

SPECIFIC FORMS OF INDIVIDUAL PRODUCTION IN SPEECH EXPRESSION - FORENSIC IMPLICATIONS

Slobodan Jovičić, PhD

Full professor in retirement, School of Electrical Engineering,
University of Belgrade, Serbia

Jelena Đorđević, MSc

Faculty of Security Studies, University of Belgrade, Serbia

Zorka Kašić, PhD

Faculty of Special Education and Rehabilitation, University of Belgrade, Serbia

Abstract: In this work, possible individual productions in speech expression are explained, together with their importance in biometric and forensic identification of a speaker. With the term specific individual productions it is meant that there is a visible variance from typical speakers production in some of the analyzed occurrences in speech. Even though, the native speakers of one language acquire common articulation base both with the environment in the language they are exposed to, personal biological and psychological characteristics of every individual cause various forms of individuality in speech production. Apart from personal biological and psychological characteristics, variance might be caused by specifics of communicative situation. In forensic analysis individuality is researched and it can be expressed in a few features which might be important markers. This work discusses the most important individual characteristics, for example: voice quality, articulation quality, basic tone pitch, specific features of certain speech segments articulation (voices and voice connections, strange words, together with prosody characteristics) tempo, rhythm and scope of international variations and loudness variations). The speaker may produce variances in some of the given markers, and may vary in a few markers because some of them by cause and effect are mutually connected, as for example tempo and articulation quality are connected. With various examples in the real productions (cases) demonstrating a greater number of such speech manifestations, it was emphasized how great their importance is in biometric and above all for forensic analysis in case of speaker identification.

Keywords: speech expression, specific production forms, segment variance, prosody variance, forensic identification of a speaker.

INTRODUCTION

Speech communication is a very complex process. Its sense is in multidimensional interaction among people with which there is a transfer of not only conscious information, but also information on statement, intention, emotions, mood, and other interpersonal relations. Under *conscious information* we mean giving speakers idea to his co-speaker through voice and speech, expressed verbally, using adequate linguistic code. This type of communication enabled a human to have social, creative and intellectual development, i.e. - civilization development (Jovičić, 1999).

When a human speaks he or she behaves as a source of acoustic (speech) signal which is heard by a listener. He or she detects and understands in the intellectual plan thanks to linguistic code which is used both by speakers and listeners. Therefore, *linguistic information* transfer is in question. Anyhow, through this and such communicational channel there are transfers of many other *nonlinguistic information* which is found in literature as *extralinguistic information* and *paralinguistic information* (Mozziconacci, 1998; Scholtz, 2002; Jovičić et al., 2006; Schuller and Batliner, 2014). Extralinguistic information is characterized by the speaker as a source of information, that is, he or she is identified as a specific entity in human population. It is about physical and physiological features like gender, age, pathology in voice, health condition, specific manifestations in voice and speech, etc., which points to speakers identity. On the other hand, paralinguistic information describes speakers complex psycho-emotional state at the moment of speaking. It is about affective manifestations which are transferred through the variety of vocal qualities, like emotions, intentions, adaptations to the topic and ambience of conversation, etc. All the given information is transferred through speech channels simultaneously and that makes up richness of speech expression. The fact that it is not like the case, that only linguistic information transferred to speech would be “robotized”, sounding as if it has no emotions. Examples of people who were born deaf and who did not have feedback acoustic channel, they speak with unnatural prosody, without speech expression of emotions and other attributes which makes spontaneous speech natural.

On the other hand, this information enables speech of a human to appear as one of biometrical features (Woodward et al., 2003) which automated systems of authentication and identification of a person are based on. With such features voice and speech became very interesting for forensic analysis and recently they have become very important arguments in Legal Practice (Rose, 2002; Hollien, 2002; Jessen et al., 2007). If biometrics and forensics are based on the same voice and speech features there is an important difference between them. Biometry is applied in situations which are prior to events like access to restrictive resources or subject verifications, by the rules they happen in controlled conditions, with the cooperation of subjects, by the rule, it is automatized. Forensics activates after occurrences like crime act, preconditions for forensic analysis are not controlled and subjects are often non cooperative. However, the fact is that biometric methods are used in forensics, as well, but the use of expert knowledge means something forensic cannot do without.

Intra-speakers variations in speech production are a problem for biometrics and forensics. On the other hand greater intra-speakers variations lower inter-speakers distance and they can greatly influence reliability of identification of a person. However, there is a very specific production in individual voice and speech which can be very important forensic markers and they can contribute greatly to expert assessment and decision on subject identity. We will show one set of such manifestations which are not present in voice and speech of every subject, but some of them, because of which they are very strong individual forensic marker.

SPEECH PRODUCTION MODEL

In order to better understand specific forms in speech production, let us first look at the model of speech production shown in Figure 1 (Jovičić and Zdravković, 2019). A subject who wants to communicate some information to its surroundings must first in its conscience form an idea what it is that he or she wants to verbally communicate. Then this idea is coded in cognitive level, with linguistic language code in which the subject communicates. This process is very complex, because of this *linguistic programming* means all the mentioned linguistic information: linguistics, paralinguistics and extralinguistics. As Figure 1 shows, some of these pieces of information have influence in certain segments of speech production model. The next step in the formed linguistic message involves inserted certain prosodic rules which should "make alive" speech message through verbal expression. With it the preparation completion for generating motor commands through nervous impulses leading to articulation organs produce speech signal by fine action.

