Investigating the Status of a Rare Cross-Linguistic Contrast: The Case of Romanian Palatalized Postalveolars

Laura Spinu
CUNY Kingsborough Community College

Recommended Citation
https://academicworks.cuny.edu/kb_pubs/165
Investigating the status of a rare cross-linguistic contrast: the case of Romanian palatalized postalveolars

Laura Spinu

Department of Communications and Performing Arts,
City University of New York - Kingsborough Community College,
2001 Oriental Blvd., Room E313, Brooklyn, NY, 11235*

(Dated: January 17, 2018)

Abstract

This study examines a rare cross-linguistic contrast, that between plain and secondarily palatalized postalveolar fricatives, through (i) an acoustic analysis of the production of 31 Romanian speakers and (ii) a perception experiment with a different group of 31 native speakers. Evidence of acoustic separation between plain and palatalized forms was found for 27 of the subjects, suggesting that the contrast is produced by the majority. This is consistent with previous reports of native speakers, collected in 1961. These findings were supported by the results of the perceptual experiment, which showed that native speakers exhibit moderate sensitivity to this contrast. An examination of each of the two genders’ production separately suggests that a process of neutralization may be in progress, more strongly realized by males compared to females. Aside from documenting this phenomenon in Romanian, an explanation is sought for its longevity, and it is proposed that grammatical restructuring offers the best account for the observed facts.

* Keywords: postalveolar fricatives; palatalization; speech production; speech perception
I. INTRODUCTION

It has been noted that in phonemic inventories, postalveolars usually pattern with either plain or palatalized consonants but not both (Kochetov 2002). This has been attributed to the low salience of the secondary palatalization contrast at this place of articulation (Kochetov 2002), which may be related to gestural timing, specifically to the overlap or blending of the palatalization gesture with that of the primary place of articulation (Zsiga, 2000). More generally, as noted by Mester and Itô (1989), the realization of the palatalized form of any consonant at the coronal place of articulation tends to involve, other than the secondary gesture, additional articulatory changes compared to the plain form, e.g. a change in primary place of articulation and increased burst release duration in the case of stops, to the point where the contrast is effectively between a (plain) stop and a (palatalized) affricate. According to Mester and Itô (1989), the characterization “palatalized” is strictly speaking only accurate for noncoronals, i.e. labials and velars, as palatalization of coronals changes their primary place of articulation to palatal or alveopalatal. It has been suggested that this may be due to the difficulty of achieving simultaneous targets at the dental and palatal regions, leading to intermediate articulations between the two (Kochetov, 2002), or to phonologically-driven strategies for phonetic enhancement in order to preserve a perceptually vulnerable contrast (Spinu, Vogel & Bunnell, 2012). Either way, the perceptual consequence of these changes is increased salience of the plain-palatalized contrast.

The typologically rare contrast between plain and palatalized postalveolar fricatives ([ʃ] - [ʃ]) is present in Romanian. This contrast has low to moderate perceptual salience, as demonstrated by the results of experimental work with native speakers (Spinu, 2007, 2012). Despite these findings, an acoustic study of the stimuli employed in the perception experiments showed that the distinction between plain and palatalized consonants is not statistically significant in postalveolars, unlike other places of articulation (Spinu et al., 2012). These findings to some extent are consistent with previous typological descriptions attesting the rarity of the plain-palatalized contrast at this place (Kochetov, 2002), but also raise questions in light of the observed mismatch between perception and production. Specifically, can any evidence be found of the realization of the secondary palatalization contrast at this place of articulation by employing more refined acoustic methods? Similarly, would using different methodology uncover increased perceptual sensitivity to this contrast? Given
the results of a 1961 study suggesting that the plain-palatalized contrast at the postalveolar place may have been present at the time in the production of Romanian speakers (Șuteu, 1961), the answers to these questions have the potential to provide both a diachronic and synchronic perspective of secondary palatalization in Romanian fricatives at this place of articulation (and also compared to other places). This paper helps us gain more insight into the typology of secondary palatalization. Furthermore, it contributes new data from an understudied language to sociolinguistic theories of language change and adds to the body of work on phonological contrast neutralization.

II. BACKGROUND

A. Secondary palatalization

Secondary palatalization is present in languages from diverse families, such as Polish, Russian, Hungarian, Irish, Mandarin, Mongolian, Navajo, Isthmus Mixe, etc. (Bateman, 2007; Bhat, 1978). A survey of a random sample of 117 languages found secondary palatalization in 27% of them (Bateman 2007), suggesting that typologically it is not a rare phenomenon. In terms of phonological status, there are two main patterns, with secondary palatalization being either (1) distinctive, as in Russian, where consonants with secondary palatal articulations are part of the phonemic inventory, contrasting with plain ones e.g. [glup] stupid vs. [glupʰ] depth, or (2) non-distinctive, as in Japanese (Vance, 1987), where secondary palatalization is assumed to be a surface realization of underlying consonant-high front vowel or consonant-glide sequences, e.g. /kar-itaɪ/ [karʰitaɪ] shear (Kochetov & Alderete, 2011).

In terms of phonotactics, neutralization of the plain-palatalized contrast is encountered in final (coda) position, as well as in pre-consonantal position, more often with labials than with coronals (Kochetov 2002). As for articulation, secondarily palatalized consonants are characterized by fronting and raising of the tongue body towards the hard palate, timed with respect to the primary articulation. The timing of the primary and secondary gestures was found to vary with speaker and syllabic position (Kochetov 1998, 2002).

Acoustically, palatalized consonants are generally longer than the plain ones, with stops having a strident-like release, which causes lower F1 and higher F2 on neighboring vowels, as compared to plain consonants. The perception of the plain-palatalized contrast was found to
be influenced by the primary place of articulation, with the contrast being disfavored (less salient) at the labial place compared to [+anterior] coronal in experiments with Russian stops (Kavitskaya, 2006; Kochetov, 2002). In a more recent experiment with Romanian fricatives, however, the reverse pattern was found, with the plain-palatalized contrast in labiodentals and glottals/velars being more salient than in both [+anterior] and [-anterior] coronals (Spinu et al., 2012).

B. Previous findings on secondary palatalization in postalveolars

While true palatalized palatals appear to be unattested (Operstein, 2010) and are deemed impossible from both a phonological and articulatory perspective (Hall, 1997, but see Campbell, 1974, for a differing view), the situation with postalveolars is not categorical. Kochetov (2002) notes that in a given language, postalveolar segments usually pattern with either plain or palatalized consonants, but not both. However, some loanwords in Polish show palatalization of (retroflex) postalveolar fricatives before the high front vowel /i/ to palatalized laminal postalveolar fricatives, which contrast acoustically with alveolo-palatal fricatives (Zygis & Hamann, 2003). Moreover, it was reported that Livonian contrasts /ʃ/ and /f̩/, while Mordvin contrasts /c/ and /c̪/ despite the posited non-existence of true palatals (Campbell, 1974, cf. Van de Weijer, 2011).

Morphological palatalization affects all consonants in Isthmus Mixe, including the postalveolar fricative (Dieterman, 2008). An acoustic analysis found distinctions between the plain and palatalized postalveolar forms in duration, spectral peak, and formant transitions, with higher F2 and F3 for the palatalized consonants (Dieterman, 2008). Dieterman notes that, despite extensive research, she has not been able to find any languages besides the Oaxacan Mixe family in which the entire consonant inventory may be modified by secondary palatalization manifesting a morpheme. As we will see, while Romanian palatalization is morphologically-induced for the most part, it differs from Isthmus Mixe in that it applies to most but not all consonants in the inventory.

Russian is one of the few languages with a 4-way contrast involving palatalized sibilant fricatives, specifically: palatalized dental/alveolar /ʃ/, palatalized postalveolar (prepalatal) /f̩/, non-palatalized dental/alveolar /s/, and retroflex (apical postalveolar) /ɕ/ (Timberlake, 2004). A recent acoustic study (Kochetov, 2017) found that the palatalized versus non-
palatalized contrast was distinguished by F1 and (especially) F2 at the onset, the midpoint, and in some cases at the offset of the following vowel. Fricative duration only marginally distinguished /β/, commonly described as a geminate, from the other consonants. The author’s conclusion is that Russian voiceless sibilant fricatives are robustly distinguished by spectral differences in this language.

C. Secondary palatalization in Romanian

Not found elsewhere in the Romance family, contrastive secondary palatalization in Romanian occurs only in word-final position, as it is commonly associated with (but not restricted to) the presence of two affixes associated with an underlying [-i]: the plural of certain nouns and adjectives (1a) and the second person singular in the present indicative of verbs (1b). (1c) shows an example of the contrast in postalveolars. With regard to the phonological status of secondary palatalization in this language, the widespread view is that an underlying word-final /i/ triggers palatalization on the preceding consonant and is then deleted (Chitoran, 2002), resulting in a surface contrast between plain and palatalized consonants. Thus, secondary palatalization in Romanian is not considered phonemic as it is in Russian.

