated for both languages, only with different destination and departure points. This is a revisited version of Elliott's map task, combined with Kinoshita's map task, modified for the purposes of measuring temporal aspects in spontaneous speech (Elliott, 2001; Kinoshita, 2001).

The Acoustic Analysis

Out of 200 minutes of the obtained acoustic material, about 120 minutes of speech were analyzed acoustically. The average number of speech samples (tokens) per speaker, per session per language was 8.05 with the average token duration of 23.37 seconds. Pauses were annotated manually, the total number of pauses being 4007. The phone computation was carried out with the help of broad phonetic transcription using IPA symbol chart. The average number of phones per token was 173.75. Transcription included ungrammatical events, "semi-words" such as exclamations and interjections (Trouvain & Troung, 2012), and dysfluences in speech such as false starts, repetitions of sounds, syllables and words, slips of the tongue and other, as all of these comprise natural, spontaneous speech (Künzel, 1997). The acoustic material was annotated by the author by means of the visualization of two windows: waveform and spectrogram, and with the help of auditory feedback. The annotations were exported as XML and then imported into Microsoft Excel to calculate the duration of each token and each annotated pause.

When annotating the acoustic material, the following criteria were applied:

(i) Number of phones was the number of realized phones (Trouvain et al., 2001; Trouvain, 2003)

(ii) The threshold for pauses was set to 100 ms (de Pijper and Sanderman, 1994; Künzel, 1997)

(iii) If a pause consisted of an unfilled (silent) and a filled portion, both durations were added together and counted as one event (Künzel, 1997)

(iv) In word-final and word-initial sound segments that were notably lengthened, the 'normal' duration of the segment was assessed by calculating the mean value of this segment in several environments in both sessions and the 'surplus' portion was counted as a filled pause (Künzel, 1997)

(v) The duration of a silence ending in a word that starts with a plosive was assessed when the average duration of closure phase of the plosive was subtracted from the entire length of that silence.

(vi) The average duration of closure phase of plosives was calculated for each speaker separately

Prior to calculation of numerical values of each of the chosen parameters, for each token, the following set of preparatory measurements was taken:

(i) Overall duration of speech sample
(ii) Total number of pauses
(iii) Number of filled pauses
(iv) Duration of every pause
(v) Sum of the durations of all pauses
(vi) Number of realized phones

The values of relevant parameters for each speaker were derived from the preparatory data in Microsoft Excel with the following formulae:

(i) ***Articulation rate (AR)*** – number of phones / (duration of speech sample – sum of the durations of all pauses)
(ii) ***Speech rate (SR)*** - number of phones / duration of speech sample
(iii) ***Degree of hesitancy (DOH)*** – number of filled pauses / total number of pauses X 100
(iv) ***Percentage of pauses in speech (PPS)*** – sum of the durations of all pauses / duration of speech sample X 100
(v) ***Average pause duration (APD)* -** sum of the durations of all pauses / total number of pauses

Once the acoustic analysis has been completed, and as soon as the raw data has been obtained, the values for each parameter for each speaker were entered into the spreadsheets designed to perform the adequate statistical analysis.

The Statistical Analysis

In order to determine the forensic significance of the chosen parameters, the following analyses were performed using Microsoft Excel 2010 with the Analyse-it plug-in.

Bayesian likelihood ratio approach

For each parameter, pairs of samples were examined (suspect sample in Serbian and question sample in English) using the Bayesian likelihood ratio formula (LR) to determine whether the formula successfully discriminates same-speaker and different-speaker pairs. The total of 40 same-speaker and 360 different-speakers combinations were examined for each parameter. Afterwards, the error rates and equal error rate were calculated to estimate the performance of the formula. The formula of Likelihood Ratio with continuous data applied in this research was introduced by Aitken, who used it for the comparison of refractive indices of glass fragments of the window broken by the offender with the glass fragments found at the suspect Aitken 1995: 180, as cited in Rose, 2002: 320):

$$\mathrm{LR} \approx \frac{\tau}{\alpha\sigma} \times e^{[-\frac{(\bar{x}-\bar{y})^2}{2\alpha^2\sigma^2}]} \times e^{[-\frac{(\omega-\hat{\mu})^2}{2\tau^2}+\frac{(z-\hat{\mu})^2}{\tau^2}]}$$

x̄ = mean of questioned sample; ȳ = mean of suspect sample
μ = mean of reference sample
σ = standard deviation of questioned and suspect samples
τ = standard deviation of reference sample
z = (x̄ + ȳ)/2
w = (m x̄ + n ȳ) / (m + n)
m = number in questioned sample
n = number in suspect sample
α = √ (1/m + 1/n)

The value for the reference sample for a particular parameter was obtained when mean and standard deviation were calculated out of the values of all the speakers for that particular parameter in Serbian. The author chose to calculate the values obtained for Serbian, as this is the participants' mother tongue.

