UDK 811.163.42'342.22
Izvorni znanstveni članak
Prihvaćeno za tisak: 16. listopada 2020.
https://doi.org/10.22210/suvlin.2020.090.02

Ana Vidović Zorić, Marko Liker
Filozofski fakultet Sveučilišta u Zagrebu
anvidovi@ffzg.hr, mliker@ffzg.hr

Speech errors and articulatory gestures: an electropalatographic investigation

One of the major findings of traditional investigations of speech errors is that a single segment is the smallest unit affected by speech errors, although the possibility that units smaller than the segment could play a role in explaining speech errors was not entirely rejected. A number of relatively recent studies using instrumental kinematic techniques for speech analysis bring evidence that errors often occur at subsegmental units of speech production, i.e. at the level of articulatory gestures. Such errors occur due to the coproduction of articulatory gestures, whereby the gestures from both the target consonant and the competing speech sound overlap. As they represent the gradient shift from one segment to another, they are often called gradient errors. Such processes are almost impossible to capture without the use of instrumental kinematic techniques, such as electropalatography (EPG). There are no instrumental kinematic studies of speech errors in Croatian speech. Thus, the aim of this paper is to use EPG to investigate speech errors produced in one Croatian tongue twister. The analysis was focused on /r/ and /l/ targets produced by 10 native female speakers of Croatian, while producing the tongue twister which facilitated speech errors in these two sounds. Each token of the target consonant was classified in one of the four categories: (1) perceptually and articulatorily correct production (P1A1); (2) articulatorily correct but perceptually incorrect production (P0A1); (3) perceptually correct but articulatorily incorrect production (P1A0) and (4) perceptually and articulatorily incorrect production (P0A0). The classification was made by the authors via auditory analysis and visual inspection of spectrograms and electropalatograms prior to the quantitative kinematic analysis. Subsequent analyses showed evidence of gradient errors, which would not be detected without the use of instrumental kinematic techniques. This investigation supports the claim that traditional method of collecting speech errors by perceptual analysis only is not sensitive enough to detect the subtleties of erroneous productions and speech motor control.

1. Introduction

Speech errors, defined as unintended, non–habitual deviations from a speech plan (Dell 1986), are an exhaustively studied phenomenon in speech production and perception research. They were considered "a window into the human mind" (Fromkin 1973: 11), i.e. the reflection of a breakdown in speech planning at higher cognitive levels of speech production process. Those traditional investigations utilized mostly speech perception techniques and phonetic transcription in detecting and annotating speech errors. One of the main conclusions drawn from these investigations, based on the observation that most errors occur at the level of a segment, was that the phoneme is the smallest unit affected by a speech error (Fromkin 1971; Shattuck–Hufnagel 1979, 1983, 1986; Shattuck–Hufnagel and Klatt 1979; Stemberger 1982). Nevertheless, most researchers who agreed with this conclusion admitted that in some cases, as in the example given by Fromkin (1971) – *glear plue sky* instead *clear blue sky*, the error could be explained not just as a phoneme substitution (in this example /k/ for /g/ and /b/ for /p/), but also as a substitution of two distinctive features (+/– voice). Since such ambiguous examples occurred rarely in the corpus, Shattuck–Hufnagel and Klatt (1979) suggested that these errors should be interpreted as a primarily segmental phenomenon. Furthermore, Shattuck–Hufnagel (1986: 141) also noted that "distinctive features (or dimensions very much like them) play a role in the representation of both vowels and consonants", but emphasized that "planning units correspond to individual segments". According to this claim, Shattuck–Hufnagel (1979, 1986) proposed the slot–and–fillers model of phonological encoding which predicted that phonological representations of segments would be copied one by one in an appropriate slot determined by suprasegmental frame. Each segment could fill only one slot, and slots were linearly arranged. Nevertheless, phonemic similarity, which is greater as two interacting segments share more distinctive features, significantly increases the possibility of segment substitutions and exchanges (Dell and Reich 1980; Levitt and Healy 1985; MacKay 1970; Shattuck–Hufnagel and Klatt 1979; Wilshire 1999). In addition to phonemic similarity, there are two other kinds of similarities which could also have a great influence on speech errors and these are positional and context similarity. Positional similarity refers to the same or similar position in the higher level units (syllable or word) of two interacting segments. For example, MacKay (1970) revealed that in his study the vast majority of reversed consonants and vowels occurred in the same syllabic position. Similar results are observed in many other studies (Boomer and Laver 1968; Fromkin 1971; Nooteboom 1969; García–Albea et al. 1989). Context similarity refers to the identical preceding and/or following segments to the interacting ones (Dell and Reich 1980; MacKay 1970).

The majority of the models of phonological planning and phonetic implementation resulting from traditional approach to speech errors were mostly modular and/or linear (for example, Fromkin 1971; Shattuck–Hufnagel 1979, 1986; Levelt 1989). Modular approach proposes that speech processing involves several discrete modules, that is, levels of linguistic planning, with no interaction between them (Levelt 1989). The output of one module is the input for the next one.

Modern approaches to speech errors often rely on acoustic analysis and instrumental kinematic techniques, such as electromagnetic articulography (EMMA), electropalatography (EPG) or ultrasound (UTI). These kinematic techniques give researchers the opportunity to gain insight into articulatory movements during speech errors. These approaches shed new light on speech errors and speech motor control in general. They confirm the findings from speech perception studies (Bond 2008; Cutler 1981; Ferber 1991) that the method of collecting errors by auditory analysis only is not sensitive enough to detect every deviation from the speech plan (Frisch and Wright 2002; Goldrick and Blumstein 2006; Goldstein et al. 2007; McMillan and Corley 2010; Mowrey and MacKay 1990; Pouplier 2007, 2008). Even if such deviations could be captured by impressionistic judgements, the classification of such speech output could not be made in terms of well–established phonological categories. These findings gave boost to reopening the question of the smallest unit affected by speech error, moreover, the smallest unit of speech production. Articulatory studies (Goldstein et al. 2007; McMillan and Corley 2010; Mowrey and MacKay 1990; Pouplier 2007, 2008) showed that a target and an intrusive muscular activity could appear simultaneously within one segment. Goldstein et al. (2007) presented an EMMA study where subjects were asked to repeat two–word phrases with alternating syllable onset consonants, as in *cop–top*. The results showed that some /k/ tokens contained an intrusive tongue tip raising gesture, typical for the competing /t/, which was coproduced with the expected tongue dorsum raising gesture, typical for the intended /k/. The reverse situation was found in some /t/ tokens, which contained an intrusive tongue dorsum raising gesture together with tongue tip raising gesture typical for /t/. Goldstein et al. referred to this type of errors as a "gestural intrusion errors". On the other hand, they also confirmed that sometimes articulatory targets were just reduced and not coproduced. These muscular intrusions or reductions of intended gestures could be activated in various degrees.