(1)

a. [domn] gentleman – /domn+i/ → [domnʲ] gentlemen
b. [sar] I jump – /sar+i/ → [sarʲ] you jump
c. [arkaʃ] archer – /arkaʃ+i/ → [arkaʃʲ] archers

Tables I and II provide the phonemic inventory of Romanian consonants and the phonetic inventory of palatalized consonants, respectively.

In a 1961 study (Șuteu, 1961), questionnaires regarding the pronunciation of Romanian words were sent out by mail to 920 informants of various professional backgrounds (e.g. clerks, physicians, engineers, lawyers, priests, and artists) but crucially no linguistically-trained individuals (e.g. teachers, professors, scientists, writers, or journalists). All informants were living in Bucharest and never received additional reminders or requests to complete and return the questionnaires so as to avoid the elicitation of superficial answers that might have been provided as a result of even the mildest pressure (p. 295). Most in-
formants were college graduates. The author describes this group as homogeneous in terms of socioeconomic and educational status and speaking the standard form of the language. 314 questionnaires were completed and returned to the author. Each questionnaire included two parts comprising a total of 88 demographic and language use questions, most of which could be answered by a simple ‘yes’ or ‘no’. One question in particular (Question II.7) is of relevance to the current study because it addresses the pronunciation of a word ending in a postalveolar fricative: “Do you pronounce the singular and plural of the word ‘moș’ (old man) in the same way?” This word, which is pronounced [moʃ], ends in a postalveolar fricative in its singular form and takes on the suffix ‘-i’ in its plural form. The spelling of the plural form is ‘moși’. Of 309 answers received, 94.4% stated that they did not pronounce the two forms in the same way and, according to the author, many specified that they produced the plural form with a short or weak i-sound at the end. Because this study is based on self-described pronunciation without acoustic analysis the results are difficult to interpret. While these reports are inconsistent with Schane’s (1971) observation that a depalatalization process applies to palatal consonants in Romanian (i.e. [ʃ], [ʒ], [tʃ]) such that they do not surface with secondary palatalization, there is a possibility that the participants in the 1961 study were influenced by orthography and morphological factors, and their perception of how they pronounce these words is not accurate.

Acoustically, Spinu et al. (2012) found no significant differences between Romanian plain and palatalized postalveolars at the group level, though the descriptive statistics showed some differences in mean duration and spectral measures (i.e. cepstral coefficients 0 through 5). The lack of significant differences in some of these measures was also observed with other plain-palatalized pairs, most notably [z]-[ʒ]. A classification of palatalization performed using a linear discriminant analysis yielded 19.5% correct classification for the plain postalveolar and 96.7% for the palatalized one, leading to the conclusion that the two forms are for the most part indistinguishable from each other, both displaying the acoustic properties characteristic of secondary palatalization. These results were replicated in a more recent study (Spinu & Lilley, 2016) comparing the effectiveness of cepstral coefficients and spectral moments in the classification of fricatives. It was found that the measures that best capture the distinctions between plain and palatalized forms are as follows: for the cepstral method, coefficients 4, 3, and 2 (in that order), all extracted from the third and last temporal region of the fricative, and for the spectral method, spectral moments 3, 2, and 1, also extracted
from the last fricative region. In a multinomial logistic regression analysis with consonant
palatalization as the dependent variable and the top 3 measures from each set as continuous
explanatory variables, the correct classification rate was 85.17% for cepstral coefficients and
80.16% for spectral moments. These analyses, however, collapsed fricatives from four places
of articulation together. For the postalveolar segments, correct classification was close to
the chance level for the plain form, while the palatalized form was classified correctly over
80% of the time, similarly to Spinu et al. (2012).

With respect to perception, Spinu et al. (2012) examined plain and palatalized fricatives
from four places of articulation: labiodental, alveolar, postalveolar, and glottal/velar (re-
ferred to as ‘dorsal’ in this publication). The stimuli consisted of 640 fricatives produced
by 16 speakers, accompanied by twice as many fillers. They were presented in a context-
neutral carrier sentence, and the subjects performed a forced-choice task in which, following
the audio presentation of a sentence, they had to press a key corresponding to one of two
words displayed on the screen, one in the singular form (ending in a plain consonant, e.g.
‘pantof’) and one in the plural form (phonetically ending in a palatalized consonant, and
spelled with a final Ci sequence, according to the orthographic conventions of Romanian,
e.g. ‘pantofi’). Spinu & Lilley (2016) determined that the Romanian dorsal fricative is re-
alized at different places of articulation word-finally, depending on whether it is plain – in
which case it is realized as mostly velar, with 13% glottal realizations – or palatalized, in
which case it is realized exclusively as a palatal fricative. The difference in primary place
of articulation between the plain and palatalized dorsal may have arisen for the purpose
of contrast enhancement (Stevens & Keyser, 2010; Stevens, Keyser, & Kawasaki, 1986).
Regardless of the factors behind its occurrence, it is likely connected to the high perceptual
salience of this contrast. The results for accuracy, reaction time, and sensitivity (measured
as the d’ statistic) revealed the same patterns: (1) dorsals tended to be the most favorable
hosts for the palatalization contrast with higher accuracy and sensitivity values, and (2) the
postalveolar place of articulation was the least favorable, with significantly lower accuracy
and longer reaction times than all other places.
III. EXPERIMENTAL WORK

The findings reported above suggest that the status of the secondary palatalization contrast in Romanian postalveolars remains unclear to date. On the one hand, over 90% of Romanian speakers in Şuteu’s study believed that they were articulating the contrast in 1961. While their perception could be influenced by a number of factors and not necessarily reflect reality, it is nevertheless a fact that their intuition is at odds with the results of a subsequent acoustic study using cepstral coefficients (Spinu et al. 2012), in which no significant differences were found between plain and palatalized segments at the group level. On the other hand, the results of the 2012 study do reveal mean differences between plain and palatalized postalveolars in terms of duration (129.8 ms for the plain segment and 142.4 ms for the palatalized one) and some of the cepstral coefficients examined. Furthermore, in a perceptual experiment reported in the 2012 study, listeners’ sensitivity to the secondary palatalization contrast in postalveolars (measured as the d’ statistic) was .79. A d’ close to zero is interpreted as a lack of conscious access (Vermeiren & Cleeremans, 2012) while a value of 1 corresponds to 69% correct both for cases when the signal is present and when it is absent. This indicates that Romanian listeners display some sensitivity to this contrast.

Taken together, these findings suggest that, while elusive, acoustic differences between plain and palatalized Romanian postalveolars do exist and closer investigation is necessary in order to uncover them. The main goal for the current paper is thus to shed more light on the acoustic and perceptual properties of this contrast.

In the following sections, the properties of plain and palatalized postalveolar fricatives are examined from an acoustic perspective, as well as in a new perceptual experiment. The questions addressed are as follows:

1. Can any acoustic differentiation be found between plain and palatalized postalveolars?

2. Do Romanian speakers display higher sensitivity to this contrast in circumstances that would cause them to focus more on the information conveyed by the presence/absence of secondary palatalization?

The current study thus adds to previous work in a number of ways. First, the focus is now explicitly on the secondary palatalization contrast in postalveolars and to this purpose a series of new statistical analyses are conducted on previously collected data. Second, a closer
look will be taken for the first time at gender patterns as well as the individual patterns exhibited by speakers. Finally, the results of a new perceptual experiment, with different methodology than the 2012 paper, are reported.

A. Acoustic study

The data analyzed here are a subset of a larger corpus collected in Spinu (2010). The analysis of the entire set of Romanian fricatives and a more detailed account of the experimental procedure are reported in Spinu et al. (2012) and Spinu & Lilley (2016).

1. Hypothesis and additional research questions

This study’s main hypothesis is concerned with whether the contrast is present in the speech of Romanians. Based on the self-reports from the earlier study (Șuteu, 1961) and the perceptual findings of Spinu et al. (2012), the prediction is that (1) there are acoustic differences between plain and palatalized /ʃ/.

Assessing the strength of these differences and the extent to which individual speakers produce them will permit a tentative first picture of the neutralization status of this contrast. Regarding potential gender differences, sociolinguistic studies have long observed that women use more forms of standard language than men (Gregoire 2006). The opposite pattern was also found, however, in other studies describing women as leaders of language change (Shin, 2013) and the existence of what Labov calls the gender paradox: “women conform more closely than men to sociolinguistic norms that are overtly prescribed, but conform less than men when they are not” (Labov, 2001: 293; Nevalainen & Raumolin-Brunberg, 2003: 112). There is thus no clear expectation possible in this regard, but the question of which gender – if any – produces stronger acoustic cues to the secondary palatalization contrast in postalveolars will be explored in what follows.