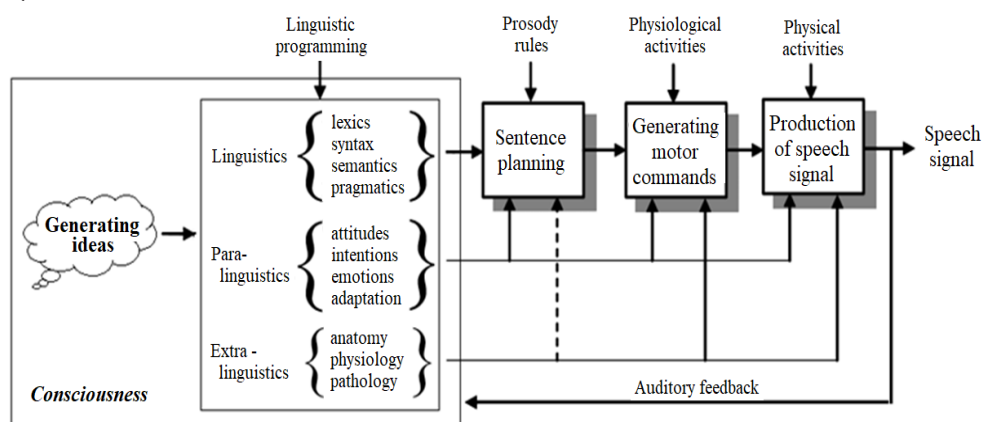


Figure 1: *Speech production model.*

In any segment of voice and speech production a subject may intentionally and most often unintentionally bring additional information. For the listener this information may be but most often is not of any communication importance.

Because of that, we named them in this work as "specific forms in speech expression" which may be of great importance for biometrics and in forensics they may appear as a key forensic markers. We have classified these specific forms in expression in three categories: *specific manifestations in articulation*, *distortions in articulations* and *language particulates (individual particularities in spontaneous speech)*. In further text we will show many examples of these specific forms in verbal expression, examples from forensic practice, with interpretation of their origin and importance for forensic analysis.

SPECIFIC MANIFESTATIONS IN ARTICULATION

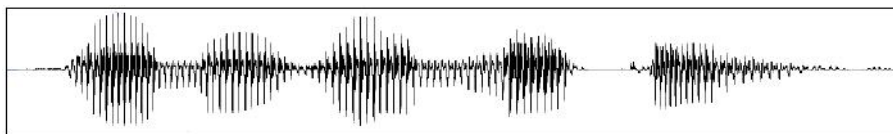
Under the term *articulation organs* we mean all speech organs which participate in speech signal production from lungs to nostril openings and mouth openings. *Articulation* is a process of articulation organs activity. In production of speech signal the glottal system, mouth cavity and nasal cavity are of primary importance. Functioning of these subsystems of speech apparatus together bring *articulation quality* and *voice quality*. In Jovičić (2008) the concepts are defined in this manner:

"*Articulation quality* defines smaller or greater muscular tension of articulator during speech production (Laver, 1980). So, we could conditionally talk about "strong" (tense), normal and "loose" (relaxed, unclear) articulation. For example, expressions with imperative semantic features and threatening expressions are characterized by extremely tense-strong articulation. Strong articulation occurs in speech expression of strong emotions, while relaxed-loose articulation is one of the more important proofs of weak emotions expression, but it also indicates depression, hangover, etc. So, articulation quality as an individual characteristic of the speaker expresses his or her temperament and/or automatized habits and as such certainly and partially participates in general voice quality. *Voice quality* is individual characteristic of a human and as specific auditory colouring of voice is recognisable feature in continual speech. Subjective impression on voice quality is contributed by features generated by phonatory system, phonetic habits, and social and regional context and situation. Voice quality can be willingly controlled in order to be used as linguistic, and most often paralinguistic speech means of communication (for linguistic distinction and phonological contrasts in some languages, for realisation of diverse regional distinctions, and functions as paralinguistic features in speech expression of emotions) but it is, at the same time, an innate individual characteristic of a speaker. Quality of voice is obviously multidimensional characteristics which may be viewed from various aspects - a constantly present fundamental feature of speech which identifies the speaker as an individual as a changeable category during speech process."

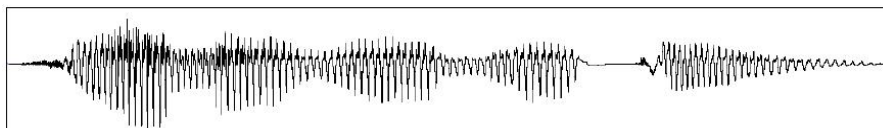
These are a few examples of various and specific manifestations in voice quality and articulation.

In Figure 2, waveforms of one and the same sentence produced with various voice quality is shown: neutral voice, breathy voice, creaky voice, harsh voice and tense voice. Very different forms can be seen in speech signal. For example in creaky voice and harsh voice we see important irregularities, while with tense voice importantly different signal dynamics in comparison with neutral voice exists. These characteristics in speech signal may be important forensic markers in characterization of the subject voice as their inherent characteristic or characteristics of a specific psycho-emotional or health state.