Acoustic studies also confirmed (Frisch and Wright 2002; Goldrick and Blumstein 2006; Goldrick et al. 2016) that errors need not pertain to the whole segment exclusively (i.e. phoneme substitutions, omissions or exchanges), but they could be produced with a partial presence of an acoustic feature atypical for intended segment, or as a partial absence of the expected feature. For instance, Frisch and Wright (2002) measured the amount of periodicity in the total duration of the fricatives /s/ and /z/, produced in tongue twisters which contain target segments at the same syllable position, for example: *sit zap zoo sip*. Besides categorical errors, in which /s/ was completely produced as /z/ and vice versa, they also found that some /s/ tokens were partially voiced and some /z/ tokens were partially devoiced. Frisch and Wright referred to this kind of errors as gradient errors, as opposed to categorical errors, and identified their source at a sub–feature level. It was hypothesized that gradient errors reflected gestural intrusions or reductions in the articulatory implementation of the phonetic plan (Goldstein et al. 2007; Pouplier and Goldstein 2005).

The existence of gradient errors is discussed within two major theoretical frameworks: (1) gestural speech production framework, developed by Browman and Goldstein (1992) (e.g. Goldstein et al. 2007; Pouplier 2007; Pouplier and Goldstein 2010) and (2) models of cascading activation (Goldrick and Blumstein 2006; McMillan and Corley 2010; Rapp and Goldrick 2000). Models of cascading activation presume that the activation of each competing unit at the higher stages of speech processing is transmitted to the lower stages (Goldrick and Blumstein 2006). In the example given by Goldrick and Blumstein, at the lexical processing stage, apart from the target unit *calf*, other semantically related candidates such as *cow*, *cub*, *lamb* and *foal* would be partially activated at the same time. The same principle of activation also operates at the lower processing stage – phonological level, which means that, apart from the target /æ/ from *calf*, an intrusive /aʊ/ from *cow* as well as /ʌ/ from *cub* would be also activated. While cascading activation between semantic, lexical and phonological level of representation has been extensively studied, the interactions between phonological planning and articulatory implementation remain relatively unknown. Goldrick and Blumstein (2006) tried to reconstruct this aspect of speech processing within the framework of cascading activation theory. Their acoustic analysis showed that erroneous segment reflected traces of the target segment. For example, if the target utterance *keff geff* was produced as [kɛf kɛf], a certain amount of voicing feature would often be found in erroneous /k/ tokens. This finding allowed them to conclude that errors reflected the simultaneous activation of both, the target and the competing phonological representation, resulting in gradient errors. Goldrick and Blumstein (2006) tried to eliminate one of the major objections to the notion of gradient errors – the claim that it was not clear whether the source of the traces in gradient errors was a competing unit of the higher processing level or the adjacent segment in the coarticulation process. For that purpose, they measured the amount of voicing in erroneous [k] productions and found that voicing did not significantly differ in the two conditions – in *geff geff* erroneously produced as [gɛf kɛf], or in *keff geff*, erroneously produced as [kɛf kɛf]. If the assumption about coarticulatory influence had been correct, then in errors like *keff geff* produced as [kɛf kɛf], erroneous /k/ should not contain voicing feature at all, which it evidently did. Based on this evidence, Goldrick and Blumstein concluded that the source of voicing in the segmental error was in the cascading activation of the target segment, and not in the previous neighboring segment. Another account of gradient errors was given within the gestural speech production theory, developed by Browman and Goldstein (1992). This theory assumed that phonological representations were not static symbols, but rather dynamic modes, called articulatory gestures, which were considered to be the atomic units of speech production (Browman and Goldstein 1992). Articulatory gestures were, as Browman and Goldstein further explained, the result of vocal tract movements, which could be coupled and combined in many ways to achieve linguistically relevant goals. They were entirely physically defined, by their spatio–temporal properties. According to this view, these spatio–temporal properties were already defined within phonological representation of each segment, which meant that the division between an abstract phonological level and a physical phonetic level of speech production was no longer required (Pouplier and Goldstein 2010). Therefore, if the gestures were indeed basic speech units of speech, then they would be individually affected by the error, and that was exactly what articulatory studies found – the coproduction of two gestures, the unintended and the target one (Goldstein et al. 2007; Pouplier 2007; Pouplier and Goldstein 2010).

To date there are no instrumental kinematic data on speech error productions in Croatian. It is important to study these processes in different languages, because cross–linguistic comparisons can shed new light on long–standing research problems in linguistic and phonetic sciences (Kendall et al. 2015; Wells–Jensen 2007). Therefore, in this paper, we analyze unintentional speech errors produced in a tongue twister in Croatian by means of electropalatography (EPG). The aim was to use an instrumental kinematic technique to analyze unintentional speech errors and investigate their categorical versus gradient nature. The EPG is well suited for this type of investigation, because it is the only instrumental physiological technique which provides detailed analysis of one of the most important aspects of speech production physiology – tongue–to–palate contact. Based on the literature review above, it is our hypothesis that evidence of gradient speech errors will be found in Croatian.

2. Methodology

2.1. Participants

Ten native female speakers of Croatian language originally from north–western Croatia and living in the capital city of Zagreb for the minimum of 2 years prior to recording. They had no hearing and speech impairments and their age ranged between 19 and 22 (mean age 20.2). They were all students of the Faculty of Humanities and Social Sciences at the University of Zagreb. All participants were treated in accordance with the guidelines of the Ethics Committee of the Faculty of Humanities and Social Sciences at the University of Zagreb.