2. Stimuli

To better understand the properties of postalveolar fricatives in Romanian, their behavior is compared to that of other plain-palatalized pairs. The stimuli collected for the original
corpus (Spinu, 2010) consisted of pairs of words ending in a plain and a palatalized version of five fricatives from four places of articulation. The originally collected consonants were:

**Labiodental:** /f/ and /v/

**Dental:** /z/

**Postalveolar:** /ʃ/ 

**Glottal/velar:** /h/~/x/ – Spinu & Lilley (2016) determined that the realization of this segment in word-final position is as either a velar fricative or glottal fricative in its plain form, and exclusively as a palatal fricative when palatalized. Previous descriptions varied between a glottal fricative (Chitoran, 2002), a glide (Ruhlen 1973), and a velar fricative, particularly before liquids or in word-final position (Mallinson, 1986; Sarlin, 2014). In this paper, /h/ is used for consistency with previous work and to reflect what is believed to be the phonemic representation of these sounds.

The stimuli consisted of pairs of real Romanian words that differed only in whether their final consonant was plain or palatalized. There were four pairs of words for each consonant. All words were disyllabic, with final stress, and were presented in a context-neutral carrier sentence, as shown in (2). The full set of stimuli is provided in Table III. Because of the restrictions on word shape and status, the vowel preceding the target consonant was not strictly controlled for, resulting in an unbalanced set of [e, a, o, u]. High front vowels were excluded because their presence might have resulted in coarticulatory effects (they are common triggers of primary and secondary palatalization in other languages). The uneven distribution of vowels poses certain limitations to the current study, discussed in more detail in subsection III.A.4. To avoid these limitations, the focus is on a subset of the consonants (/f/, /ʃ/ and /h/), specifically words in which the fricatives are preceded by either [a] or [o]. The words that are part of this subset are highlighted in boldface in Table III.

(2) Am sã aleg cuvântul [kodaf]/[kodafʃ] când voi fi gata. ‘I will choose the word [kodaf]/[kodafʃ] (‘slacker.M.sg./slacker.M.pl.’) when I am ready.’

The selection of a following velar stop may at first appear counterproductive to the purpose of the study, as this context tends to be impoverished for obtaining place information for consonants. However, a following vowel would not have been feasible due to the phenomenon of coda resyllabification typically encountered in Romanian and other Romance
languages, whereby a word-final consonant becomes an onset to a following vowel (in the case of word-final secondary palatalization, the palatal element might be resyllabified as a glide onset). The velar stop following the targets had previously been determined in an unpublished exploratory study not to induce anticipatory co-articulation effects on the target consonant (in contrast to a following labial).

The subjects produced twice as many fillers in addition to the target sentences. The fillers were also paired, showing inflections other than singular or plural, so as to distract the subjects from the target pairs. The items were presented three times to each subject, in a randomized order for each block. Thus, a full set of recordings contained 120 target items per subject: 5 consonants $\times$ 4 words per consonant $\times$ 2 forms per item (plain and palatalized) $\times$ 3 repetitions.

3. Subjects and procedure

Thirty-one native speakers of standard Romanian participated in the experiment. None of the participants had a history of speech or hearing disorders. Their linguistic and demographic background was recorded via questionnaires administered prior to testing. No participants were excluded from the study based on the info provided. They were recruited via email announcements sent to undergraduate groups at universities in Bucharest, Romania, with the help of faculty members. The subjects were 10 males and 21 females, ranging in age from 19 to 30, with an average age of 21.7 years. They were all tested individually in a quiet room in Bucharest and received compensation for their time. The stimuli were presented and recorded using the InvTool program (Yarrington, Gray, Pennington, Bunnell, Cornaglia, Lilley, Nagao & Polikoff, 2008). Each sentence was displayed on a computer screen (using Romanian orthography) and the subjects were instructed to read the sentence as naturally as possible. Before beginning the actual experiment, the subjects completed a practice session with 20 items.

4. Data Analysis

All of the subjects except one produced the full set of three repetitions; the remaining subject only produced two repetitions. Six items were rejected due to disfluencies, leaving
3,674 items for acoustic analysis. Each segment was acoustically analyzed to obtain the duration and average spectral properties expressed as the first six coefficients of the Bark cepstrum, which describes the amplitude and shape of the speech spectrum in terms of a set of compact orthogonal components (Bunnell, Polikoff & McNicholas, 2004).

Less common in the phonetics literature, perceptually weighted cepstral coefficients are routinely used in speech recognition (Deller et al., 1993; Rabiner & Schafer, 2011), providing useful descriptions particularly for vowels and obstruents. They have also been used in clinical phonetics studies, with better results than spectral moments in the classification of stop release bursts (Bunnell et al., 2004). In the acoustic phonetics literature, Ferragne & Pellegrino (2010) recommended cepstral coefficients for computing distances between vowels. Their method yielded excellent results in estimating the acoustic distance between 13 accents of the British Isles.

While vowels and stop release bursts are acoustically different from fricatives, more recent studies demonstrate that the advantages of cepstral coefficients also apply to these segments. Mel-frequency cepstral coefficients (MFCC) were used to classify voicing in fricatives in British English and European Portuguese (Jesus & Jackson, 2008). More recently, Kong, Mullangi & Kokkinakis (2014) obtained 85% correct classification for three places of articulation in fricatives using a set of 13 MFCCs. In Spinu & Lilley’s (2016) study of Romanian fricatives, cepstral coefficients provided high correct classification rates and outperformed spectral moments in the classification of place of articulation, voicing, palatalization, and gender. Alternatives include measures such as adjacent vowel formants, duration, center of gravity, and intensity of fricative noise (Kochetov 2017). In a recent comparison of such measures with cepstral coefficients 0-5 in the classification of plain and palatalized Russian voiceless sibilant fricatives (including [p]), the latter outperformed spectral measures across the board (Spinu, Kochetov & Lilley, ms.).

In the current study, the sequence of phonetic symbols comprising the transcription of each sentence was first aligned with the corresponding signal in the recordings. The initial automatic segmentation of the wavefiles was performed by maximizing the likelihood of a forced alignment of trained Hidden Markov models (HMMs) to the signal using the Viterbi algorithm (Viterbi, 1967). Each file was then visually inspected and alignment errors were corrected manually for the segments under investigation. Both the waveform and the wideband spectrogram of each token were used in verifying the segmentation. Fricative onset
was defined as the point at which high-frequency energy first appeared on the spectrogram, and/or the point at which the number of zero crossings rapidly increased (Jongman, Wayland & Wong, 2000). The end of each segment (plain or palatalized) was defined as the intensity minimum immediately preceding the silence during the closure of the following stop. Sample waveforms and spectrograms of one male and one female speaker’s productions are shown in Figure 1.

The durations of the target consonants and preceding vowels were determined, and then a series of feature vectors comprising six Bark cepstral coefficients (DC and the first five cosine terms - see Appendix A) were extracted from the segments of interest using overlapping Hamming analysis windows, 20 ms wide and spaced 10 ms apart. Next, Hidden Markov Models (HMMs) were used to divide the segments into three temporal regions of internally minimized variance. The goal was to maximize the differences between adjacent states and minimize the differences among feature vectors within the same state. HMMs thus partition the time-varying structure of segments into a series of piecewise approximately stationary regions, mimicking the dynamic nature of the segments. These separate temporal regions offer the possibility to examine the strength of the acoustic cues to palatalization progressively throughout each segment. The main reason for using HMM-defined regions is the focus on secondary palatalization. The degree of sequentiality of the secondary palatalization gesture was found to vary by speaker (Kochetov 2002), with some speakers producing the primary and secondary gestures almost simultaneously. Especially for segments with acoustic properties that are similar to those of secondary palatalization, such as postalveolars, averaging over entire segments may obscure the presence of secondary palatalization. Examining their acoustic properties region by region, when the regions are determined based on internal variance, makes it more likely that, if at all sequential, the acoustic consequences of the palatalization gesture might be captured.

To train the HMMs, the segments were first divided into regions of approximately equal duration, and the initial six means and six variances of those regions provided the initial parameters of the model states. Then the boundaries between the regions were recalculated so that the total likelihood of the data is maximized with reference to the current model parameters. The means and variances of the new regions were obtained. This process was repeated until no feature vectors were reassigned (Viterbi, 1967; Baum et al., 1970).

After dividing each target segment into regions, the means of the features over all of
FIG. 1. Male speaker: waveforms and spectrograms for plain and palatalized consonants in two vowel contexts. For each spectrogram, the first two productions contain the vowel [a] and the last two the vowel [o].
FIG. 1. Female speaker: waveforms and spectrograms for plain and palatalized consonants in two vowel contexts. For each spectrogram, the first two productions contain the vowel [a] and the last two the vowel [o].
the vectors in each region were calculated and used as input to the statistical analyses. This resulted in 36 different measures: 6 cepstral coefficients × 2 segments (fricative and preceding vowel) × 3 regions inside each segment. In the remainder of the paper, these measures will be labeled by composite names containing the specific coefficient used (e.g. C1), the segment from which it was extracted (C or V), and the region from which it was extracted (1, 2, or 3), as in C1.C.3, which refers to the first cepstral coefficient extracted from the third consonantal region. Two corpora of fricatives were constructed as follows: (1) all fricatives (3,674 total) – this used the entire fricative corpus, but the information from preceding vowels was excluded. This was due to the fact that the uneven distribution of the vowels preceding the fricatives introduced confounds that could not be controlled for. Even with the vocalic information excluded, there is a possibility that coarticulatory effects carried into the frication portion and the presence of the vowel confound thus interfered with the classification. To address this, (2) a smaller subset was selected, with fricatives from three places of articulation: labiodental, postalveolar, and glottal/velar. These were all voiceless and most importantly the vowels preceding them were balanced between [a] and [o]. There were 1,103 fricatives in this subset. In what follows, the two corpora are referred to as (1) ‘the ALL corpus’ and (2) ‘the fšt corpus’. It is the latter corpus that most figures and related discussions are based on.