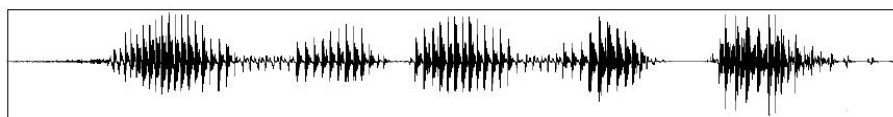
Neutral voice



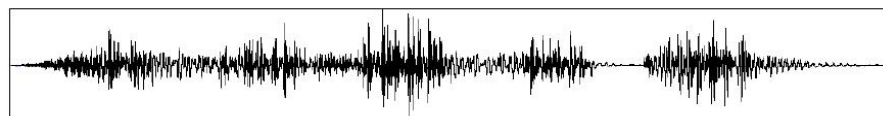
Breathy voice



Creaky voice



Harsh voice



Tense voice

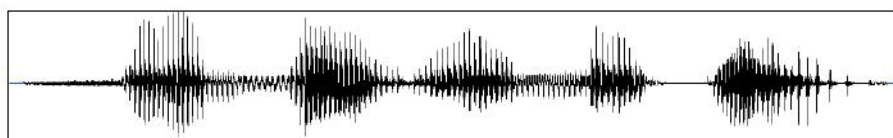


Figure 2: Waveforms of the same sentence with different quality in pronunciation.

The next examples shown in Figure 3 are characteristics of an older person, a person under physical and psychological burden. The occurrence of *creaky voice*, Figure 3a, most often happens at the end of pronounced segment/voice, when it happens that articulation relaxes and especially voice chords relax. In that case fundamental voice frequency (F0) abruptly lowers most often close to half of its nominal value (in spectrogram it can be seen by important scattered glottal im-

pulses). Usual occurrence of creaky effect in voice of a person maybe one of forensic markers.

Inspirium in voice, Figure 3b, happens with stronger inhale most often caused by obstruction problems in the respiratory tract, when it is independent from speech tempo. However, it can happen sporadically with speakers of fast tempo when because of speech acceleration the speaker cannot breathe so easily. Similar noise effect can happen as a typical individual characteristic in expirium form at the end of particular words, Figure 3c.

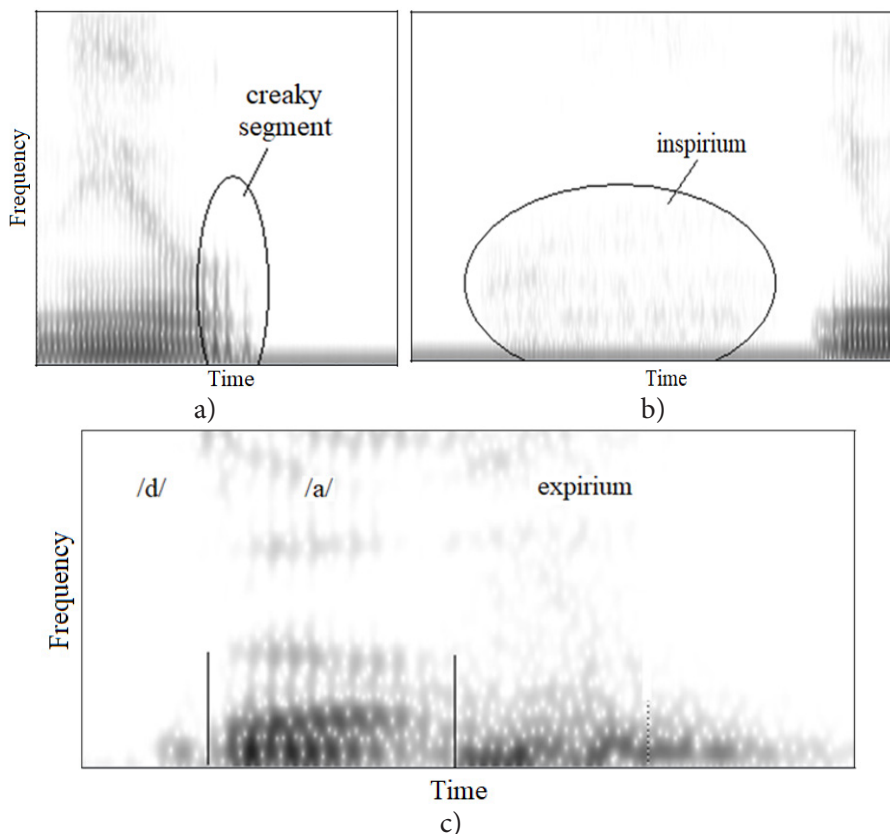


Figure 3: *Some specific manifestation in voice:*
 a) creaky voice, b) inspirium and c) expirium.

In Figure 4a an example of distortion in plosive /k/ is shown. Explosive segment of plosive /k/ in co-articulation with vowels most often has too short acoustic impulses. This explosive segment happens after opening occlusive barrier which with velar voice /k/ is made by back part of the tongue. With production of this voice, the tongue strongly glues to hard palate and it is necessary to have certain air pressure from glottis in order for it to unglue. Then the giving impulses happen by separating specific tongue structures from the palate. When a person has difficulties over control of tongue muscles, then stronger pressure is needed

in order to separate tongue from the palate, while stronger acoustic impulses are generated.