2.2. Speech material

The analyzed speech material is a part of the simultaneous acoustic and electropalatographic corpus of Croatian speech (R–kor). The corpus consists of the recordings from 10 female speakers, who were recorded every six months for a period of three years. The corpus includes recordings of sentences, tongue–twisters and quasi–spontaneous speech. The analysis for the present investigation was based on the Croatian tongue twister *Kralj Karlo i kraljica Klara krali klarinet* [kraʎ karlo i kraʎitsa klara krali klarinet] (Eng. *King Karlo and queen Klara were stealing a clarinet)* produced in two sessions in the same year. The analyzed speech sounds were /r/ and /l/ tokens in Croatian words /klara/, /krali/ and /klarinet/ (Eng. *Klara*, *steal*, *clarinet*). A speech sound arrangement in that tongue twister facilitates the elicitation of speech errors in consonants /r/ and /l/, as these consonants are in a similar articulatory context and are articulatorily similar as well. Sound /r/ is in Croatian literature (Horga and Liker 2016) described as a voiced apico–alveolar trill, and /l/ is voiced apico–alveolar lateral approximant. Nevertheless, the /l/ productions in Croatian vary both across and within speakers (Horga and Liker 2016), and this variability needs further research. Each participant repeated the tongue twister for six times. Target consonants in self–repairs of the target words as well as in multiple attempts to pronounce those words correctly were also included in analysis. The total of 409 tokens were analyzed (221 for /r/ and 188 for /l/).

2.3. Instrumentation and procedure

The EPG signal was recorded using the WinEPG System, with the EPG data sampled at 100 Hz. Acoustic data were recorded simultaneously using the WinEPG System and M–Audio MobilePre external USB sound card/pre–amplifier with the sampling rate of 44 100 Hz. The Articulate palate (Figure 1) was used, which consists of 62 electrodes adjusted to anatomical characteristics of each participant (Wrench 2007).

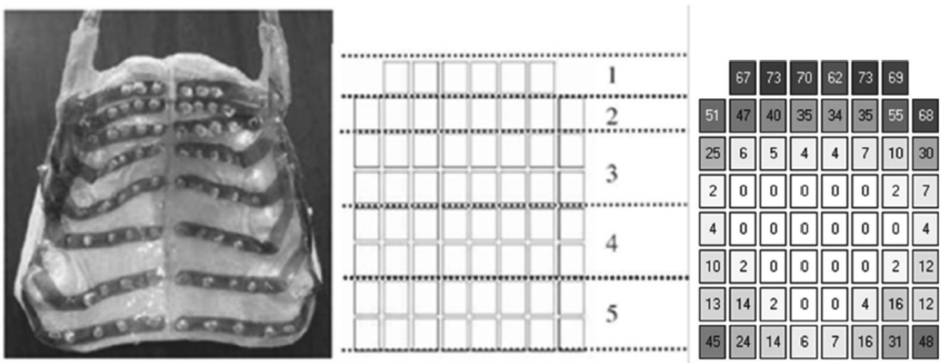

Figure 1. The Articulate palate and the articulatory zoning scheme: 1–dental, 2–alveolar, 3–postalveolar, 4–palatal, 5–velar zone (adapted from Wrench 2007). Picture on the left is the photography of the palate, schematic diagram in the middle shows the zoning scheme where each square represents one electrode, while the printout on the right shows one electropalatogram. Darker shades of grey indicate increased average tongue–to–palate contact, which is quantified by the percentage number in each square

Participants were instructed to read the target tongue twister *Kralj Karlo i kraljica Klara krali klarinet* (Eng. *King Karlo and queen Klara were stealing a clarinet)* in a speech rate they felt as comfortable.

2.4. Annotation and data analysis

Segmentation, annotation and data preparation were performed in Articulate Assistant (Wrench et al. 2002). If lateral approximant /l/ was produced correctly, segmentation and annotation were made according to articulatory signal and articulatory criteria (Horga and Liker 2016). In that case, the beginning of dentoalveolar lateral approximant /l/ was identified at the frame with full electrode activation across the second and/or first row of electrodes and the end of /l/ was identified when there was absence of full electrode activation across these rows. Also, incomplete lateral contact was required for a target to be identified as /l/. An example of the segmentation and annotation of one /l/ token in the word *Klara* is shown in Figure 2.

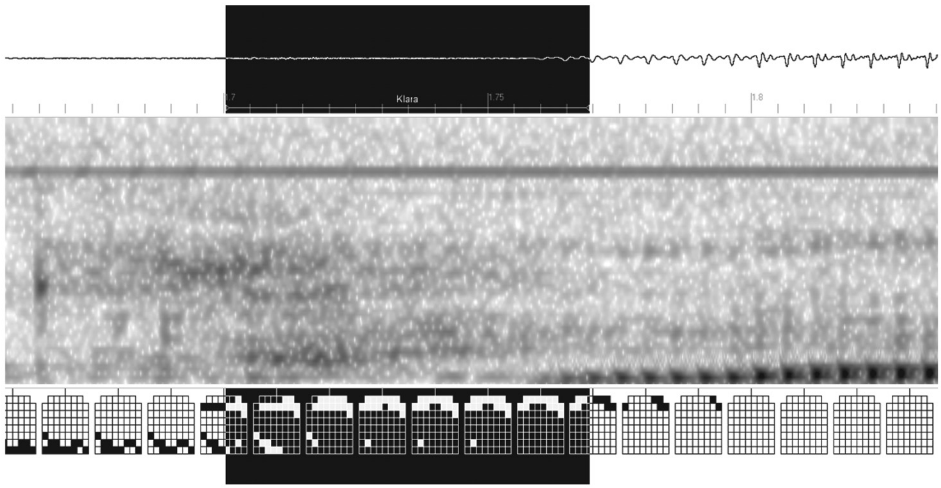

Figure 2. Illustration of annotation criteria for the dentoalveolar lateral approximant /l/

If /l/ was produced incorrectly, annotation and segmentation of the produced sound were made according to acoustic signal and perceptual cues.

The segmentation and annotation of the alveolar trill /r/ was somewhat more complicated when compared with /l/. Electropalatographic cue for the alveolar trill was a short and incomplete electrode activation pattern in the second row of electrodes, accompanied with a complete lateral contact. Since electropalatographic criteria for the segmentation of trills are relatively unreliable due to low frame rate, acoustic cues were also observed during the annotation process. Acoustic cue for the occurrence of /r/ was a sudden and short decrease in the acoustic energy of harmonic sound, which is the consequence of the tip of the tongue touching the alveolar ridge. In the spectrogram, the beginning of /r/ was assigned to the point in which intensity of the formants frequencies of the previous vowel started to drop and the end was assigned to the point in which formants intensity of the next vowel become more prominent. Changes in periodicity at the points of transition between /r/ and surrounding sounds were also taken into account during the annotation process. An example of segmentation and annotation of one /r/ token in the word *Klara* is shown in Figure 3.