Statistical analyses were carried out in the R environment (R Core Team, 2012) using the lme4 package (Bates, Maechler, Bolker & Walker, 2014) in order to find out how well duration and the cepstral coefficients extracted from different regions of the fricatives and preceding vowels predicted palatalization status. A binomial logistic regression with mixed effects was conducted on the ALL corpus to classify palatalization with 18 acoustic measures (6 coefficients × 3 fricative regions) as continuous explanatory variables and subject, word, and gender as random effects to control for the influence of different mean ratings associated with these variables. Duration and vocalic measurements were not included due to the fact that the vowels preceding the target consonants formed an unbalanced set as discussed in section III.A.2. To further avoid this vowel confound, a logistic mixed effects model was also conducted on the fšt corpus to classify palatalization with subject, word, gender, and vowel as random effects. The fšt corpus was balanced in terms of vowel distribution, with half of the target fricatives preceded by [o] and the other half by [a], so the 6 coefficients from the third vocalic region (adjacent to the fricative) were included together with duration and
the 18 consonantal measurements for a total of 25 predictors. Correlation analyses were performed between duration differences (in plain versus palatalized) and cepstral coefficient differences. Next, subsets of measures were used in additional logistic mixed effects models with the fijn dataset as follows: only measures from the first fricative region, only measures from the second fricative region, and only measures from the third fricative region, for a total of 6 predictors each time. These analyses were conducted in order to examine the discriminability between plain and palatalized forms progressively throughout the segment. Finally, one-way ANOVAs were conducted for each subject separately to see if each of the predictor variables (duration and cepstral coefficients from different regions) can reliably discriminate palatalization in postalveolars.

5. Results

In the binomial logistic mixed effects model for the entire (ALL) dataset, a significant effect on palatalization was found for 10 out of the 18 combinations of cepstral coefficients and temporal regions: C1.C.1 (p<.01), C2.C.3 (p<.001), C3.C.1 (p<.05), C3.C.2 (p<.05), C3.C.3 (p<.001), C4.C.1 (p<.01), C4.C.3 (p<.001), C5.C.1 (p<.01), C5.C.2 (p<.001), C5.C.3 (p<.001). The top 5 predictors (with the highest coefficients) were, in descending order, C4.C.3, C3.C.3, C3.C.2, C2.C.3 and C4.C.1. For the fijn dataset, a significant effect on palatalization was found for 7 out of the 25 measures: C0.C.1 (p<.05), C0.V.3 (p<.05), C2.C.3 (p<.05), C3.C.3 (p<.001), C4.C.3 (p<.001), C4.V.3 (p=.01) and C5.C.1 (p<.05).

Classification of palatalization was 90.96% accurate for the ALL dataset, and 90.3% accurate for the fijn model. For ALL, the correct classification rates by consonant ranged from 66.3% correct for plain [j] to 100% correct for both plain and palatalized /h/. For fijn, the lowest correct classification rate was 73.9% for plain [j], and the highest was 100% for both plain and palatalized /h/. Overall, the classification of palatalization in postalveolars was higher than chance (68.7% for the ALL set and 74.2% for the subset), indicating the existence of acoustic differences between the plain and palatalized forms.

The results for duration and the different cepstral coefficients were reported in Spinu et al. (2012), to which the reader is referred for more details. For duration, while palatalized consonants were always longer than the plain ones, the difference was only significant (p<.05)
in the case of /h/ and /v/. The results for the different cepstral coefficients were very similar, in that larger differences were noted between plain and palatalized labiodentals, and especially glottals/velars, while smaller differences (if any) were found between the plain and palatalized [+anterior] and [-anterior] coronals. For both duration and cepstral measures, postalveolars stood out as the segments with the least differentiation between the plain and palatalized form. In the current study, a series of correlation analyses were used to investigate the relationship between duration and the six cepstral coefficients. It was found that differences in duration between plain and palatalized forms correlate significantly with differences in C1 ($R^2=12.25\%$) and C3 ($R^2=16.81\%$). The findings suggest that subjects who produce smaller durational differences between plain and palatalized forms also tend to produce hypoarticulated speech more generally, in which acoustic cues related to the secondary palatalization contrast (and possibly other aspects of the speech signal) are also diminished.

Figure 2 displays three-dimensional plots of the top three predictors (that is, C4 and C3 from the last consonantal region and C4 from the last vocalic region) in the fh subset, for each consonant separately. The top three predictors were defined as the variables with the largest coefficients and a minimum required level of significance ($p \leq .01$). The separation of the tokens into relatively distinct areas is apparent, though more so for the labiodentals and glottals/velars than for the postalveolars.

Figure 3 shows the durations of the different regions identified via the procedure described in III.A.4 for plain and palatalized forms of labial (labiodental), postalveolar, and glottal/velar segments. While the regions of plain consonants appear very similar, more variability is noted for palatalized forms, with the third region being longer for the labiodental consonant (compared to the plain form) and much longer ($>120$ ms) for the glottal/velar. Thus, the presence of palatalization in labiodentals and glottals/velars causes a shift in the third region of the segment, which does not appear to be the case for postalveolars.

Figure 4 displays palatalization classification results for logistic mixed effects models trained either on all parameters (including vocalic information) or just parameters extracted from a single fricative region, for each gender separately. The glottal/velar classification rates are the highest throughout, suggesting that the difference in primary place of articulation
FIG. 2. Three-dimensional scatterplots of the 3 most informative acoustic measures for classification of palatalization. Each point represents a single fricative token in the dataset. The first part of each axis label stands for the coefficient (e.g. C4), the middle part stands for the segment from which it was extracted (C or V), and the last part stands for the region from which it was extracted (1, 2, or 3). E.g. C3.C.3 = the third cepstral coefficient extracted from the third region of the fricative. CON = consonant.
FIG. 3. Region duration (ms) for plain and palatalized forms at three places of articulation. Each segment was divided into three regions following the procedure described in Section III.A.4.

(plain = velar, palatalized = palatal) robustly encodes the secondary palatalization contrast throughout the segment. No differences are noted between the two genders. For the labiodentals, the final region yields the highest correct classification rates, which may be due to a more sequential realization of the secondary palatalization gesture in the case of this place of articulation, and there are hardly any differences between the two genders. Finally, postalveolars also show an increase in the last region but only for female productions. Unlike for labiodentals and glottals/velars, a clear difference can be seen in the postalveolar data between the classifications based only on the last region and those based on all fricative regions plus vocalic information, especially for the female data. This suggests that in the case of postalveolars, vocalic transitions contain important information pertaining to the secondary palatalization contrast.

Lastly, one-way ANOVAs were conducted for each subject separately to find out whether each of the 25 variables discriminates between plain and palatalized postalveolars. Because fewer items were included in these analyses (i.e. 12 tokens for each plain and palatalized segment, and 8 for the speaker who only completed two of the three blocks), the sample size is substantially reduced. Under these circumstances, near-significant values (between .05 and .1) were also considered. It was found that only one of the 31 subjects did not produce either significant or near-significant differences between plain and palatalized postalveolars in any of the measures examined. Three other subjects produced near-significant differences, but no significant ones. The remaining 27 subjects all produced significant differences in some of
FIG. 4. Palatalization classification results for models trained either on all 25 parameters (black columns) or just parameters extracted from a single fricative region (grey columns). The column labels show the gender (M or F) and the region (R1, R2, or R3). The label ALL is used for the entire parameter set. CON = consonant.

...the measures. For these subjects, the average number of measures with significant differences is 3.2 (out of 25 measures); one subject showed significant differences in 9 measures. There was a great deal of variability in which measures differed significantly between the two forms. The only measure for which neither significant nor near-significant differences were found in any of the subjects was c0 extracted from the third vocalic region.