Analysis of variance - ANOVA

One-way analysis of variance (ANOVA) was performed for each of these parameters to calculate whether between-speaker variation is higher than within-speaker variation across languages.

Overall likelihood ratio

Likelihood ratios of the parameters that didn't exhibit statistically significant correlation were multiplied to obtain the overall likelihood ratio (OLR). Then, the equal error rate estimation for each group of parameters was given. To determine which of the parameters do not correlate, we performed Pearson's correlation for the data set in Serbian as well as in English.

Research results: findings

Due to the extensiveness of the calculation process and the spreadsheets containing likelihood ratio calculations, the author is presenting only the summary of the results. Table 1 contains the summary of the reference values for each parameter for the data in Serbian. The reference values below are compiled by calculating mean and standard deviation for all speakers' values for the given parameter in their mother tongue. This constitutes a closed set of reference values, where the person who we need to identify is a member of a limited group of people. Although closed set cases are rare in forensic reality, it is useful to compile such sets for the purposes of experiments.

Table 1. Mean and standard deviation values of reference population for each parameter.

AR		***SR***		***DOH***		***PPS***		***APD***	
mean	***SD***	***mean***	***SD***	***mean***	***SD***	***mean***	***SD***	***mean***	***SD***
11.59	1.05	8.04	1.00	40.78	15.14	30.80	5.32	0.68	0.16

Table 2 summarizes the error rates and equal error rates for each of the tested parameters. If we take articulation rate, for example, we can see that in 26 cases out of 40, same speakers were falsely identified as different, which means that error rate for articulation rate in same speaker pairs is 26 / 40 x 100 = 65%. This ER implies that AR as a parameter is rather unreliable in cases of language mismatch when samples in Serbian and English are compared. On the other hand, if we take a look at different speaker pairs, we can perceive that in 109 out of 360 cases, different speakers were incorrectly identified as same, which constitutes the error rate of 109 / 360 x 100 = 30%. In cases where same-speaker error rate is higher than different speaker error rate, equal error rate (EER) is obtained by lowering the threshold of acceptance and recalculating error rates until we reach the threshold for which these two numbers are equal. For articulation rate, EER is 45% for the threshold of 0.28. This implies that if we used AR as the parameter for speaker identification under the given circumstances, we would be only slightly better off than if we relied on pure chance.

Table 2. Error rates and equal error rates for all parameters.

Parameter	***Same speaker***		***Different speaker***		***EER***	***t***
	Number	***%***	***Number***	***%***	***%***	
Articulation rate	26	65	109	30	45	0.28
Speech rate	21	52.5	133	37	47	0.55
Degree of hesitancy	17	42.5	100	27.78	33	0.75
Percentage of pauses	23	57.5	143	39.72	44.5	0.8
Average pause duration	14	35	164	45.56	40	1.2

According to Table 2, the parameter which exhibits the best EER is degree of hesitancy (33% for the threshold of 0.75), with error rates of 42.5% in same speaker pairs and about 28% in different speaker pairs. Bearing in mind the percentage, we could not rely solely on DOH as a discriminant in the given case; however, it may still constitute a valuable parameter which could be combined with other parameters, at least for certain types of speakers. Degree of hesitancy is followed by average pause duration with EER of 40% for threshold of 1.2. This parameter is specific as it performs better in same speaker pairs, 35%, as opposed to about 46% error rate in different speaker pairs. The ER in different speaker pairs indicates that APD is a rather poor discriminant, however, the EER implies that there may indeed be something speaker specific about this parameter that extends across languages.

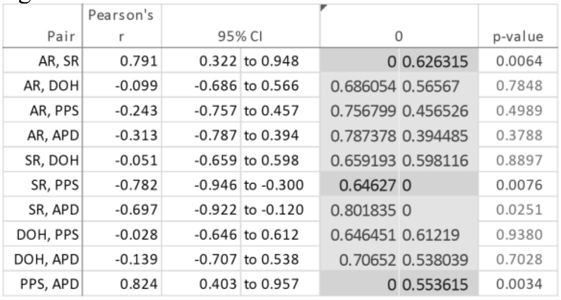
The analysis of variance performed in this research was considered statistically significant for $p \leq 0.01$. Observing Table 3, we can note that the parameters that exhibit greater between-speaker than within-speaker variation across languages include degree of hesitancy, percentage of pauses in speech and average pause duration. The ANOVA results imply that these three parameters are indeed speaker specific regardless of which language is spoken.

Table 3. ANOVA between speakers and within speakers across languages for all parameters, $p \leq 0.01$.