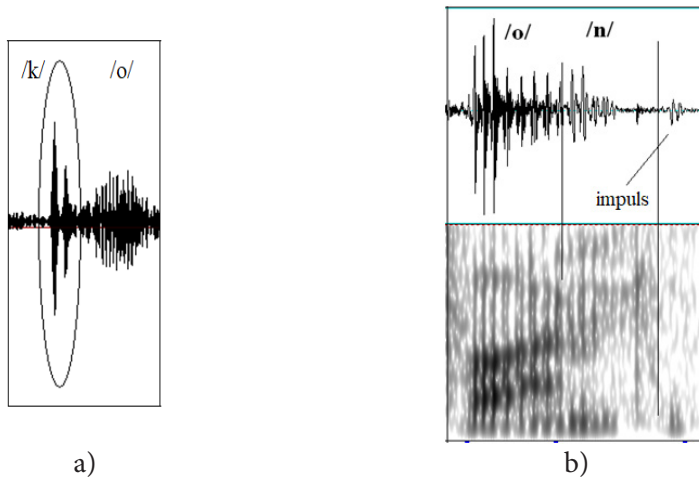


Figure 4: a) *Intensive realization of the explosive burst of the plosive /k/ and*
b) *impulse disturbance in final alveolar voice /n/.*

In Figure 4b we show waveform and spectrogram of speech signal of final parts of words /telefon/. In speech signal after final nasal voice /n/, we noticed an impulse which systematically happened with given subject in more positions in questioned and suspected recordings. The subject was established to wear braces and understanding correlation of tooth braces and seen impulse was such: after pronunciation of an alveolar voice /n/ the top of the tongue touches braces, which is vacuumed to the palate, in the place where it normally touches alveoli. At the end of voice expression /n/ the tip of the tongue is separated from the braces, draws and then returns to the first position. With this an impulse signal is generated which in the recording appears as a disturbance.

DISTORTIONS IN ARTICULATION AND IN VOICE

With the production of specific voice, the articulation organs take place in a certain position in the vocal tract. This position defines acoustic characteristics of produced voice. If the position of the articulation organs is not adequate that induces the change of acoustic picture of the given voice. The changed acoustic picture in softer form shows *distortion in articulations*, and if it is about important change of acoustic picture then *pathology in articulation* happens and consequently pathology in voice. One should mention that distortions in articulation may sporadically happen in normal speech too, especially with fast speakers or speakers with transient and specific psycho emotional state. However, forensic

practice has proved that innate or acquired distortion of articulation is a strong individual characteristic of a speaker. Next examples demonstrate that there is a variety of forms of exceptions in voice articulation

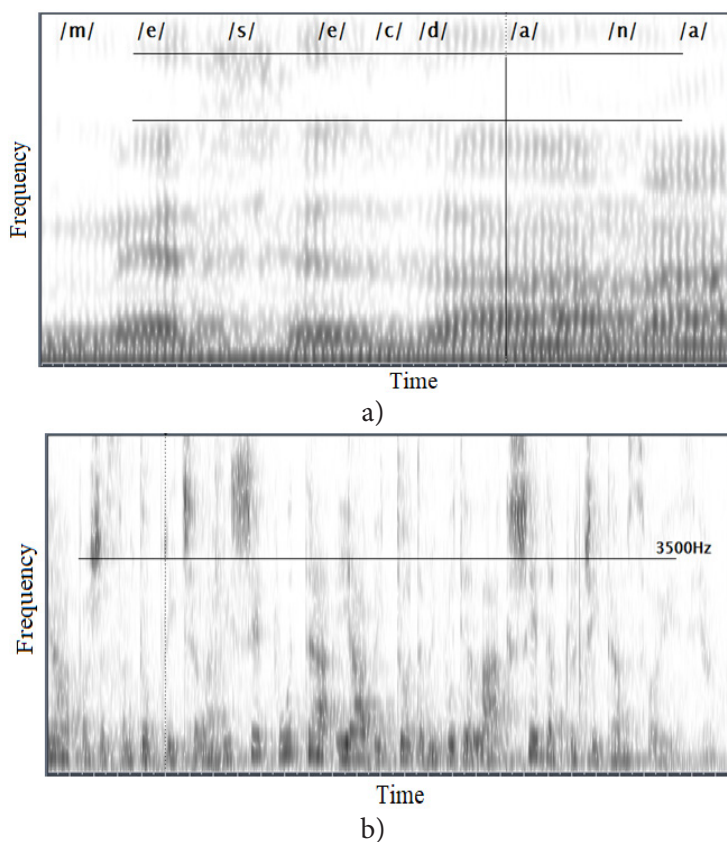


Figure 5: a) *nasalization* and b) *sigmatism* in speech signal.

A spectrogram in Figure 5a shows one speech segment of a subject which has expressive effect of nasalisation. This effect happens when the uvula, which is a very movable organ positioned in the depth of the pharynx in the place where the mouth cavity and nasal cavity connect, does not perform its function. However, with normal production of all voices except nasals, the uvula closes the nasal cavity, so air flow from the lungs comes through the mouth cavity. With the nasal production, the mouth is closed and air flow comes away from the vocal tract through the nose. In that case the uvula opens the nasal cavity, so air flow goes away from the vocal tract through the nose. When a problem in functioning of the uvula occurs, it is relaxed and it cannot qualitatively close the nasal cavity with production of such voices, so air flow comes through mouth cavity and nasal cavity at the same time. The presence of nasal cavity with production of all other voices has the effect of weakening the part-of-speech spectrum. It is shown by a very bleak part of spectrum between two parallel lines, Figure 5a. Nasalization

effect is characteristic of older persons, but also of persons who have the problem in physiological control of normal functioning of the uvula and it is very stable distortion of articulation.