6. Summary

The results of the analyses reported in previous subsections indicate that there is evidence of acoustic separation between plain and postalveolar Romanian fricatives, thus confirming the hypothesis from III.A.1, contrary to the findings of Spinu et al. (2012). The differences between the two studies are likely due to the refinement in the analysis methods employed: division into regions in order to capture the dynamic nature of the contrast, addition of vocalic measurements as predictors of palatalization, and the use of logistic mixed-effects models for classification. A secondary finding was that there are certain differences in classification rates when the productions of males and females are analyzed separately, but only for the postalveolar segments. These findings indicate that males may be more prone to hypoarticulating the secondary palatalization contrast at this place of articulation.
B. Perception study

The fact that acoustic differences were found between [ʃ] and [p] casts some doubts over the relatively low sensitivity results obtained in previous perceptual experiments. Not only is this contrast produced by native speakers of Romanian, it is also acquired, which suggests that it must be perceptually salient to some extent. There is a possibility that the set-up of previous experiments did not encourage listeners to focus on it and thus failed to reveal an existing perceptual pattern. In the perception experiment reported in Spinu et al. (2012), the target words were presented in a carrier sentence that was semantically and grammatically neutral with respect to their status as singular (no palatalization on the word-final consonant) or plural (with palatalization on the word-final consonant). It is possible that some of the perceptual patterns were obscured or diminished by the fact that the target items were presented in structures in which either a singular or plural interpretation would have yielded a grammatical structure. Hawkins (2010) argues that “while systematic phonetic variation in the spoken signal can strongly influence perception, it cannot do so unless the listener is able to relate the variation in sound patterns to different functions/meanings”. This is because linguistic information in the signal maps to more than just “low-level” abstract units (e.g. phonemes, features, or words), many of which are non-phonological (e.g. bound morphemes, function words, content words, auxiliary verbs, pronouns, etc.). No formal unit of linguistic structure is prior to any other during speech processing: “meaning (including function) is prior, and all potential units that allow meaning to be quickly understood are valuable” (Hawkins, 2010).

Given that normally there is rich agreement in Romanian (subject-predicate, nouns-determiners/modifiers, objects-pronominal clitics), the question arises whether the presence of morphological information in normal speech might lead listeners to pay more attention to the information conveyed by palatalization. The experiment described in the following section investigates the perception of the plain-palatalized contrast in the presence of morphological cues either consistent or conflicting with the information (i.e. grammatical number) encoded by the presence or absence of palatalization. It may be the case that, while subjects are able to hear the contrast between plain and palatalized postalveolars, they do not particularly focus on it unless a certain degree of functional load is involved. If this experimental set-up makes the presence or absence of palatalization more detectable, par-
particularly where mismatches are involved, subjects might demonstrate higher sensitivity to the secondary palatalization contrast in postalveolar fricatives compared to previous studies.

1. Hypothesis

The hypothesis formulated here is partially based on the phenomenon of “mismatch effect” or “mismatch cost” observed in neurolinguistic studies. Incongruent phonological or morphological cues were found to elicit larger brain responses than correct conditions (Leminen et al. 2016, Arbour 2012, Archibald & Joanisse 2011, Scharinger et al. 2010). For example, phonologically inconsistent words produced greater activation in medial frontal gyrus/anterior cingulate cortex (Binder et al., 2005; Tan et al., 2001). Behaviorally, incongruent stimuli were found to affect subjects’ responses in lexical decision tasks (Janke & Kolokonte 2015), result in processing disruptions such as longer reading times (Haskell & McDonald 2003, Pearlmutter et al. 1999, Marslen-Wilson & Warren 1995), and slower grammatical category decisions (Arciuli & Monaghan 2009). Most importantly, behavioral responses were more accurate and faster for phonological mismatches with different sounds than for phonological matches in Lee et al. (2012). Since the results from previous sections indicate that the contrast is acoustically present and produced by the majority of the speakers, the expectation is that conflicting morphophonological cues will result in enhanced perceptual sensitivity to the secondary palatalization contrast in postalveolar fricatives. Thus, a sensitivity value higher than previously obtained by Spinn et al. (2012), i.e. $d' = .79$, is expected.

2. Stimuli

Each of the target words appeared in four different types of constructions. The plain and palatalized consonants were placed in carriers that either matched or did not match them with regard to the information about grammatical number conveyed by the presence or absence of secondary palatalization. Four types of constructions were used, as follows:

**plain matched**: the target word is in the singular; the additional number cue is singular (e.g. [un koko] one rooster).
plain mismatched: the target word is in the singular; the additional number cue is plural, thus conflicting with the information provided by the absence of palatalization (e.g. [patru koko] *four rooster).

palatalized matched: the target word is in the plural; the morphological cue is in the plural (e.g. [patru koko] four roosters).

palatalized mismatched: the target word is in the plural; the additional cue is in the singular, in conflict with the information provided by the presence of palatalization in the target word (e.g. [un koko] *one roosters).

The construction of the stimuli for this experiment involved two steps. The first step consisted of recording the target words in grammatical sentences. The same words as in the production experiment described in section III.A.2 were used (Table III), for a total of 40 target items (5 segments × 2 palatalization conditions × 4 words). Twice as many fillers were included, for a total of 120 sentences. The set of stimuli was repeated 3 times, with short breaks in between the repetitions. The order of the sentences was randomized for each presentation.

Fifteen different speakers (3 males, 12 females, mean age = 20.6) produced the sentences for this experiment. These speakers were selected among a larger pool of potential speakers, n=31. These speakers were not the same as those who participated in the production experiment reported in section III.A, but they were drawn from the same demographic pool via the same methods. They were all native speakers of the standard dialect of Romanian. None of the subjects had any known speech or hearing disorders.

Each speaker went through a short practice session prior to recording the experimental items. The practice stimuli contained 20 sentences that were very similar to the ones used in the experiment. Each practice sentence was preceded by a prompt recorded by a male native speaker, to demonstrate the appropriate speaking rate to the speakers. The speakers were asked to replicate the speaking rate they had heard, to the extent possible. Audio prompts were not used in the recording of the actual test items. Both the practice session and the experimental session were administered in a quiet room in Bucharest, using a Sony laptop and a head-mounted microphone for recording. The software for presenting and recording the sentences was InvTool (Yarrington et al., 2008). No automatic manipulation of the sound files took place.
The second step consisted of splicing out the targets from the sentences in which they were recorded and using them to create the sentences to be used in the perception study. The alternative would have been to record all items directly. Since this experiment used both grammatical and ungrammatical stimuli, however, ungrammatical sentences might cause the speakers to (consciously or unconsciously) hyperarticulate or exhibit other effects such as unusual intonation due to the unfamiliarity with the stimuli. Because of this possibility, all sentences recorded were grammatical, and the ungrammatical items were constructed through cross-splicing of the target items from the original (morphologically appropriate) context into a morphologically mismatched context. Furthermore, to avoid the presence of splicing as a crucial difference between grammatical and ungrammatical sentences, the grammatical sentences were also constructed by splicing out the target items from the sentences in which they had been originally recorded and inserting them into appropriate contexts. Each target sentence contained exactly one morphological cue with regard to the number status (singular or plural) of the target words.

Only one repetition of each sentence per subject was used in the perception experiment. Consequently, the total number of target sentences was 1200 (5 segments $\times$ 4 words $\times$ 2 palatalization conditions $\times$ 2 contexts $\times$ 15 speakers). No fillers were used. Two lists of 600 stimuli were prepared, each containing half of the recordings. For each of the speakers, half of the targets were added to one list and the other half to the other list, such that the matched and mismatched condition for one word did not come from the same speaker.

3. Subjects and procedure

Thirty-one subjects (11 males, 20 females, mean age = 24.25) participated in this experiment, and were paid for their participation. None of the subjects reported having any current or past speech or hearing disorders. These subjects had not participated in any of the production or perception experiments described in previous sections.

In order to limit the duration of the experiment to approximately one hour, each subject heard only one of the two lists of stimuli. The subjects were sequentially assigned to one of the lists, e.g. Subject 1 to List 1, Subject 2 to List 2, Subject 3 to List 1, etc. The order of presentation of the stimuli was randomized for each subject. A short break was taken after the first 300 sentences.
The experiment was designed and administered using the E-Prime Software, and the subjects listened to the sentences over headphones. The task consisted of pressing a key on a keyboard to indicate whether a given sentence was perceived as being acceptable or unacceptable (i.e. likely or unlikely to have been uttered by a speaker of Romanian). To eliminate the possibility of interference with motor dominance, the response keys were counterbalanced.

A practice session was administered before the experimental block so as to familiarize the subjects with the task. The practice session contained 20 sentences, half matched and half mismatched.

4. Data analysis

The dependent variable sensitivity was computed for each subject for all five consonants, and the values were compared in a repeated-measures within-subjects ANOVA with the independent factors of consonant and palatalization. The statistical test for sensitivity is known as the d’ (d prime) statistic in the context of Signal Detection Theory (cf. Wickens 2002).

In the perceptual experiment reported here, bias was taken into account by using both the number of hits (how many times a signal was correctly identified; in our case, how many of the mismatched targets were perceived as mismatched), and the number of false alarms (that is, how many times a signal was incorrectly identified; in our case, how many of the matched targets were identified as being mismatched). The d’ scores by consonant for each subject were computed, and a repeated-measures ANOVA was conducted with sensitivity (d’ scores) as the dependent variable and consonant and palatalization as independent variables.