Parameter	***Between gr.***	***Within gr.***	***F***	***P-value***	***F crit***
Articulation rate	7.075013	8.131697	0.966726	0.515617	3.020383
Speech rate	8.645442	7.683149	1.250275	0.364411	3.020383
Degree of hesitancy	4034.1	928.743	4.826236	0.010884	3.020383
Percentage of pauses	414.3933	93.68459	4.914757	0.010202	3.020383
Average pause duration	0.292746	0.049407	6.583486	0.003434	3.020383

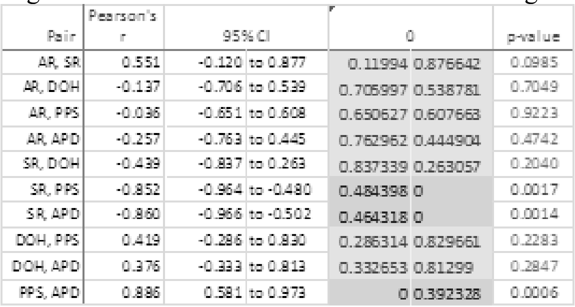
Having tested the performance of each of the parameters separately, the author set to determine whether some of the parameters could be combined to obtain the overall likelihood ratio and thus estimate the probability of observing the evidence given the prosecution and defense hypothesis with higher precision. The analysis of correlation (Figure 1 and Figure 2) revealed that degree of hesitancy is the only parameter that is completely independent of other parameters; therefore, its LR can be combined with LR of any of the four parameters. On the other hand, speech rate arose as the most dependent of all. Thus, apart from degree of hesitancy, it could not be combined with other parameters to obtain the OLR. Additionally, the interdependence is perceived between percentage of pauses in speech and average pause duration.

Figure 1. Pearson's correlation for the data in Serbian.



Pair	Pearson's r	95% CI		0		p-value
AR, SR	0.791	0.322	to 0.948	0	0.626315	0.0064
AR, DOH	-0.099	-0.686	to 0.566	0.686054	0.56567	0.7848
AR, PPS	-0.243	-0.757	to 0.457	0.756799	0.456526	0.4989
AR, APD	-0.313	-0.787	to 0.394	0.787378	0.394485	0.3788
SR, DOH	-0.051	-0.659	to 0.598	0.659193	0.598116	0.8897
SR, PPS	-0.782	-0.946	to -0.300	0.64627	0	0.0076
SR, APD	-0.697	-0.922	to -0.120	0.801835	0	0.0251
DOH, PPS	-0.028	-0.646	to 0.612	0.646451	0.61219	0.9380
DOH, APD	-0.139	-0.707	to 0.538	0.70652	0.538039	0.7028
PPS, APD	0.824	0.403	to 0.957	0	0.553615	0.0034

Figure 2. Pearson's correlation for the data in English.



Pair	Pearson's r	95% CI		0		p-value
AR, SR	0.551	-0.120	to 0.877	0.11994	0.876642	0.0985
AR, DOH	-0.137	-0.706	to 0.539	0.705997	0.538781	0.7049
AR, PPS	-0.036	-0.651	to 0.608	0.650627	0.607663	0.9223
AR, APD	-0.257	-0.763	to 0.445	0.762962	0.444904	0.4742
SR, DOH	-0.439	-0.837	to 0.263	0.837339	0.263057	0.2040
SR, PPS	-0.852	-0.964	to -0.480	0.484398	0	0.0017
SR, APD	-0.860	-0.966	to -0.502	0.464318	0	0.0014
DOH, PPS	0.419	-0.286	to 0.830	0.286314	0.829661	0.2283
DOH, APD	0.376	-0.333	to 0.813	0.332653	0.81299	0.2847
PPS, APD	0.886	0.581	to 0.973	0	0.392328	0.0006

The overall likelihood ratio results can be observed in Table 4. The OLR calculation revealed that the EER is best (32.5%) when LR of degree of hesitancy is multiplied by LR of average pause duration. Other combinations of parameters did not demonstrate any improvement in EER in comparison to the performance of degree of hesitancy on its own. The reason for this is poor individual performance of the parameters involved. In spite of the fact that EER does not improve with the combination of parameters, it is undisputable that discrimination of speakers in different-speaker pairs is significantly improved with the increase in the number of parameters combined. Average error rate for false alarms for three parameters is 11.95%, for two parameters it is 19.03% and for a single parameter it equals 36%. The fact that we can diminish the percentage of false alarms has very positive implications for forensic speaker identification. However, contrary to different-speaker comparisons, in same-speaker pairs with the increase in the number of combined parameters, the chances for missed hits increase as well. The average error rate for same speakers i.e. the percentage of missed hits with three parameters equals 67.5%, with two parameters it is 57.5% and with a single parameter it amounts to 50.5%. Thus, with combination of several temporal parameters we increase the chance for missed hits but diminish the chance for false alarms. Thus, we can conclude that the overall likelihood ratio performance would improve with the improvement of individual performance of each of the involved parameters in same-speaker comparisons.