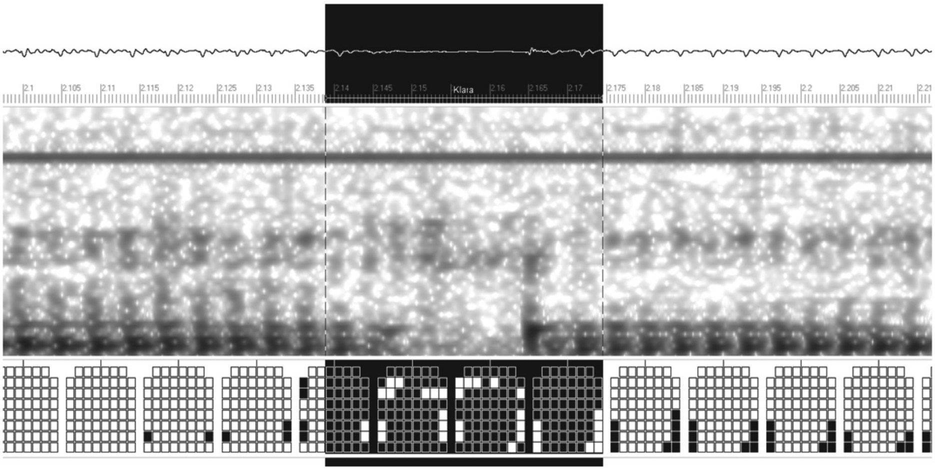

Figure 3. Illustration of annotation criteria for the alveolar trill /r/

If /r/ was produced incorrectly, annotation and segmentation of the produced sound were made according to acoustic signal and perceptual cues.

The annotation of every target consonant was discussed between the authors of this paper and each token was classified in one of the following categories:

P1A1 – articulatorily and perceptually correct production

P0A1 – articulatorily correct but perceptually incorrect production

P1A0 – articulatorily incorrect but perceptually correct production

P0A0 – articulatorily and perceptually incorrect production

The criteria for articulatory and perceptual correctness are based on a well–established fact that articulatory–auditory (and acoustic) cues in speech are in non–linear relation. Therefore, a certain speech sound can be auditory perceived as acceptable, while at the same time its articulatory configuration is not typical (e.g. regarding place or manner of articulation) and vice versa. Since all acceptable variations of speech sounds are not determined for Croatian speech, the decision on articulatory correctness is based on two criteria – place and manner. For the alveolar trill to be classified as correctly produced it had to be produced with an incomplete tongue–to–palate contact in the alveolar region, which is present on at least one electropalatogram within the annotation and is not uninterrupted (not continuous throughout the annotation). For the alveolar lateral approximant to be classified as articulatorily correctly produced, it had to show full electrode activation across one or two rows of the alveolar region and incomplete contact in the lateral regions. In the case of sound substitution or deletion, substituted or deleted target consonant was assigned as in (4).

Eight electropalatographic parameters were included in the analysis. The amount of tongue–to–palate contact in each of the eight rows of electrodes was calculated as a ratio of the total number of electrodes in each row and the number of activated electrodes. The result was an index ranging in value between 0 and 1, where the higher number indicates increased tongue–to–palate contact at a particular row of electrodes. These eight parameters were calculated at temporal midpoint of each annotation. The center of gravity measure (CoG) was also analyzed, but it did not prove to be sensitive enough for the purposes of this investigation, since places of articulation of /l/ and /r/ were relatively similar. The CoG measure is an EPG index widely used for quantifying the location of the highest concentration of the contacted electrodes along the front–back axis of the palate (Hardcastle et al. 1991), however, it can be insufficiently sensitive for the differentiation between very similar productions with low amount of contact (e.g. Gibbon and Nicolaidis 1999; Liker 2018).

MS Excel was used for statistical analysis and data visualization. One–way ANOVA (single factor ANOVA) was used to test the difference in average EPG contact in each row of electrodes between different production categories of /l/ and /r/. For the lateral approximant /l/ the difference was tested between three categories (P1A1, P1A0 and P0A0), while for the trill /r/ the difference was tested between four conditions (P1A1, P1A0, P0A1 and P0A0). Alpha was set to 0,05.

3. Results

There were 188 tokens of dentoalveolar lateral approximant /l/ in total. P1A1 productions occurred 138 times (73%), P1A0 occurred 31 times (17%), while P0A0 occurred 18 times (10%). There were no productions of /l/ which were classified as articulatorily correct but perceptually incorrect (P0A1). Figure 4 shows electrode activation in each row of electrodes averaged for each production class across all speakers and all productions. P1A1 productions are correct productions of /l/ and the data show typical EPG configuration – maximum contact in the first two rows of electrodes, i.e. at the dentoalveolar place of articulation, very few contacts in the mid palate due to laterality and slightly increased contact at the back of the palate in the velar area due to lateral bracing of the tongue. The P0A0 production shows low amount of contact at the place of articulation and the place of articulation is reduced to the alveolar area. This resembles /r/ productions, because trills have very small amount of contact at the place of articulation and /r/ is alveolar in Croatian, as opposed to /l/ which is dentoalveolar. This class of productions is also produced with increased contact in the mid–palate, possibly indicating reduced laterality. P1A0 production falls somewhere between these two extremes. The differences between the three different production categories are not statistically significant ($F(2, 21) = 0{,}782$, $p = 0{,}47$)).

Figure 5 shows average electropalatograms which support these data.

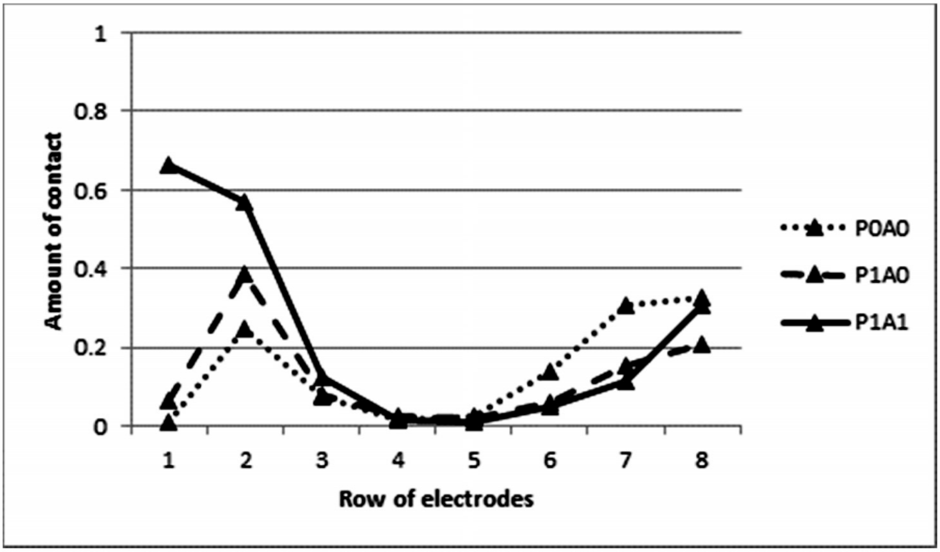

Figure 4. EPG contact in each row of electrodes in P0A0, P1A0 and P1A1 conditions averaged across all repetitions of /l/ and all speakers