5. Results

Figure 5 shows the sensitivity (d’) scores for each consonant, broken down by palatalization status. The repeated-measures ANOVA showed that consonant identity had a significant effect on sensitivity, F(4, 27) = 82.59, p<.001. No main effect of palatalization or the interaction of consonant and palatalization was found. Pairwise comparisons with the Bonferroni adjustment for multiple comparisons show that listeners’ sensitivity to /f/
differed significantly from /z/ and /ʃ/, sensitivity to /v/ differed significantly from /ʃ/ and
/h/, sensitivity to /z/ differed significantly from /ʃ/ and /h/, sensitivity to /ʃ/ differed sig-
nificantly from all other places, and sensitivity to /h/ differed significantly from all places
but /f/.

Generally speaking, the larger the d’ value, the higher the sensitivity to a certain con-
trast. A d’ value of zero means that trials with the target cannot be reliably distinguished
from trials without the target, whereas a d’ of 4.65 indicates a nearly perfect ability to dis-
tinguish between trials that include the target and trials that do not include the target. The
latter appears to be the case of the glottal/velar fricative, which had the highest sensitivity
scores for both the plain and the palatalized forms (d’=4.51, averaged over the two). The
postalveolar /ʃ/ was the consonant to which the subjects were least sensitive, with a d’ score
of .93, averaged over the two categories (plain and palatalized). This score corresponds to
moderate sensitivity.
6. Summary

The results of the perception experiment were consistent with those of the production experiment. The analysis of sensitivity scores revealed the same general patterns: glottal/velars tended to be the most favorable hosts for the palatalization contrast, with the highest sensitivity values. The postalveolar place of articulation was, by contrast, the least favorable. The d’ value for this place was higher than previously obtained (Spinu et al., 2012), showing that the inclusion of matched and mismatched cues to the presence of palatalization causes listeners to pay more attention to it. Thus, the hypothesis was supported by the findings.

IV. DISCUSSION

A. Main findings

The current study shows that for postalveolar fricatives in Romanian: (a) the plain versus palatalized form can be distinguished reliably based on cepstral measurements (though not to the same extent as with other places of articulation), (b) the secondary palatalization contrast is acoustically realized at this place by 27 out of 31 speakers, (c) women appear to produce the contrast more robustly compared to men, and (d) the contrast is perceptually salient, with listeners displaying moderate sensitivity to it, compared to high sensitivity at the other places examined. The specific ways in which men produce lower acoustic differentiation between plain and palatalized forms remain unclear, but they might involve the reduction or absence of the secondary palatal gesture. Future articulatory studies could clarify whether this is indeed the case.

B. The evolution of the contrast in Romanian

The main reference point with which these findings can be compared is the 1961 study (Șuteu, 1961) regarding the pronunciation of Romanian words. According to Șuteu, 292 of 309 speakers (all from Bucharest, Romania) reported making a distinction between the singular and the plural form of a word ending in a postalveolar fricative. While helping to establish a diachronic perspective, this study was not controlled experimentally, and
it involved self-described pronunciation, without acoustic analysis. The possibility arises that there is no exact overlap between the speakers’ intuitions and their pronunciation. This being the case, the comparison between today’s state of affairs and that in 1961 - and any discussion of potential neutralization - is tentative at best, as no benchmark of the extent of palatalization and the robustness of the contrast in Romanian spoken in the past can be established. The only thing we know for certain is that today males produce somewhat less reliable cues to postalveolar palatalization than females, and that in itself may be interpreted as a sign of either incipient or steady-state (partial) neutralization. Nevertheless, most speakers produce this contrast. Since the mean age of the subjects used in the production experiment reported here was 21.6 years, there could not have been any overlap in the generations that we are comparing.

C. Enhancement vs. neutralization

According to licensing by cue (Kochetov, 1999, 2002; Steriade, 1999), the distribution of a phonological contrast is sensitive to the amount of acoustic information available in a given environment, such that (a) if environment A provides more acoustic information to a contrast between two segments /x/ and /y/, the identification of the contrast by listeners is likely to be high and, as a result, the contrast will be preserved, whereas (b) if environment B provides less acoustic information to the contrast, the identification rate of /x/ versus /y/ will tend to be lower, and the contrast is more likely to be neutralized. Adding to licensing by cue, the phonetic knowledge hypothesis (Hayes & Steriade, 2004) posits that sound patterns in languages reflect the activity of grammatical constraints induced from speakers’ implicit knowledge of certain facts of phonetic difficulty. Perceptually fragile contrasts tend to undergo one of two changes: enhancement or neutralization. This is known as phonetic knowledge-based grounding. It should be mentioned, however, that this view is not uncontroversial. While certain approaches in phonology emphasize the role of physical constraints as the bases of phonological organization and contrast (Browman & Goldstein, 1995) and incorporate facts of acoustic contrast and ease of articulation as constraints in the phonological grammar (Steriade, 1999), other views emphasize the separation between quantitative (acoustic and perceptual) aspects of speech and the phonological representations which encode “relational information” between the paradigmatic contrasts of a given
natural language (Beckman, 1999; Pierrehumbert, 1990).

Historical evidence suggests that perceptually weak contrasts can take the neutralization path. Coda position, for example, has been claimed to be a weaker position for the realization of acoustic cues for consonants in general (Fujimura, Macchi, & Streeter, 1978; Ohala, 1990; Wright, 2004). Steriade (1999) lists seven phonetic properties known as cues to the voicing contrast in obstruent stops: closure voicing, closure duration, duration of preceding sonorant, F1 value in the preceding vowel, burst duration/amplitude, voicing onset time, and F0 and F1 values at onset of voicing in a following sonorant. In coda position, when preceded by a sonorant, three of these are potentially missing: burst duration/amplitude, voicing onset time, and F0 and F1 values at the onset of voicing in a following sonorant. When preceded by an obstruent, only the first two of these cues remain available. It is thus not surprising that coda laryngeal contrasts are typologically more marked than onset laryngeal contrasts. The voiced-voiceless distinction has been neutralized in Russian obstruents (and many other languages) in this position. In the specific case of secondary palatalization, the plain-palatalized contrast with labials in coda position tends to be neutralized cross-linguistically (Kochetov, 2002).

While somewhat difficult to quantify what counts as ‘more’ or ‘less’ acoustic information, the type of information that results in moderate perceptual sensitivity to a contrast, as was found with secondary palatalization in postalveolars in Romanian, would likely fall towards the low end of the scale. This would predict neutralization, and thus the possibility that postalveolar palatalization is subject to a change in progress in Romanian is quite plausible. In fact, neutralization of this contrast may be in progress, as attested by the fact that males appear to realize it to a lesser extent than females.

As far as enhancement is concerned, cases of the plain-palatalized contrast becoming enhanced are well attested. In Russian, plain consonants were described as velarized (Halle, 1959, Trubetzkoy, 1969) or, more recently, as either uvularized or velarized (Litvin, 2014), which is interpreted by Padgett (2001) – following Trubetzkoy (1969) – as a means of enhancing the secondary palatalization contrast. In Romanian, the fricative [s] contrasts with [ʃ] and, to a lesser extent, [z] contrasts with [ʒ]. Spinu & Lilley (2016) also show that the plain-palatalized contrast in fricatives at the glottal/velar place of articulation is implemented as a velar for plain forms and palatal for palatalized ones. It is not clear why /ʃ/ did not follow a similar route. Possible enhancement strategies would include (a) strengthening
to an affricate, as attested for Catalan sibilants which become affricates word-initially and after a consonant, e.g. \[\text{[r} \rightarrow \text{t}, \text{z} \rightarrow \text{d}]\] (Lavoie, 2001); (b) replacement of the secondary articulation with a full-fledged vowel, as in Italian, in which there is no secondary palatalization and the plural affix (identical to the Romanian one underlyingly) is realized as a vowel, e.g. ‘wolf’ is \[\text{[lup]}\], compared to Romanian \[\text{lup}\^] (Lampitelli, 2014); or (c) fortition to a full-fledged stop as in Burushaski \[\text{x} \rightarrow q, \text{g} \rightarrow k\] (Lavoie, 2001).

To summarize this subsection, while there is no evidence of enhancement, there is a possibility that the plain-palatalized contrast in postalveolar fricatives is undergoing incipient neutralization in Romanian, though the available data suggest this may either be a very slow process or very recently initiated.

D. Grammatical restructuring

While the evolution of this contrast over time cannot be established with certainty, the experimental findings reported here indicate that the contrast is nowadays present in the production of Romanian speakers. This situation warrants closer investigation of the factors that may override its relatively weak acoustic distinctiveness (compared to other places) and conspire to its maintenance. As explained in section II.C, the contrast carries a high functional load being associated with the presence of a morphological marker in the vast majority of cases. Enhancement might thus be predicted to be favored but, as discussed in the previous section, no indication of enhancement of the secondary palatalization contrast at this place can be found.