Table 4. Overall Likelihood Ratios with error rates and equal error rates.

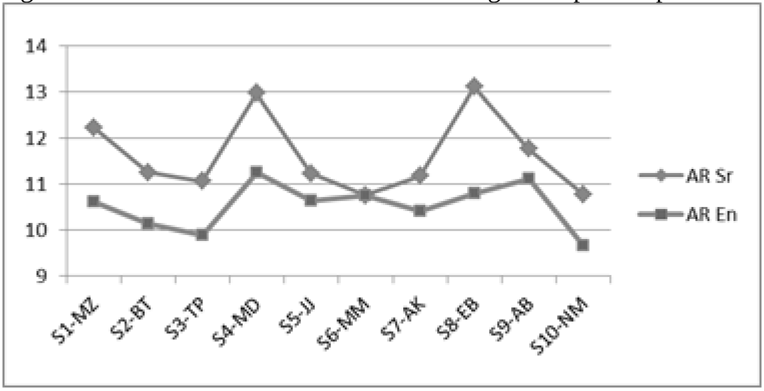
Combination of LR	***Same speaker***		***Different speaker***		***EER***	***t***
	Number	***%***	***Number***	***%***	***%***	
All parameters	24	60	33	9.17	38	0.0008
DOH x APD	19	47.5	80	22.22	32.5	0.3
DOH x PPS	18	45	68	18.89	33	0.2
AR x DOH x PPS	28	70	37	10.28	37.5	0.005
SR x DOH	23	57.5	62	17.22	39.5	0.03
AR x DOH	28	70	57	15.83	40	0.013
AR x DOH x APD	26	65	49	13.61	42.5	0.003
AR x APD	24	60	77	21.39	42.5	0.12
AR x PPS	26	65	67	18.61	44	0.05

Research results: analysis

In order to better understand the performance of individual parameter, and to discover what hinders their performance as forensic discriminants, we are going to take a closer look at the data obtained for those parameters.

To start with, we are going to try to investigate why articulation and speech rate do not seem to be constant across languages. Namely, mean AR for Serbian is 11.63 phones per second, standard deviation 0.82, while for English it is 10.52 phones per second with SD of 0.48. As far as SR is concerned, mean SR for Serbian is 8.07 phones per second with standard deviation of 0.87, while for English, this number is 7.05 with SD of 0.59.

Figure 3. Mean AR values for Serbian and English in phones per second.



In Figure 3 above, we can observe the mean values of articulation rate (AR) for each of the participants for both Serbian and English. By looking at the figure, we can confirm that with the exception of Speaker 6 (S6-MM), whose mean AR values for Serbian and English overlap, the rest of the speakers exhibit rather different means for the two languages. Namely, AR for English is notably lower than for Serbian. The fact that mean AR and SR in Serbian are predominantly higher than in English could mislead us to conclude that the participants tend to speak slower in the foreign language than in the mother tongue. However, lower AR and SR in English may be the result of some other phenomena, such as the structure of the languages themselves. Namely, in English there are segments that have longer duration such as vowels /a:/, /æ/, /ɔ:/, /i:/ or /u:/, whose pronunciation could significantly affect the number of segments per second in comparison to Serbian. Furthermore, English is a stress-timed language, which implies that duration of an utterance depends mostly on the number of stressed syllables, regardless of the number of unstressed syllables in between (Chun, 2002). Therefore, the elision or dropping of segments in an English utterance, as part of connected speech, may not necessarily result in shorter duration of that utterance, which in turn results in lower articulation or speech rate. On the other hand, Serbian is a syllable-timed language, which implies that addition of every syllable (or segment) would prolong the duration of that utterance. Thus, the reason why we did not obtain the expected results may be the fact that Serbian and English are incomparable in terms of number of phones per second.

On the other hand, The parameter that proved to be the most successful as a discriminant in cases where suspect and questioned samples are in different languages is degree of hesitancy with an error rate of 42.5% in same-speaker and 28% in different-speaker pairs. Equal error rate for degree of hesitancy was estimated to be 33%. These results indicate that the ratio of silent and filled pauses in speech is a feature that goes beyond the language spoken, that is, most of the participants are rather consistent in the amount of filled pauses they use when speaking in either their mother tongue or a foreign language. However, the problem with this parameter in the current research is its rather high error rate in same-speaker pairs, which implies that there is a certain degree of intra-speaker variation, and that the chances for missed hits are rather high. As the analysis of variance indicated that the percentage of filled pauses varies more between-speakers than within-speakers across languages, we were able to conclude that the source of variation within-speakers is not only the language spoken. Another possible source of intra-speaker variation of DOH could the condition under which the speech sample was produced. Namely, we were able to observe that percentage of filled pauses for certain speakers changed with the speaking task. To test the variation of DOH between speakers and within speakers across different types of tasks, we performed the analysis of variance (Table 5). The ANOVA results indicate that intra speaker variation across tasks is still lower than inter-speaker variation; however, p>0.01, therefore, we cannot consider this variation statistically significant. Thus, we should infer that the performance of degree of hesitancy was indeed hindered by the conditions under which the speech was produced, at least to some extent.