The next distortion in articulations is produced by irregular position of the tongue in the mouth cavity. Because of that, the geometry distortion of the resonant cavities in the anterior part of the mouth cavity and irregular spectral deformations in the spectrum of speech signal are seen. This distortion in articulation is called *sigmatism*. Depending on the position of the tongue in the mouth cavity, there are various forms of sigmatism, but that is a question for speech-language pathologists (SLPs) and we will not take into account this topic in this paper. In Figure 5b, one form of sigmatism is shown when the upper part of a spectre of speech signal is importantly strengthened, beyond 3500Hz, and which is auditi-vely perceived as “lisp- ing”.

A special form of sigmatism is the so-called *stridents*. It happens very often with speakers and this is why we have pointed it out. With pronunciation of dental and palatal fricatives and affricates tip of the tongue is in the correct position, groove in the middle of the tongue exists but it is too deep so that during the friction, the sound similar to whistling occurs. Groove may disappear as a consequence of not enough activities and synchronisation of the tongue muscles work. It can be very intensive and unpleasant to listen. On the other hand, potential occurrence may happen in the normal communication, but if it happens system-ically in speakers speech then it may be an important forensic marker.

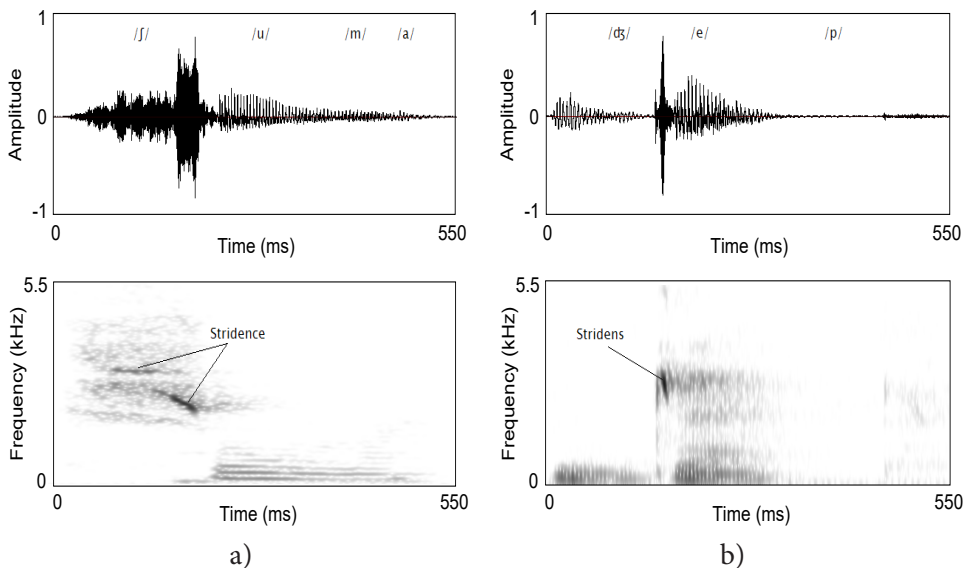


Figure 6: Two examples of stridents.

Figure 6 depicts different forms of stridents. With fricative /ʃ/, Figure 6a, strident occurs as two resonant occurrences changeable in time and in spectrum. With affricates /dʒ/, Figure 6b, stridents occur as very short and intensive whistle.

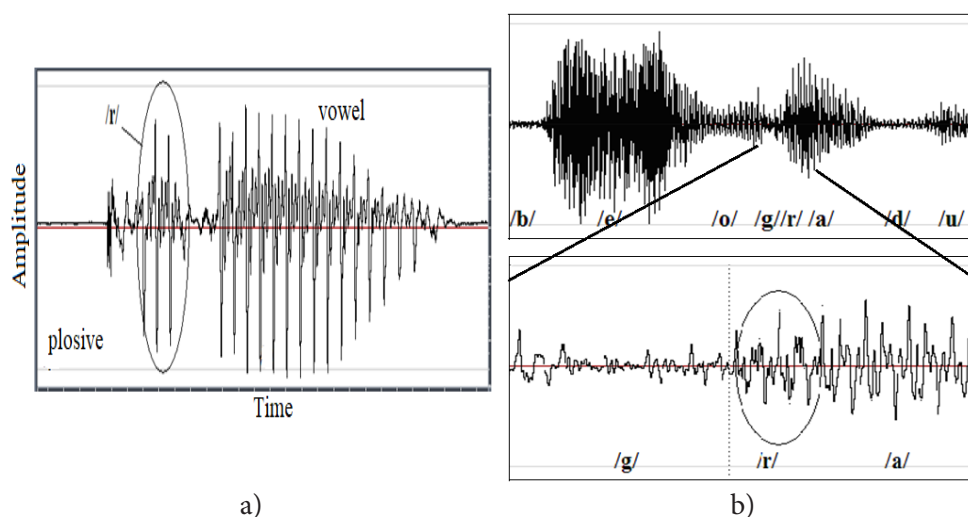


Figure 7: a) normal realization of /r/ and b) atypical realization of /r/.

The next example of articulation distortion is *rhotalism*, in other words untypical pronunciation of vibrant /r/. In normal speech voice /r/ is realized in one, at best three impulses, as a consequence of vibrating the tip of the tongue. In Figure 7a, normal realization of voice of the /r/ is shown with one impulse. However, with persons with label loose articulation or with a problem in muscle control of the tongue, articulators are not able to separate voice /r/ from the next voice. In Figure 7b, the case of co-articulation of voices /r/ and /a/ are shown. In magnified speech segment, it is visible /r/ that is completely integrated with a vowel /a/ and in the context, voice /r/ is still audibly well heard. With harder distortion, when the tip of the tongue is not movable, there can be an omission of voice /r/. With such pathological cases, they may occur in other voices too.