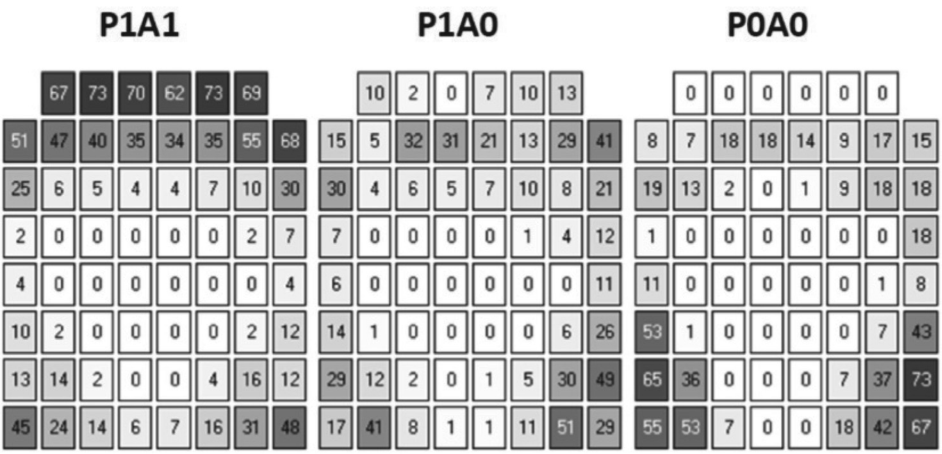

Figure 5. Average electropalatograms in P0A0, P1A0 and P1A1 conditions averaged across all repetitions of /l/ and all speakers. Shades of grey indicate average amount of contact at a particular electrode and the number indicates the average percentage of contact at a particular electrode

Between–speaker variation was considerable. Some speakers produced no speech errors and all productions were classified as P1A1 (F2), some speakers produced only P1A0 errors (F3) and some only P0A0 errors (F4, F7, F10), but the majority of speakers produced all three classes of productions (F1, F5, F6, F8, F9).

Alveolar trill /r/ was produced 221 times and all four classes of productions were observed. P1A1 was produced 163 times (74%), P0A1 four times (2%), P1A0 34 times (15%) and P0A0 20 times (9%). Figure 6 shows electrode activation in each row of electrodes averaged for each production class across all speakers and all productions of /r/. The data show that all four types of productions have similar contact patterns. P1A1 productions are correct productions of /r/ and the data show typical EPG configuration – maximum contact in the second row of electrodes, which is the location of alveolar placement, very few contacts in the fourth row of electrodes and increased contact in the palatal and velar areas due to strong lateral bracing of the tongue, required for successful apical trill production. It is also noticeable that there is a relatively small amount of contact at the alveolar place of articulation, because a trill is produced without full electrode activation across the whole row. The P0A1 and P1A0 have almost identical tongue–to–palate contact patterns, while P0A0 productions are noticeably different. These tokens are produced with similar amount of contact at the dental and the alveolar articulatory zone, i.e. dentoalveolar place of articulation, with posterior EPG contact decreased when compared with other three classes of productions. Thus, P0A0 r–productions have contact–patterns somewhat similar to correct l–production. Differences between the four production categories are not statistically significant (F(3,29) = 0,151, p = 0,927). Average productions are illustrated in Figure 7.

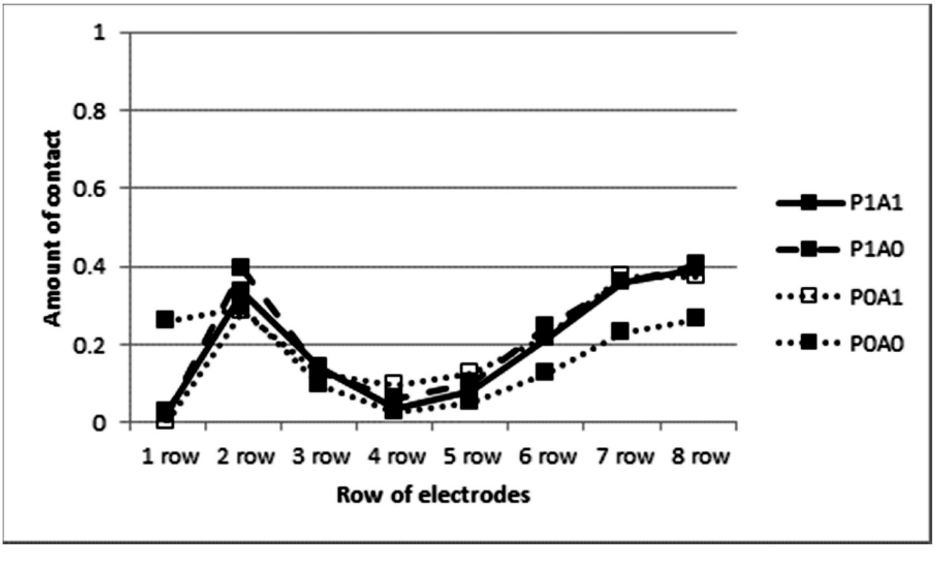

Figure 6. EPG contact in each row of electrodes in P0A0, P1A0, P0A1 and P1A1 conditions averaged across all repetitions of /r/ and all speakers