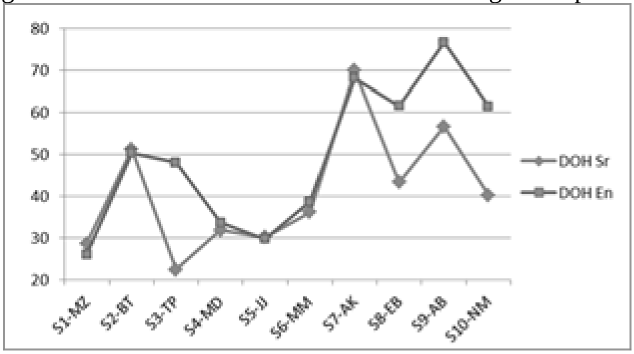
Table 5. DOH - ANOVA between speakers and within speakers across tasks.

Source of Var.	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***
Between Groups	3936.157	9	437.3508	4.094764	0.019187	3.020383
Within Groups	1068.073	10	106.8073			

In Figure 4, we can observe the mean values of degree of hesitancy (DOH) per speaker across languages. Mean DOH for

Serbian is 41% with SD of 13.86, while mean DOH for English is 49.37% with SD of 16.41. Although the two means appear to be close, we can perceive that the range of individual values varies notably with different speakers; S1-MZ exhibits as little as 26% of filled pauses in English, while S9-AB exhibits as much as 78% of pauses.

Figure 4. Mean DOH values for Serbian and English in per cent.



As shown by the numeric values, and as depicted in Figure 4, percentage of filled pauses is on average higher in the foreign language than in the mother tongue. If we observe the graph, we can note that 6 out of 10 speakers exhibit rather similar DOH values for Serbian and English, as opposed to AR where only one speaker had similar values across languages. Speakers whose mean DOH for Serbian and English almost overlap include S1-MZ, S2-BT, S4-MD, S5-JJ, S6-MM and S7-AK. The speaker with the greatest distance between the percentages of filled pauses in the two languages is S3-TP, who exhibits few filled pauses when speaking in her mother tongue (22.38%) but much more when speaking in the foreign language (48%). The mean values of EB's, AB's and NM's DOH appear to differ across languages rather regularly. Namely these speakers' mean DOH for English is approximately 1.4 times higher than for Serbian. The explanation that some of the speakers exhibited higher percentage of hesitation pauses when speaking in the foreign language may lay in the fact that prior to this research, the participants weren't subject to any proficiency or fluency tests. Therefore, these speakers may simply not be proficient enough to exhibit the same amount of hesitation pauses when speaking English.

Research limitation

Although we have shown that some of the parameters explored in this research may be valuable in forensic speaker identification in cases of language mismatch, we ought not to take the results for granted and must be aware of numerous limitations of this research. First and foremost, none of the parameters exhibit the performance of 100% correct identifications or discriminations; what is more, error rates are rather high for the parameters to be completely reliable. Also, we should note that each parameter will perform perfectly well for certain speakers and at the same provide completely misleading conclusions for others. Another important issue is the fact that Bayesian Likelihood Ratio formula is most effective when distribution of data is normal; however, in this group of speakers, not all the parameters exhibit normal distribution of values. The next key issue with regard to the current research is the fact that this was a closed set comparison, as reference sample was calculated from the mean values of the participants themselves. To be precise, in forensic reality, the means of the suspect and questioned sample are almost never contained in the reference population. In the end, we assumed that all the speakers are equally proficient in English based only on their qualifications, without prior fluency or proficiency tests. However, certain differences might exist between speakers. These differences might not be large enough for the speakers not to obtain their qualifications, but could still be large enough to affect the research results. Finally, we should bear in mind that only one researcher (the author) performed the annotation of the acoustic material as well as that the annotations were done manually. Therefore, the risk of subjectivity and personal mistake are not to be disregarded.

Comparison to previous research

The author is not aware of any similar research exploring speaker identification using temporal parameters under the circumstances of language mismatch. Namely, as mentioned earlier, until very recently, forensic comparison under such circumstances had been considered impossible. The following citation of Rose's sheds some light on the reasons why it is difficult to implement traditional forensic comparison methods under conditions of language mismatch (Rose, 2002: 342):

> "It is known that languages can differ in potentially important forensic parameters. […] A little research has been done into the speech of bilinguals, and bidialectals, which shows that they tend to preserve these linguistic differences. Unfortunately, not enough is known yet about bilingual speakers to say whether any voice quality remains the same across two samples of the same speaker speaking in two different languages or dialects. Most likely it will depend on how good a command the speaker has of both varieties. Until we have a much better knowledge of this area, cross-linguistic forensic comparison is clearly counter-indicated."