In the end, one should point out that multiple distortions might happen in practice. In our current research we have analyzed perception in distortion of duration and intensity of fricative /ʃ/. This kind of research is important in speech therapist practice with SLPs work in habilitation and rehabilitation of children with problems in speech. However, they are important for forensic experts as well who in listening to recordings should see specific qualities in speech of a person and accept these as forensic markers or not.

In Figure 8, two dimensional (2D) identification surface is shown in perception of variation of duration and intensity of fricative /ʃ/ in the word /ʃuma/. The result is important because it shows what are the boundaries of normal value of

a certain feature in analyzed voice, as an interaction of distortion of one feature to another.

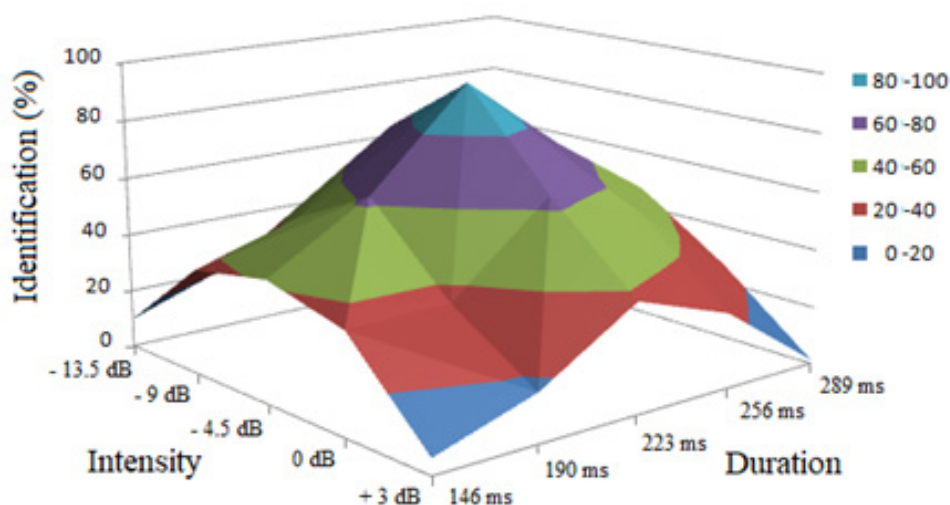


Figure 8: 2D identification surface in perception of variations in duration and intensity.

INDIVIDUAL SPECIFIC QUALITIES IS SPONTANEOUS SPEECH

Individual variations in the plan of speech expressions are simply saying personal speaking features of every human according to which he is different from other humans, because great similarities of speech features occur only in monozygotic twins. The group of individual speech characteristics presents *phonetic capacity* of one person. Phonetic capacity of every individual determines anatomical and physiological characteristics of his system of speech organs, but also the ability or inability to use natural advantages and abilities or inability to compensate weaknesses with phonetic compensations (Kašić, 2000b). With forensic analysis of questioned or suspected samples of speech, from linguistic aspect the individual markers in spontaneous speech are most often determined. Therefore, this questioned samples most often happen in a dialogue which occurs spontaneously between a speaker and co-speaker. Suspected samples occur in controlled conditions, so they are not usually spontaneous. But experts may manage in suspected recordings (according to their experience and according to knowledge of spontaneous occurrences in speech) a separate parts of spontaneous speech in order to compare forensic markers in both questioned and suspected recordings, because of inclusion or disconnection of similarities between this questioned and suspected speaker.

VARIANCE IN THE LEVEL OF SEGMENTS

The specificities of speech quality and articulation of particular sounds, which are determined by acoustic analysis, are analyzed in the previous part of this work. In given examples of variance of voices pronunciation, characteristics are following in certain speech segments in spontaneous speech. Most often special manner of speech of specific consonant clusters occurs as a marker (Kašić, 1997). Speakers of shtokavian dialect with imprecise articulation, often pronounce clusters with connection between dental and alveolar sounds in a specific way. When dental fricative and alveolar lateral are in connection, then the other element of a cluster, that is lateral, is not precisely pronounced (or not pronounced at all), so there are cases *misim, sično, sabo...* instead of *mislim, slično, slabo...* etc. However, when dental plosive is in touch with alveolar nasal then the pronunciation lacks the first cluster member that is plosive. Examples for this happened in real cases (*nevno, nevnik, jeno, svejeno...* instead of *dnevno, dnevnik, jedno, svejedno...*). Most typical cases with specific pronunciation of consonant clusters are given. In the analysis, all types of clusters are subject in observation as an individual characteristics especially in the strangest types of pronunciation. Irregularities in speech of clusters are an important marker for native speakers of shtokavian dialect, because of the phylogenesis of this type of composition, derivation and forms of morphological system of shtokavian dialect, in this articulation base there is the numerous two elements and more complex consonant clusters (Kašić, 2000b).

Special type of divergence in pronunciation of segments are the remains of automatism of dialect articulation base. Namely, articulation base is automatized up to the age of seven, so with speakers who have separated geographically and in speech from their native dialect, a dialect expression of certain segments may happen (Kašić, Đorđević, 2009b). This marker is recorded in famous cases with English language native speakers (Rose, 2002). Dialect markers with native speakers of shtokavian dialect, such as: non differentiating of affricates couples, voiceless sonant /v/, specific pronunciation of lateral /l/, fricative /š/ affricate /č/, open and closed pronunciation of vocal of middle range, reduction of vocal in short syllables ... are valuable in excluding similarities between speakers, while with proving similarities there must be used with special care, that is, there is little evidence of value in proving similarities (Kašić, Đorđević, 2009a).