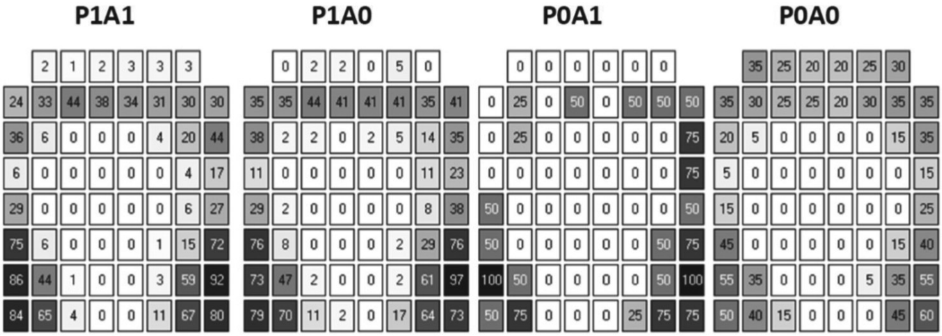

Figure 7. Average electropalatograms in P0A0, P1A0, P0A1 and P1A1 conditions averaged across all repetitions of /r/ and all speakers. Shades of grey indicate average amount of contact at a particular electrode and the number indicates the average percentage of contact at a particular electrode

All four classes of r–production were not found in all speakers. The majority of speakers had three classes of productions – P1A1, P0A1 and P0A0 (F1, F4, F5, F7), followed by speakers producing all four classes (F3, F6, F8), and only three speakers with different combinations of two classes of productions – P1A1 and P1A0 (F2 and F9); P1A1 and P0A0 (F10).

4. Discussion and conclusion

The results show that the majority of target consonant productions, 73% for /l/ and 74% for /r/, were classified as P1A1, i.e. they were articulatorily correctly produced and they were auditory perceived as typical for Croatian.

There are no tokens of /l/ and very few of /r/ (2%) classified as P0A1. This result shows that articulatory performance and configuration of a vocal tract typical for a specific speech segment will be perceived mostly as that speech segment. Contrary, a relatively high percentage of P1A0 tokens (17% for /l/ and 15% for /r/) shows that untypical articulatory realization of a specific speech segment will not be always perceived as an error. This result supports the claims that auditory perception is not sensitive enough to detect all deviations and subtleties of speech production processes (Bond 2008; Cutler 1981; Ferber 1991), and certain instrumental techniques could improve speech error analysis in terms of reliability and precision. This suggests that conclusions about phoneme as the smallest unit of speech (Fromkin 1971; Shattuck–Hufnagel and Klatt 1979; Stemberger 1982) based on the analysis of the traditional speech error corpora collected by listening and phonetic transcription could benefit from detailed articulatory data provided by the instrumental physiological techniques such as EPG. It should be noted that it does not mean that phoneme substitutions, i.e. categorical errors, are not possible. Moreover, P0A0 tokens of both target consonants can be described exactly as categorical substitution of one target consonant to another. As it is shown in Figure 4, P0A0 /l/ tokens have EPG contact typical for /r/, which indicates that this category represents segmental / categorical errors. Similarly, P0A0 /r/ tokens have pattern of electrode activation very similar to production of typical /l/ (Figure 6). However, beside categorical errors, some articulatorily erroneous tokens could be described as gradual errors, e.g. P1A0 /l/ tokens. Figures 4 and 5 illustrate this claim. Figure 4 shows that P1A0 tokens have EPG contact patterns typical for both target consonants: negligible contact in dental region (first electrode) and low amount of contact in alveolar region (second electrode), which is typical for /r/, but also very few contacts in the mid palate, which is typical for /l/, due to its lateral articulation. Average electropalatograms (Figure 5) additionally illustrate this gradual transfer (P1A0) from one phoneme category to another: from typical /l/ production (P1A1), to typical /r/ production (P0A0). P1A0 category in Figure 5 also shows a laterality level somewhere between P1A1 and P0A0 /l/ tokens, which is also the confirmation of a gradual change. However, it should be noted that the difference in EPG contact patterns between the four categories of /l/ and /r/ tokens is not statistically significant.

All categories of /r/ tokens, with the exception of P0A0, show very similar and highly variable pattern of electrode activation. Increased EPG placement variability in the production of the alveolar trill is expected (e.g. Díaz–Campos 2008; Recasens and Pallarès 1999) and can be at least partly explained by the dynamics of its production (i.e. repeated tongue to palate contacts at the place of articulation are caused by aerodynamic constraints and not by controlled articulatory movement of the tip of the tongue) and by a low apico–predorsal coupling (Recasens and Pallarès 1999). On the other hand, trill productions increase articulatory constraints placed on the tongue dorsum (Recasens and Espinosa 2009; Recasens and Pallares 1999; Recasens et al. 1997), which results in relatively low EPG contact variability and high tongue body constraint at the posterior portion of the palate (see Figure 7). When these expected patterns of articulatory gestures are not produced, as in P0A0 in Figure 7, and when this is combined with full electrode activation across dentoalveolar region, it is only then that the deviation in the articulatory gesture is perceived as an error.

/l/ tokens provide even more evidence of gradual errors, most notably P1A0 patterns, which represent the coproduction of the unintended and the target articulatory gesture in the realization of a single speech segment. This finding is also compatible with the results of some previously published articulatory and acoustic investigations of errors (Frisch and Wright 2002; Goldrick and Blumstein 2006; Goldstein et al. 2007; McMillan and Corley 2010; Mowrey and MacKay 1990; Pouplier 2007, 2008). Furthermore, the existence of gradual errors does not support those models of phonological encoding which predict that only whole phoneme could be allotted to the specific slot in the string of phonological units. Rather, one slot could be filled with the portions of two competing segments, and these portions could be articulatory gestures. Those studies, supported by the data from the present investigation, prompt us to revise the models of speech production which are based on the analysis of speech error corpora collected by auditory analysis and transcription in light of the findings obtained by instrumental laboratory investigations. It is beyond the scope of this paper to define whether the source of these coproductions is at the higher symbolic level of phonological planning, as it is proposed by models of cascading activation (Goldrick and Blumstein 2006; McMillan et al. 2010; Rapp and Goldrick 2000), or in the dynamical relationship between vocal tract gestures, as it is suggested by the gestural approach to speech production (Goldstein et al. 2007; Pouplier 2007; Pouplier and Goldstein 2010).

This investigation has its limitations, which prevent us from drawing definitive conclusions based on the data presented here. Although sample sizes of five to ten speakers are quite common in instrumental laboratory studies, this experiment should be replicated on other groups of speakers and more diverse speech samples before general conclusions are made. Furthermore, EPG is the only instrumental laboratory technique which allows us to analyze one of the most important aspects of articulation (i.e. tongue–to–palate contact patterns). However, EPG does not provide information on parts of the tongue which are not in contact with the palate (Gibbon and Nicolaidis 1999; Hardcastle et al. 1991). It would be beneficial to combine EPG with other instrumental laboratory techniques such as ultrasound tongue imaging (UTI) in future studies. Finally, our investigation was based on the temporal midpoint of the investigated sound. Tongue–to–palate dynamics could reveal interesting coarticulatory processes connected with speech errors, but that is a completely separate topic for a new investigation.