What changed this trend was the appearance of software for automatic speech recognition, which rapidly evolved into software for speaker recognition. With the development of this technology, the researchers' attention was more and more focused onto speaker identification in multilingual environments (Luengo, et al., 2008). Nowadays, there are many studies that analyze the efficiency of speaker recognition programmes by training the speaker models with recordings in one language and testing them in another (Durou, 1999), (Bhattacharjee & Sarmah, 2012), (Faundez-Zanuy & Satué-Villar, 2006), (Kumar et al., 2009), (Künzel, 2013).

Temporal aspects of speech have rarely been explored in forensic phonetic literature. The most influential study on the topic is an older research by Künzel (1997). What Künzel did was to analyze the temporal aspects of speech of 5 male and 5 female students, in three speaking conditions - spontaneous speech (in an interview), semi-spontaneous speech (a report on the interview), read-out speech - and in two recording conditions – direct recording and recording over the telephone. The parameters that he was concerned with included speech rate (SR), articulation rate (AR), percentage of pauses in speech (PPS), pause free interval, number of syllables between pauses, ratio of silent and filled pauses (SFP), and ratio of pauses with and without respiratory activity (RESP). What he discovered was that AR, with the equal error rate (EER) of 38%, is more reliable as a forensic phonetic parameter than SR, with the EER of about 50%. In addition, he perceived that the values of SR and AR are notably higher in read-out than in spontaneous and semi-spontaneous speech. (Künzel, 1997) More recent studies on tempo as a parameter in forensic speaker identification were performed by Cao and Wang (2011) and Gold (2012). In the former study, 101 Chinese speakers were recorded with the aim to test the inter- and intra-speaker variation of AR. The measuring unit was the number of realized monosyllabic Chinese characters per second. The results indicate that "the global articulation rate (GAR) parameter does not successfully discriminate some of the speakers with GAR values in the central area. However, for those speakers who strongly deviate from the central trend, the GAR becomes a salient discriminatory parameter" (Cao & Wang, 2011). Similar conclusions were drawn by Gold (2012), who measured the AR and standard deviation of 100 male speakers of English and calculated the LR for same speaker and different speaker pairs. The results that she obtained indicate that AR performs much better as the parameter with same speakers (90%) and rather poor with different speakers (46%) (Gold, 2012).

The lack of research in the area is the factor strong enough to encourage scientists to continue investigating the field. With the current study the author hopes to raise interest of other experts and inspire them to draw their attention towards exploring temporal parameters in forensic phonetics. Also, the author believes that this study may inspire other researchers to come up with new ways to overcome the problems of cross-lingual comparison of speakers.

Comparison to previous research

The current research aimed to explore whether traditional forensic speaker identification could be performed if questioned and suspect sample are in different languages (Serbian as mother tongue and English as a foreign language) provided that temporal parameters are used and that the speakers are proficient in both languages. Having applied the auditory-acoustic analysis and standard forensic phonetic statistical procedures, we tested the performance of articulation rate, speech rate, degree of hesitancy, percentage of pauses in speech and average pause duration as forensic parameters under the above-described circumstances.

By completing the current research, we have proven that forensic speaker comparison across languages is not completely "counter-indicated" as usually described. Also, we have shown that temporal parameters, especially degree of hesitancy and average pause duration, exhibit rather large intra-speaker variability in comparison to inter-speaker variability across languages. Furthermore, we have proven that speech rate and articulation rate are incomparable across Serbian and English in terms of phones per second. By exploring forensic speaker identification in cases of language mismatch, we have answered many questions but at the same time raised the new ones and thus opened much room for further exploration in the area.

Bearing in mind the research results, future work in the area should strive to compare different languages using similar methodologies. For instance, it would be useful to discover whether the same differences within speakers as expressed here will occur regardless of the structure of the languages in question. We could compare two syllable-timed languages to find out whether AR and SR manifest in the same manner as when one syllable-timed and one stress-timed language are used. Furthermore, future work should set a goal to determine what kind of normalization could be applied, if at all, to neutralize the discrepancies between the compared languages. Moreover, a useful continuation of this research would be to perform evaluation of the selected parameters with LR for the same participants in their mother tongue only and compare the results with the data obtained in the current study. Also, an alternative pathway would be to obtain the likelihood ratio results of the selected parameters when the participants are bilingual speakers, who adopted both languages before the expiration of their critical period. Finally, as proposed by Luengo et al., (2008) we should strive to implement and combine the findings of the current study with the results obtained by automatic speaker recognition software to ameliorate the performance of such software.