VARIANCE IN THE LEVEL OF SUPRA SEGMENTS

In the scope of prosodic features of speech, that is supra segments (syllable stress, tempo, loudness, intonation, rhythm), which by definition are produced by the variance of basic acoustic parameters (Kašić, 2000a) we determine important individual markers in identification and profiling of speakers.

When *syllable accent* is in question with a native speakers of shtokavian dialect, individual markers may happen in automatisation in one of the existing regional accent types - four accent system, two accents type with declining accents or one accent - dynamic. Each one of the existing types has its subtypes. However, regional types and double forms of long accents are not a reliable marker because of general conspicuousness. This general conspicuousness enables speakers, while giving suspected samples for expert analysis, to sensibly mask this characteristics. Very often speakers use the exceptional regional characteristics for masking this personal speaker characteristics in questioned samples of speech. So, during the analysis of syllable accident in forensic analysis, it is necessary to determine the clear criteria of differentiating between regional occurrences and individual markers in order not to be trapped for identification. Regional accent characteristics because of early automatisation are a valuable marker when determining the profile of an unknown speaker, and in identification they can help for excluding similarities between speakers. With speakers who in great measure acquired standard speech, a forensic phonetician may use inconspicuous remains of automatism in articulation base for identification; one of the most important markers in this domain is automatism of one of double forms of short accents because both forms are standard, and since the difference between short accents for laymen is almost indistinguishable, both speakers and co-speakers are unaware of, so they do not try to make a distinction in acquiring the standard, and they cannot use it in masking that speech.

Speech *tempo* as automatized personal speech characteristics and features of the temperament (Kašić, 2000a), which can be fast, moderate and slow, gives reliable and inescapable possibilities of determining individual markers in certain sequences of productive continuous speech. And when the speaker tries to disguise and slow the tempo, he/she very often very quickly passes to his or her natural fast speech. The speed of tempo is most often measured by average number of pronounced syllables in unit of time. Even though in our articulation base of shtokavian dialect more than two voices can occur and occur, this measure is in real cases proven as a reliable, proving fastness of speech.

Very often a fast tempo is caused by imprecision of quality of articulation. With fast speakers it is mirrored in the occurrence of non producing certain sounds, sometimes in occurrence of overlapping of syllables, and sometimes only the first part of the word is pronounced and then the second word is pronounced. With very fast speakers it is perceived that this occurrence happens before breathing pause producing unexpectedly a lengthy vowel (*gdee, bratee, samoo, maloo...*). Comparison to the given occurrences with fast speakers contribute to reliability of the speaker identification. A very slow tempo may lead to articulation deformities which also may be one of the forensic markers.

Rhythm, in the most general sense is organised alternating of sound segments and pauses in speech (Kašić, 2000a) that gives the possibility of determining individual markers in speech production. Fluent rhythm means continuous speaking

with adequate breathing breaks which follows syntactic and semantic structure of an expression. However, the given type of fluency is not always possible to make out, so the speakers try to find the solution for too long breaks in speech in different ways. Variance of manners solving too long breaks in spontaneous speech enables determining more reliable individual markers in this domain. In too long breaks speakers often, as automatized habit, produce various speech segment which can represent only voice or word, syntagm, and sometimes the whole expression. In real cases of sample analysis of speech the given possibilities were recorded: /ə/ *neutral vowel* - short or prolonged, *ovaj, onaj, ovoga, vidi, znaš, kako se zove.....* In Figure 9 the spectrogram of /ə/ *neutral vowel* is shown in trying to find the solution to the break in speech.

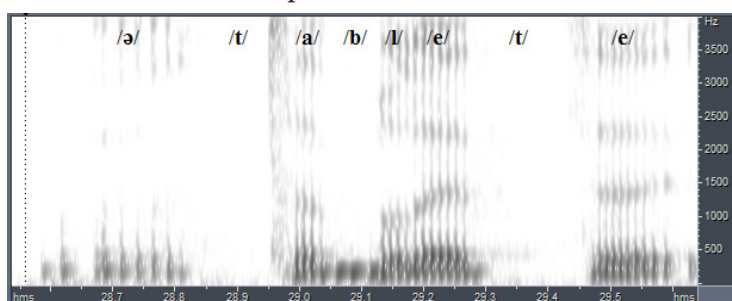


Figure 9: Occurrence of *neutral vowel* (Jovičić, 2008).

Because these sequences are usually automatized, with their analysis one can reach important individual markers. Therefore, these productions may be compared in questioned and suspected recordings, and give possibilities that the voices in syllables in them compare two voices in the similar surroundings in this questioned and suspected samples of speech if there is the possibility that the speaker masks him or herself with language means.

Rarely there occurs the complete non-fluency or stuttering, according to statistics, stuttering occurs 1 of 100 speakers (Golubović, 2000). This type of non-fluency is characterized by long breaks, difficulties with the beginning of a production and similar. In Figure 10 there is a presentation of a spectrogram of “successive” repeating of one sequence through trying to solve the two long break.