The issue of morphological conditioning brings forth another possibility for the maintenance of a weak contrast, specifically that of grammatical restructuring (Kochetov, 2002). A grammar constructed by a learner can be restructured under pressure from higher-level phonological categories and morphological alternations (Kochetov, 2002). Indeed, Kochetov suggests that deviations from the general patterns observed across languages may be due to properties of the lexicon and grammar of the languages in question. For example, specific non-phonetic characteristics of languages may in some cases override the general markedness patterns established on the basis of phonotactic, articulatory, and perceptual properties of palatalization. Thus, a particular contrast might be maintained in a less favorable environment if the pressure from additional factors is sufficiently strong, especially with regard to
productivity and morphophonological transparency (Kochetov, 2002). Other than productivity, the strength of this pressure depends on the relative salience of these morphological categories (Pierrehumbert, 2001). Highly productive, morphologically transparent alternations are predicted to have stronger effects. The plain-palatalized contrast in less favorable phonotactic environments can be maintained if a lexical item is involved in transparent alternations, or if it signals certain morphological categories. The role of these factors in the maintenance of phonological contrasts should not be overestimated, however, as in most cases the factors that induce neutralization seem to override paradigm uniformity, e.g. syllable coda depalatalization in standard Bulgarian, where palatalized labials and coronals have been depalatalized even in the presence of substantial paradigmatic evidence from alternations (Carlton, 1990). In Polish, the palatalization contrast in labial consonants was similarly neutralized in the coda environment (Kochetov, 2002).

The persistence of Russian palatalized consonants in medial clusters (the most unfavorable environment) may be an example of grammatical restructuring. Also, in the Nova Nadezhda dialect of Bulgarian, all palatalized stops are allowed in word-medial clusters, but these result from the addition of highly productive inflectional or derivational affixes (Kochetov, 2002). Isthmus Mixe also has plain and palatalized postalveolars which are morphologically conditioned, just as in Romanian (Dieterman, 2008). In the case of Romanian, the high functional load carried by the plain-palatalized distinction might override more general markedness effects. While this might explain why the plain-palatalized contrast with postalveolars has been retained in Romanian, it does not provide an answer as to why it has not been enhanced.

E. How is this contrast acquired?

The lack of enhancement is relevant from the point of view of acquisition. If adults cannot reliably perceive the plain-palatalized contrast in postalveolars, presumably children cannot either, since their perceptual system was shown to be similar to that of adults after the age of 1. This is due to a retuning of attention and ensuing gradual decline of the ability to distinguish non-native phonological contrasts once language-specific knowledge is acquired (Aslin, Jusczyk & Pisoni, 1998; Polka & Werker 1994; Werker & Lalonde 1989; Werker & Tees 1984). Together with the morphological factors, orthography may also play an impor-
tant part in the maintenance of this contrast, since any secondarily palatalized consonant is
spelled as a ‘Ci’ sequence. If the role played by orthography is crucial, the distinction would
presumably be absent before learning the correct spelling. A future longitudinal study could
establish if the contrast in postalveolars is acquired before the age of literacy (based on mor-
phological patterns alone) or after learning that the plural (palatalized form) is rendered
with an additional ‘i’ in orthography. Anecdotal evidence suggests that when Romanian
children start learning how to write, failing to add the plural ‘i’ at the end of words ending
in a /ʃ/ is a relatively common mistake. This would be a case of late acquisition due to
external pressure. If acquisition is not delayed, it may be the case that the visual differences
in the articulation of the plain and palatalized forms (i.e. lip rounding in the plain form,
and lip spreading in the palatalized one) play an important part in the intergenerational
transmission of this contrast.

F. Limitations

As previously discussed, some of the limitations of this study were imposed by the make-
up of the original corpus and the original research goals associated with it. The most
important of these has to do with the unbalanced distribution of the vowels preceding the
fricatives, which resulted from the decision to include real, frequently encountered, disyllabic,
stress-final words of Romanian. This problem was addressed by investigating a balanced
subset separately, consisting of about a third of the original data. It was found that the
overall patterns for production and perception alike were similar in both the full corpus and
the subset. The subset permitted the examination of vowel effects, which would not have
been possible otherwise.

A second methodological issue concerns the cepstral coefficient set employed in the acous-
tic analyses. As pointed out by an anonymous reviewer, the spectral properties of fricatives
are strongly influenced by the vocalic context in which they are produced, especially in the
case of labiodental and glottal fricatives. The primary information that cepstral coefficients
will model in coda [f], [v] and [h] is the shape of the tract formed to produce the preced-
ing vowel, which will persevere into the coda fricative; very little information in six-feature
vectors will correspond to the labial or glottal constrictions that primarily characterize the
non-palatalized variants of these fricative pairs. The subset /f|h partially avoided this issue
because only labiodental [f] is subject to these effects (the reader is reminded that Romanian /h/ is realized mostly as a velar or palatal fricative in the context investigated here). Even though the reduced set employed here was successful with various classification tasks applied to corpora of Romanian (Spinu & Lilley, 2016), Russian (Spinu, Kochetov, & Lilley, ms.), and Greek fricatives (Lilley, Spinu, & Athanasopoulou, ms.), it was not an ideal choice for the current study due to its focus on postalveolars, in which all six coefficients employed tended to converge. Inasmuch as the originally formulated goal of uncovering acoustic differences between plain and palatalized postalveolars was attained, the usefulness of the restricted set employed here should not be underestimated. However, the use of a standard 14-coefficient parametric model, which may be able to tease apart consistent differences in tract shaping characterizing plain versus palatalized fricatives, is recommended for future work.

Finally, the lack of acoustic measurements to accompany the results of the 1961 study makes it impossible to compare the status of secondary palatalization in Romanian postalveolars at these two distinct points in time; therefore, all discussion of potential neutralization can only be tentative. The possibility of a real comparison may become available with the collection and annotation of corpora of Romanian spoken at different points in time. In recent years, efforts have been made to join together corpora of spoken Romanian (often referred to as a less-resourced language) produced in various communicative frameworks, as a first attempt to foster resources for corpus-based linguistic studies (Vasilescu et al., 2014). If corpora of Romanian broadcasts from the 1960s become available in the future, it will become feasible to address the issue of neutralization of the plain-palatalized contrast, among others. Furthermore, future articulatory studies (for example, electromagnetic midsagittal articulometer systems) are needed to help settle some of the remaining questions involving the realization and neutralization of this contrast. Unrelated to postalveolars, another question raised by the current work is whether the palatalized form of /h/, which has been referred to as [cʰ] throughout this paper (for consistency with the paradigm), is actually accompanied by a palatalization gesture, thus making it a true palatalized palatal, speculated to be impossible (Hall, 1997), but marginally attested in Mordvin (Campbell, 1974). An articulatory study would be crucial in determining whether [cʰ] exists in Romanian.
V. CONCLUSIONS

This study documents a rare cross-linguistic contrast, that between plain and palatalized postalveolar fricatives in Romanian. While experimental evidence shows that this contrast conforms to typological predictions of being acoustically and perceptually weaker compared to other places of articulation, most speakers’ productions of the plain and palatalized form are acoustically distinct. This situation may have persisted for the past 50 years. Some evidence of neutralization was found with four of the speakers, and the productions of males yielded lower classification rates compared to those of females. The fact that to date this contrast has been neither neutralized (possibly due to morphological restructuring) nor enhanced (when other places within the same paradigm are enhanced) is a good example of the lack of 1-to-1 correspondence between the phonetic factors triggering neutralization and actual neutralization patterns attested in individual languages. In a sense, this contrast appears too important to lose (presumably because of its high functional load) but also stable enough so as not to become enhanced, which may cause it to remain in a state of inertia, almost as if it were below a threshold of phonetic/phonological ‘maneuverability’. Following Kochetov (2002), the best conclusion is that an explanation for neutralization, and phonotactic patterns in general, should not be sought only in phonology or only in phonetics, but in the interaction of phonetic factors with the phonological grammar – and possibly beyond.

ACKNOWLEDGMENTS

This research was supported in part by doctoral dissertation research improvement grant #0720231 from the National Science Foundation. Special thanks are owed to Jason Lilley and Florin Spinu for their help with data analysis.