To conclude, despite all the limitations of the current study, and in spite of the fact that the results we obtained are not unambiguous and are far from definite, we should understand that the importance of this research lies in the fact that it attempted to break the deadlock on the issue of cross-lingual forensic comparison. Namely, although traditional forensic comparison of samples in different languages has been characterized as impossible in the related literature, this study proved that there is much room for exploration in the area.

References

Aitken, Colin G. C. 1999. *Statistics and the Evaluation of Evidence for Forensic Scientists*. Chichester: Wiley, 1995.

Bhattacharjee, Utpal, and Kshirod Sarmah. 2012. GMM-UBM Based Speaker Verification in Multilingual Environment. *International Journal of Computer Science Issues 9, no. 6 (2012)*: 373-380.

Braun, Angelika. 1995. Fundamental frequency – how speaker-specific is it? In *Studies in Forensic Phonetics*, edited by Angelika Braun and Jens Peter Köster, 9-23. Trier: Wissenschaftlicher Verlag, 1995.

Cao, Honglin, and Yingli Wang. 2011. A forensic aspect of articulation rate variation in Chinese. *The Proceedings of the International Conference on Phonetic Sciences (ICPhS XVII).* Hong Kong. 396-399.

Durou, Geoffrey. 1999. Multilingual Text-independent Speaker Identification. *Proceedings of Multi-lingual Interoperability in Speech Technology (MIST). Leusden, The Netherlands: PN.* 115-118.

Elliott, Jennifer. 2001. Auditory and F-pattern variation in Australian Okay: a forensic investigation. *Acoustics Australia 29, no. 1* (2001): 37-41.

Faundez-Zanuy, Marcos, and Antonio Satué-Villar. 2006. Speaker Recognition Experiments on a Bilingual Database. *Proceedings of IV Jornadas en Tecnologias del Habla (4JTH). Zaragoza, Spain*. 261-264.

Fónagy, Ivan, and K Magdics. 1960. Speed of Utterance in Phrases of *Different Lengths. Language and Speech 3, no. 4 (1960)*: 179-192.

Gfroerer, Stefan, and Isolde Wagner. 1995. Fundamental frequency in forensic speech samples. In *Studies in Forensic Phonetics,* edited by Angelika Braun and Jens Peter Köster. 41-48/ Trier: Wissenschaftlicher Verlag.

Gold, Erica. 2012. Articulation rate as a discriminant in forensic speaker comparisons. *UNSW Forensic Speech Science Conference 2012.* Sidney, Australia.

Goldman-Eisler, Frieda. 1968. *Psycholinguistics. Experiments in Spontaneous Speech*. London/New York: Academic Press.

Grosjean, François, and A Deschamps. 1975. Analyse contrastive des variables temporelles de l'anglais et du français; vitesse de paroles et variables composantes, phénomènes d'hésitation. *Phonetica 31 (1975)*: 144-184.

Gut, Ulrike, Jürgen Trouvain, and William J Barry. 2007. Bridging research on phonetic descriptions with knowledge from teaching practice – The case of prosody in non-native speech. *In Non-Native Prosody. Phonetic Description and Teaching Practice,* edited by Urlike Gut, Jürgen Trouvain and William J Barry. 1-21. Berlin: De Gruyter Mouton.

Gut, Urlike. 2003. Prosody in second language speech production: the role of the. *Zeitschrift für Fremdsprachen Lehren und Lernen 32 (2003)*: 133-152.

Hollien, Harry. 1990. *The Acoustics of Crime: The New Science of Forensic Phonetics*. New York: Springer.

Hu, Ling. 2007. Long pauses in Chinese EFL learners' speech production. *Interlinguistica 17 (2007)*: 606-616.

Jessen, Michael. 2010. The forensic phonetician: Forensic speaker identification by experts. In *The Routledge Handbook of Forensic Linguistics,* edited by Malcolm Coulthard and Alison Johnson, 702. Abingdon and New York: Routledge.

Kinoshita, Yuko. 2001. *Testing realistic forensic speaker identification in Japanese: a likelihood ratio based approach using formants*. Unpublished PhD Thesis. The Australian National University.

Köster, Olaf, Michael Jessen, Freshta Khairi, and Hartwig Eckert. 2007. Auditory-perceptual identification of voice quality by expert and non-expert listeners. *Proceedings of the 16th International Congress of Phonetic Sciences. Saarbrücken*. 1845–1848.

Kumar, Rajeev, Rajesh Ranjan, Sanjay Kumar Singh, Rahul Kala, Anupam Shukla, and Ritu Tiwari. 2009. *Multilingual speaker recognition using neural network. Proceedings of the Frontiers of Res. on Speech and Music, FRSM 2009. Gwalior, India*. 1-8.