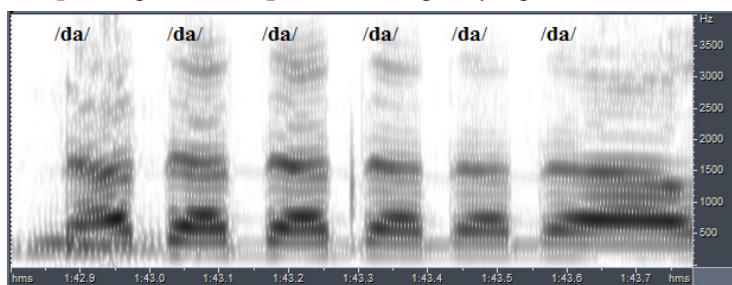


Figure 10: The occurrence of “successive” repetitiveness of the same word.

Furthermore, the number and length of pauses, as a way of beginning of articulation after a pause (attack or performance) maybe one of the distinctive characteristics. A distinctive feature is made up of a specific individual coarticulation variation caused by rhythmic international organisation of speech.

In estimation of a fluent speaker, the scope of intonation and intensity variations are analysed because in a natural spontaneous, well-organized speech, the scope of these variations is very great.

In a certain way *loudness* (speech is perceived as quiet, moderately loud or too loud), as a feature of temperament and as a feature of cultural habits which may happen (in working with other markers), determines individual characteristics of a speaker. Because of different cultural habits in the domain of loudness of speech and difference in personality characteristics which brings to various expression automatized degree of loudness in speech production, this suprasegment is also valuable in identification and determining the profile of an unknown speaker.

CONCLUSION

The paper has analyzed the most important specific forms of individual production in speech expression that are determined as forensic markers in real cases of identification and profiling of speakers. During the forensic identification of a speaker, it is inescapable to do the analysis of voice quality, articulation quality, the basic tone pitch, articulation specific activity of certain speech segments (voices and voices linkage), as with prosodic characteristics (syllable accent, tempo, rhythm and scope of intonation variations and loudness variation).

It is especially important to point out that because of intraspeakers variations in speech production, it is necessary to determine connected markers in order to determine the identity of a speaker. It is also pointed out that in forensic identification of a speaker, an interdisciplinary approach is valuable with acoustic and linguistic aspect in order to be more reliable.

Acknowledgements: This study was supported by grants OI 178027 and OI 178002 from the Ministry of Education, Science and Technological Development of the Republic of Serbia.

REFERENCES

1. Golubović S. (2000). *Poremećaji fluentnosti (Verbal fluency disorders)*. U knjizi – S. Golubović, Z. Kašić: *Segmentna i suprasegmentna organizovanost govora i poremećaji fluentnosti*. Beograd: Društvo defektologa Jugoslavije.
2. Hollien, H. (2002). *Forensic voice identification*. San Diego, Academic press.
3. Jessen, M., Franz, B., Gfroerer, S. (2007). Developments in automatic speaker recognition at the BKA. *IAFPA 2007*, Plymouth.

4. Jovičić, S. T. (1999). *Govorna komunikacija: fiziologija, psihoakustika i percepcija* (*Speech communication: physiology, psychoacoustic and perception*). Nauka, Beograd.
5. Jovičić, S. T., Kašić, Z., Đorđević M. (2006). Paralingvističke i ekstralingvističke informacije u govornim tehnologijama (Paralinguistic and extralinguistic information in speech technologies). *Proc. of Conf. ETRAN*, Beograd, 443-449.
6. Jovičić, S. T., Zdravković, S. (2019). Effect of stress on verbal expression: physiology, psychology and practical implications. *The International Conference "Languages and cultures in time and space 8"*, Novi Sad.
7. Kašić Z. (1997). Promene glasova uslovljene ritamsko-intonacionom organizacijom govora (Changes of voices conditioned by rhythmic-intonation speech organization). *Beogradska defektološka škola*, 1: 77-83.
8. Kašić Z. (2000a). Funkcija suprasegmenata u govornom izrazu (Suprasegment's function in speech). *Beogradska defektološka škola*. 2-3: 113-124.
9. Kašić Z. (2000b). *Segmentna i suprasegmentna organizovanost govora* (*Segmental and suprasegmental speech organization*). U knjizi – S. Golubović, Z. Kašić: *Segmentna i suprasegmentna organizovanost govora i poremećaji fluentnosti*. Beograd: Društvo defektologa Jugoslavije.
10. Kašić Z., Đorđević J.P. (2009a). Zašto je lingvistika postala forenzička veština? (Why did linguistics become a forensic skill?). *Istraživanja u specijalnoj pedagogiji*. Beograd: Fakultet za specijalnu edukaciju i rehabilitaciju, 469-482.
11. Kašić Z. Đorđević J.P. (2009b). Ostaci automatizma artikulacione baze kao forenzički markeri (The remains of the regional articulation base automaticity as forensic markers). *ETRAN 2009*, Zbornik radova. Vrnjačka banja, AK4.1-1-4
12. Laver, J. (1980). *The phonetic description of voice quality*. Cambridge University Press, London.
13. Mozziconacci, S. (1998). *Speech variability and emotion: production and perception*. Doctoral thesis, Technische Universiteit Eindhoven.
14. Rose, P. (2002). *Forensic speaker identification*. Taylor & Francis, London.
15. Schotz, S. (2002). Linguistic & paralinguistic phonetic variation in speaker recognition & text-to-speech synthesis. *GSLT: Speech Technology 1*, Term paper, Autumn 2002.
16. Schuller, B., Batliner, A. (2014). *Computational paralinguistics: emotion, affect and personality in speech and language processing*. Wiley & Sons, UK.
17. Woodward, Jr. J. D., Orlans, N. M., Higin, P. T. (2003). *Biometrics: Identity Assurance in the Information Age*. McGraw-Hill, Osborne Media.