Acknowledgements

This investigation is a part of a long–term research project *Coarticulation in Croatian speech: an instrumental study* (CROCO), which is funded by the Croatian Science Foundation (HRZZ, grant number: IP–2016–065367). The authors are also grateful to anonymous reviewers, whose comments greatly improved the earlier version of the manuscript.

References

Boomer, Donald S., and John D. M. Laver (1968). Slips of the Tongue. *British Journal of Disorders of Communication* 3(1): 2–12, https://doi.org/10.3109/13682826809011435

Bond, Z. S. (2008). Slips of the ear. Pisoni, David B. and Robert E. Remez, eds. *The Handbook of Speech Perception*. Malden: Blackwell Publishing, 290–310, https://doi.org/10.1002/9780470757024.ch12

Browman, Catherine P., and Louis Goldstein (1992). Articulatory Phonology: An Overview. *Phonetica* 49 (3–4): 155–180, https://doi.org/10.1159/000261913

Cutler, Anne (1981). The reliability of speech error data. *Linguistics* 19(8): 561–582

Dell, Gary. S. (1986). A spreading–activation theory of retrieval in sentence production. *Psychological review* 93(3): 283–321, https://doi.org/10.1037/0033–295X.93.3.283

Dell, Gary S., and Peter A. Reich (1980). Slips of the tongue: The facts and a stratificational model. Copeland, James, and Philip Davis, eds. *Papers in cognitive–stratificational linguistics: Rice University studies*, 66(2), 19–34.

Díaz–Campos, Manuel (2008). Variable Production of the Trill in Spontaneous Speech: Sociolinguistic Implications. Colantoni, Laura and Jeffrey Steele, eds. *Selected Proceedings of the 3rd Conference on Laboratory Approaches to Spanish Phonology*. Somerville, MA: Cascadilla Proceedings Project, 47–58

Ferber, Rosa (1991). Slip of the tongue or slip of the ear? On the perception and transcription of naturalistic slips of the tongue. *Journal of Psycholinguistic Research* 20(2): 105–122

Frisch, Stefan A., and Richard Wright (2002). The phonetics of phonological speech errors: An acoustic analysis of slips of the tongue. *Journal of Phonetics* 30(2): 139–162, https://doi.org/10.1006/jpho.2002.0176

Fromkin, Victoria (1971). The non–anomalous nature of anomalous utterances. *Language* 47: 27–52, https://doi.org/10.2307/412187

Fromkin, Victoria A. (1973). Introduction. Fromkin, Victoria A., ed. *Speech errors as linguistic evidence*. Hague/Paris: Mouton, 11–45, https://doi.org/10.1515/9783110888423.11

García–Albea, José E., Susana del Viso and José M. Igoa (1989). Movement errors and levels of processing in sentence production. *Journal of Psycholinguistic Research* 18: 145–161, https://doi.org/10.1007/BF01069052

Gibbon, Fiona, and Katerina Nicolaidis (1999). Palatography. Hardcastle, William J., and Nigel Hewlet, eds. *Coarticulation: theory, data and techniques*. Cambridge: CUP, 229–245

Goldrick, Matthew, and Sheila E. E. Blumstein (2006). Cascading activation from phonological planning to articulatory processes: Evidence from tongue twisters. *Language and Cognitive Processes* 21(6): 649–683, https://doi.org/10.1080/01690960500181332

Goldrick, Matthew, Joseph Keshet, Erin Gustafson, Jordana Heller, and Jeremy Needle (2016). Automatic analysis of slips of the tongue: Insights into the cognitive architecture of speech production. *Cognition* 149: 31–39, https://doi.org/10.1016/j.cognition.2016.01.002

Goldstein, Louis, Marianne Pouplier, Larissa Chen, Elliot Saltzman, and Dani Byrd (2007). Dynamic action units slip in speech production errors. *Cognition* 103(3): 386–412, https://doi.org/10.1016/j.cognition.2006.05.010

Hardcastle, William J., Fiona Gibbon, and Katerina Nicolaidis (1991). EPG data reduction methods and their implications for studies of lingual coarticulation. *Journal of Phonetics* 19: 251–266

Horga, Damir, and Marko Liker (2016). *Artikulacijska fonetika: anatomija i fiziologija izgovora*. Zagreb: Ibis grafika

Kendall, Diane, Lisa Edmonds, Anine Van Zyl, Inge Odendaal, Molly Stein, and Anita van der Merwe (2015). What can speech production errors tell us about cross–linguistic processing in bilingual aphasia? Evidence from four English/Afrikaans bilingual individuals with aphasia. *The South African Journal of Communication Disorders* 62(1): 1–10, https://doi.org/10.4102/sajcd.v62i1.111

Levelt, Willem J. M. (1989). *Speaking. From Intention to Articulation*. Cambridge MA: MIT Press.

Levitt, Andrea G., and Alice F. Healy (1985). The roles of phoneme frequency, similarity, and availability in the experimental elicitation of speech errors. *Journal of Memory and Language* 24(6): 717–733, https://doi.org/10.1016/0749–596X(85)90055–5

Liker, Marko (2018). Electropalatographic analysis of vowels in quasi–spontaneous speech: a preliminary investigation. *Phonetician* 115: 5–22

MacKay, Donald G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia* 8(3): 323–350, https://doi.org/10.1016/0028–3932(70)90078–3

McMillan, Corey T., and Martin Corley (2010). Cascading influences on the production of speech: Evidence from articulation. *Cognition* 117(3): 243–260, https://doi.org/10.1016/j.cognition.2010.08.019

Mowrey, Richard A., and Ian R. A. MacKay (1990). Phonological primitives: Electromyographic speech error evidence. *The Journal of the Acoustical Society of America* 88(3): 1299–1312, https://doi.org/10.1121/1.399706

Nooteboom, S. G. (1969). The tongue slips into patterns. Fromkin Victoria A., ed. *Speech Errors as Linguistic Evidence*. Hague/Paris: Mouton, 144–156, https://doi.org/10.1515/9783110888423.144

Pouplier, Marianne (2007). Tongue kinematics during utterances elicited with the SLIP technique. *Language and Speech* 50(3): 311–341, https://doi.org/10.1177/00238309070500030201