Künzel, Hermann J. 2010. Automatic Speaker Identification with Multilingual Speech Material. Abstracts, IAFPA 2010, *The 19th Annual Conference of the International Association for Forensic Phonetics and Acoustics*. Trier, Germany: Depatment of Phonetics, Trier University. 20.

Künzel, Hermann J. 2013. Automatic speaker recognition with crosslanguage speech material. *International Journal of Speech Language and the Law 20, no. 1 (2013):* 21-44.

Künzel, Hermann J. 1997. Some general phonetic and forensic aspects of speaking tempo. *International Journal of Speech Language and the Law 4, no. 1 (1997).*

Laver, John. 1994. *Principles of Phonetics*. Cambridge: Cambridge University Press.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, Massachusetts and London: The MIT Press.

Lennon, Paul. 1990. Investigating fluency in EFL: *A quantitative approach. Language Learning 40 (1990)*: 378-417.

Luengo, Iker, et al. 2008. Text Independent Speaker Identification in Multilingual Environments. *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC '08). Marrakech, Marocco: European Language Association (ERLA).*

Miller, Joanne L, François Grosjean, and Concetta Lomanto. 1984. Articulation rate and its variability in spontaneous speech: a reanalysis and some implications. *Phonetica 41 (1984)*: 215-225.

Nakasone, Hirotaka, and Steven D. Beck. 2001. Forensic Automatic Speaker Recognition. *Proceedings of 2001 Speaker Odyssey Speaker Recognition Workshop. Crete, Greece*. 1-6.

Osser, Harry, and Frederick Peng. 1964. A Cross Cultural Study of Speech Rate. *Language and Speech 7, no. 2 (1964)*: 120-125.

Paunović, Tatjana. 2011. Sounds Serbian? Acoustic properties of Serbian EFL students' speech. Edited by Eliza Kitis, Nikolaos Lavidas, Nina Topintzi and Tasos Tsangalidis. *Selected Papers from the 19th International Symposium on Theoretical and Applied Linguistics (ISTAL19). Thesaloniki: Aristotle University of Thessaloniki, School of English, Department of Theoretical & Applied Linguistics*. 357-369.

Poulisse, Nanda. 1999. *Slips of the Tongue: Speech Errors in First and Second Language Production*. Amsterdam: John Benjamins Publishing Co.

Rieger, Caroline L. 2003. *Disfluencies and hesitation strategies in oral L2 tests. Proceedings of DiSS’03: Disfluency in Spontaneous Speech Workshop, 5–8 September 2003,* *Göteborg University*. Edited by Robert Eklund. Theoretical Linguistics. 41-44. Sweden: Gothoenburg Papers.

Roach, Peter. 1998. Some Languages are Spoken More Quickly Than Others. *In Language Myths, edited by Laurie Bauer and Peter Trudgill*. 150-158. London, New York, Auckland, Toronto, Ringwood: Penguin Books.

Rose, Philip. 2002. *Forensic Speaker Identification*. London and New York: Taylor & Francis.

Trouvain, Jürgen. 2003. *Tempo Variation in Speech Production: Implication for Speech Synthesis*. PhD Thesis. der Philosophischen Fakultäten der Universität des Saarlandes, Saarbrücken.

Trouvain, Jürgen, and Bernd Möbius. 2014. Sources of variation of articulation rate in native and non-native speech: comparisons of French and German. *Proceedings of Speech and Prosody (SP7)*. 275-279. Dublin.

Trouvain, Jürgen, and Khiet P Truong. 2012. Comparing Non-Verbal Vocalisations in Conversational Speech Corpora. *4th International Workshop on Corpora for Research on Emotion Sentiment and Social Signals (ES3 2012)*. 36-39. Istanbul, Turkey.

Trouvain, Jürgen, Jacques Koreman, Attilio Erriquez, and Bettina Braun. 2001. Articulation Rate Measures and Their Relation to Phone Classification in Spontaneous and Read German Speech. *Proceedings of the Workshop Adaptation Methods for Speech Recognition: Sophia-Antipolis, France, August 29 - 30, 2001* . 155-158.

Walker, Jean F, Lisa M. D Archibald, Sharon R Cherniak, and Valerie G Fish. 1994. Articulation Rate in 3- and 5-Year-Old Children. *Journal of Speech, Language, and Hearing Research 35 (1994)*: 4-13.

Wiese, Richard. 1984. Language Production in Foreign and Native Languages: Same or Different? In *Second Language Productions*, edited by H. W. Dechert, D Möhle and M Raupach, 11-25. Tübingen: Narr.

Wu, Chen-huei. 2008. Filled Pauses in L2 Chinese: A Comparison of Native and Non-Native Speakers. Edited by K M Marjorie, Chan Kang and Hana Kang. *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*. 213-227. Ohio: The Ohio State University.