APPENDIX A

Cepstral coefficients are the sums of the products of the filtered log magnitude speech spectrum and each of the cosine waves illustrated in Figure 6. Thus, \( c_0 \) is the sum of all the spectral values multiplied by 1.0, i.e., the total spectral energy. For the higher order coefficients, each of these will be large and positive when the speech spectrum looks just like
the feature vector. The coefficient will be large and negative when the speech spectrum is
the complement of the feature vector, i.e., large where the feature vector is most negative,
and small where the feature vector is most positive. A good way to conceptualize these
coefficients is that they contrast energy in one or more parts of the spectrum with energy in
other parts of the spectrum. So, for example, C1 contrasts low-frequency energy versus high-
frequency energy; if the spectrum is low-frequency heavy (e.g. vowel) C1 will be positive,
while a spectrum with predominant high-frequency energy (e.g. voiceless sibilant fricative)
will have a negative C1, and the magnitude of the difference between low- and high-frequency
energy determines the magnitude of C1. The C2 coefficient is positive when mid-frequency
energy is weak compared to low- and high-frequency energy, and negative when energy in
the mid-frequencies is strong relative to low- and high-frequency regions. The magnitude of
the difference in energy between mid versus low + high energy determines the magnitude of
the coefficient. It is also worth remembering that because the spectrum is warped to Bark
units, terms like low, mid, and high are not linear frequency descriptors. For example, C2
is strongly positive for a vowel like [a] that has F1 and F2 close together in the region of 1
kHz, and C2 is typically weak or even negative for a vowel like [i] that has a very low F1
and high F2 and a valley between them.

FIG. 6. Bark-scaled cosine terms used in the computation of cepstral coefficients 0-5.
APPENDIX B

Correct classification rates (%) for the ALL and f\textipa{f\textipa{h}} datasets broken down by consonant and palatalization status.

<table>
<thead>
<tr>
<th>Consonant</th>
<th>Palatalization</th>
<th>ALL dataset</th>
<th>f\textipa{h} dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>f</td>
<td>plain f</td>
<td>98.6</td>
<td>98.9</td>
</tr>
<tr>
<td></td>
<td>palatalized f\textipa{d}</td>
<td>95.9</td>
<td>94.5</td>
</tr>
<tr>
<td>v</td>
<td>plain v</td>
<td>97.5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>palatalized v\textipa{i}</td>
<td>95.3</td>
<td></td>
</tr>
<tr>
<td>z</td>
<td>plain z</td>
<td>95.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>palatalized z\textipa{i}</td>
<td>89.6</td>
<td></td>
</tr>
<tr>
<td>f\textipa{j}</td>
<td>plain f\textipa{j}</td>
<td>66.3</td>
<td>73.9</td>
</tr>
<tr>
<td></td>
<td>palatalized f\textipa{j}</td>
<td>71.0</td>
<td>74.4</td>
</tr>
<tr>
<td>h</td>
<td>plain x</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>palatalized x\textipa{j}</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>OVERALL</td>
<td>both plain and palatalized</td>
<td><strong>90.96</strong></td>
<td><strong>90.3</strong></td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Postalveolar</th>
<th>Velar</th>
<th>Glottal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasal</td>
<td>m</td>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plosive</td>
<td>p</td>
<td>b</td>
<td>t</td>
<td>d</td>
<td>k</td>
<td>g</td>
</tr>
<tr>
<td>Affricate</td>
<td></td>
<td></td>
<td>ts</td>
<td>f</td>
<td></td>
<td>h*</td>
</tr>
<tr>
<td>Fricative</td>
<td></td>
<td></td>
<td>s</td>
<td>z</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trill</td>
<td></td>
<td></td>
<td></td>
<td>r</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td></td>
<td></td>
<td></td>
<td>l</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*The allophonic realizations of /h/ include both voiced and voiceless glottal and velar fricatives; in word-final position, it is realized mostly as [x].
TABLE II. Romanian palatalized consonants. NOTE: (1) [d] and [s] are never palatalized in Romanian, and alternate with [z] and [f], respectively, in the context of the plural/second person singular morpheme; (2) [t], [k] and [g] always correspond to a root-final /i/ and are not morphologically conditioned; (3) [f] and [ğ] are the palatalized counterparts for [k] and [g]; (4) [ç] is the palatalized form of /h/.

<table>
<thead>
<tr>
<th></th>
<th>Bilabial</th>
<th>Labiodental</th>
<th>Dental</th>
<th>Postalveolar</th>
<th>Palatal</th>
<th>Velar</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nasal</td>
<td>[pm]</td>
<td>[am]</td>
<td>[an]</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>trees</td>
<td>years</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plosive</td>
<td>[lp]</td>
<td>[lp]</td>
<td>[ph]</td>
<td></td>
<td>[unk]</td>
<td>[ung]</td>
</tr>
<tr>
<td></td>
<td>wolves</td>
<td>lobes</td>
<td>kid</td>
<td></td>
<td>uncle</td>
<td>angle</td>
</tr>
<tr>
<td>Affricate</td>
<td></td>
<td></td>
<td>[hps]</td>
<td>[raf]</td>
<td>[raf]</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>thieves</td>
<td>crawfish, pl.</td>
<td>you bray</td>
<td></td>
</tr>
<tr>
<td>Fricative</td>
<td></td>
<td>[zul]</td>
<td>[grav]</td>
<td>[kaz]</td>
<td>[pa], [paş]</td>
<td>[teç]</td>
</tr>
<tr>
<td></td>
<td>curls</td>
<td>grave, M.pl</td>
<td>you fall</td>
<td>steps</td>
<td>pageboys</td>
<td>Czechs</td>
</tr>
<tr>
<td>Trill</td>
<td>[pur]</td>
<td>pure, M.pl</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Approximant</td>
<td></td>
<td>[bol]</td>
<td></td>
<td></td>
<td>illnesses</td>
<td></td>
</tr>
</tbody>
</table>
TABLE III. Stimuli used in the production experiment. The forms in boldface were used in the subset concerned with the effects of the previous vowel on the palatalization contrast.

<table>
<thead>
<tr>
<th>Final /C/</th>
<th>Vowel</th>
<th>Singular</th>
<th>Plural</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>/f/</td>
<td>o</td>
<td>[pantof]</td>
<td>[pantof]</td>
<td>shoe/s</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>[vataf]</td>
<td>[vataf]</td>
<td>bailiff/s</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>[kartoof]</td>
<td>[kartoof]</td>
<td>potato/es</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>[zuluf]</td>
<td>[zuluf]</td>
<td>curl/s</td>
</tr>
<tr>
<td>/v/</td>
<td>a</td>
<td>[bolnav]</td>
<td>[bolnav]</td>
<td>sick/pl</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>[grozav]</td>
<td>[grozav]</td>
<td>great/pl</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>[firav]</td>
<td>[firav]</td>
<td>feeble/pl</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>[zugrav]</td>
<td>[zugrav]</td>
<td>painter/pl</td>
</tr>
<tr>
<td>/z/</td>
<td>e</td>
<td>[obez]</td>
<td>[obez]</td>
<td>obese/pl</td>
</tr>
<tr>
<td></td>
<td>e</td>
<td>[kinez]</td>
<td>[kinez]</td>
<td>Chinese/pl</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>[ursuz]</td>
<td>[ursuz]</td>
<td>morose/pl</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>[mofluz]</td>
<td>[mofluz]</td>
<td>grumpy/pl</td>
</tr>
<tr>
<td>/j/</td>
<td>o</td>
<td>[kokoʃ]</td>
<td>[kokoʃ]</td>
<td>rooster/s</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>[kodaʃ]</td>
<td>[kodaʃ]</td>
<td>slacker/s</td>
</tr>
<tr>
<td></td>
<td>e</td>
<td>[tireʃ]</td>
<td>[tireʃ]</td>
<td>cherry tree/s</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>[gidoʃ]</td>
<td>[gidoʃ]</td>
<td>playful/pl</td>
</tr>
<tr>
<td>/h/</td>
<td>o</td>
<td>[paroʃ]</td>
<td>[paroʃ]</td>
<td>vicar/s</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>[valaʃ]</td>
<td>[valaʃ]</td>
<td>Wallachian/s</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>[kazaʃ]</td>
<td>[kazaʃ]</td>
<td>Cossack/s</td>
</tr>
<tr>
<td></td>
<td>a</td>
<td>[monaʃ]</td>
<td>[monaʃ]</td>
<td>monk/s</td>
</tr>
</tbody>
</table>
Fig. 1. Waveforms and spectrograms (X axis: Time; Y axis: Frequency 0-7000 Hz) for plain and palatalized consonants in two vowel contexts. For each spectrogram, the first two productions contain the vowel [a] and the last two the vowel [o]. Left: male speaker; Right: female. Top: labiodental; Middle: postalveolar; Bottom: glottal/velar.

Fig. 2. Three-dimensional scatterplots of the three most informative acoustic measures for classification of palatalization. Each point represents a single fricative token in the dataset. The first part of each axis label (e.g. C1.C.1) stands for the coefficient (e.g. C4), the middle part stands for the segment from which it was extracted (C or V), and the last part stands for the region from which it was extracted (1, 2, or 3).

Fig. 3. Region durations (ms) for plain and palatalized forms at three places of articulation.

Fig. 4. Palatalization classification results for models trained either on all 25 parameters or just parameters extracted from a single fricative region. The labels show the gender and the region (or ALL in the case of the entire parameter set).

Fig. 5. Mean sensitivity scores to plain and palatalized consonants.

Fig. 6. (Appendix A) Bark-scaled cosine terms used in the computation of cepstral coefficients 0-5.