Pouplier, Marianne (2008). The role of a coda consonant as error trigger in repetition tasks. *Journal of Phonetics* 36(1): 114–140, https://doi.org/10.1016/j.wocn.2007.01.002

Pouplier, Marianne, and Louis Goldstein (2005). Asymmetries in the perception of speech production errors, *Journal of Phonetics* 33 (1): 47–75, https://doi.org/10.1016/j.wocn.2004.04.001

Pouplier, Marianne, and Louis Goldstein (2010). Intention in articulation: Articulatory timing in alternating consonant sequences and its implications for models of speech production. *Language and Cognitive Processes* 25(5): 616–649, https://doi.org/10.1080/01690960903395380

Rapp, Brenda, and Matthew Goldrick (2000). Discreteness and interactivity in spoken word production. *Psychological Review* 107 (3): 460–499, https://doi.org/10.1037/0033–295X.107.3.460

Recasens, Daniel, and Maria Dolors Pallarès (1999). A study of /ɾ/ and /r/ in the light of the “DAC” coarticulation model. *Journal of Phonetics* 27: 143–169, https://doi.org/10.1006/jpho.1999.0092

Recasens, Daniel, Maria Dolors Pallarès, and Jordi Fontdevila (1997). A model of lingual coarticulation based on articulatory constraints. *Journal of the Acoustical Society of America* 102 (1): 544–561, https://doi.org/10.1121/1.419727

Recasens, Daniel, and Aina Espinosa (2009). An articulatory investigation of lingual coarticulatory resistance and aggressiveness for consonants and vowels in Catalan. *Journal of the Acoustical Society of America* 125(4): 2288–2298, https://doi.org/10.1121/1.3089222

Shattuck–Hufnagel, Stefanie (1979). Speech errors as evidence for a serial–ordering mechanism in sentence production. Cooper, William E., and Edward C. T. Walker, eds. *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, N.J.: Lawrence Erlbaum, 295–342

Shattuck–Hufnagel, Stefanie (1983). Sublexical units and suprasegmental structure in speech production planning. MacNeilage, Peter F., ed. *The production of speech*. New York: Springer–Verlag, https://doi.org/10.1007/978–1–4613–8202–7_6

Shattuck–Hufnagel, Stefanie (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook* 3: 117–149

Shattuck–Hufnagel, Stefanie, and Dennis H. Klatt (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior* 18(1): 41–55, https://doi.org/10.1016/S0022–5371(79)90554–1

Stemberger, Joseph Paul (1982). The nature of segments in the lexicon: Evidence from speech errors. *Lingua* 56(3–4): 235–259, https://doi.org/10.1016/0024–3841(82)90012–2

Wells–Jensen, Sheri (2007). A Cross–Linguistic Speech Error Investigation of Functional Complexity. *Journal of Psycholinguistic Research* 36(2): 107–157, https://doi.org/10.1007/s10936–006–9036–5

Wilshire Carolyn E. (1999). The “Tongue Twister” Paradigm as a Technique for Studying Phonological Encoding. *Language and Speech* 42(1): 57–82, https://doi.org/10.1177/00238309990420010301

Wrench, Alan A. (2007). Advances in EPG palate design. *Advances in Speech–Language Pathology* 9(1): 3–12, https://doi.org/10.1080/14417040601123676

Wrench Alan A., Fiona E. Gibbon, Alison M. McNeill, and Sara E. Wood (2002). An EPG therapy protocol for remediation and assessment of articulation disorders. Hansen, John H.L., and Brian L. Pellom, eds. *Proceedings of ICSLP–2002*, 965–968.

Govorne pogreške i artikulacijske geste: elektropalatografsko istraživanje

Jedan od glavnih zaključaka tradicionalnih istraživanja govornih pogrešaka, koja se temelje na slušnom zamjećivanju pogrešaka u govoru i njihovu transkribiranju, je da je glasnik najmanja jedinica govora zahvaćena pogreškom. Ipak, istraživači takva pristupa nisu u potpunosti odbacili mogućnost da na neke pogreške, npr. zamjene jednog fonema drugim, ne utječu i jedinice manje od segmenta, npr. razlikovna obilježja. Suvremene studije, koje govorne pogreške analiziraju s pomoću nekih od instrumentalnih metoda za istraživanje govora, npr. elektropalatografijom, pokazuju da se govorne pogreške mogu dogoditi i na razini nižoj od jednog glasnika – tj. mogu zahvatiti artikulacijsku gestu. Takve se pogreške manifestiraju kao istodobna produkcija gesti dvaju glasnika – ciljanoga i uljeza. S obzirom da takve pogreške predstavljaju stupnjeviti pomak od jednog segmenta ka drugomu, za razliku od kategorijalnih pogrešaka, nazivaju se gradacijskim pogreškama. Cilj je ovoga rada provjeriti pojavu gradacijskih pogrešaka u izgovoru, koristeći elektropalatografsku metodu. Analiza je obuhvatila dva ciljana suglasnika – /l/ i /r/, koje je izgovorilo deset izvornih govornica hrvatskoga jezika, ponavljajući brzalicu *Kralj Karlo i kraljica Klara krali klarinet* uobičajenim tempom. Svaka je pojavnica ciljanih suglasnika svrstana u jednu od četiriju kategorija: (1) perceptivno i artikulacijski i ispravan izgovor (P1A1); (2) artikulacijski ispravan, ali perceptivno neispravan izgovor (P0A1); (3) perceptivno ispravan, ali artikulacijski neispravan izgovor (P1A0) i (4) perceptivno i artikulacijski neispravan izgovor (P0A0). Artikulacijska ispravnost određena je tipičnim elektropalatogramom ciljanoga glasnika, a perceptivna ispravnost time zvuči li ciljani glasnik kao tipično ostvarenje toga glasnika u hrvatskome jeziku. Analiza je pokazala prisutnost i gradacijskih i kategorijalnih pogrešaka. Time je ovo istraživanje potvrdilo rezultate nekih drugih instrumentalnih istraživanja, koje pokazuju kako tradicionalna metoda koja se temelji isključivo na slušanju i zapisivanju pogrešaka nije dovoljna da zamijeti sva odstupanja od govornoga plana. Ono također ukazuje na potrebu revidiranja postojećih modela fonološkog kodiranja u proizvodnji govora, uzimajući u obzir gradacijske pogreške.

Keywords: electropalatography (EPG), gradient speech errors, categorical speech errors, articulatory gestures

Ključne riječi: elektropalatografija, gradacijske govorne pogreške, kategorijalne govorne pogreške, artikulacijska gesta