Perceptual salience and palatalization in Russian

Darya Kavitskaya
Yale University
darya.kavitskaya@yale.edu

This paper reports on a gating experiment testing the relative salience of sonority, place of articulation, voicing and palatalization. The experiment showed that while sonorant segments are identified by the listeners in the highest percentage of cases than non-sonorant ones, the hypothesis that secondary features will be perceptually less salient than primary features is not supported: the cues for palatalization are at least as perceptually salient for the speakers of Russian as cues for voicing and place of articulation.

1. Introduction
Since the advent of the idea that phonological segments can be further analyzed into distinctive features (Jakobson 1939, building on Trubetzkoy 1939), there was a great deal of research which developed various aspects of feature theory, from the early proposals, e.g. by Clements (1985), Sagey (1986), McCarthy (1988) to later developments, as, for example, Padgett (1995), Avery and Idsardi (2001), Clements (2001). The studies in feature theory cover a wide range of topics, such as the nature of featural representations, the issue of binary vs. privative nature of certain features, and various proposals with respect to the organization of features into hierarchical classes, that is, feature geometry; the acoustic properties of features have been invoked in the discussion of all of these topics. While it has been claimed from the early days of the feature theory that distinctive features are necessarily correlated with phonetic properties of sounds, the exact correlation is still a subject of debate. Defining precise acoustic and articulatory properties of features has been problematic, since phonological features are by definition abstract and often have more than one acoustic correlate1.

This paper focuses on one aspect of a proposal which posits a classification of distinctive features on the basis of their acoustic and articulatory properties. This proposal is advanced in a series of works by Kenneth Stevens; it is based on Halle and Stevens (1991), continued in Stevens (1994), and further developed in more recent work, such as Stevens (2000, 2001, 2002). The research program developed in Stevens’ work provides a phonetically-grounded approach
to the hierarchical organization of features and makes a number of predictions with respect to the notion of salience of distinctive features. The subsequent sections will show that while some of these predictions in are borne out, the others appear not to hold at least for the data tested. Even though the present experiment cannot fully settle the issue of the relative salience of distinctive features, it nevertheless advances our understanding of the matter.

This paper is organized as follows. Section 2 reviews the basics of the Steven’s (1994) model of distinctive features and outlines the predictions that it makes with respect to feature salience. Section 3 presents an experiment designed to test the predictions of the model and discusses the experimental setup, data and methods used. The results, along with the general discussion, are presented in Section 4, which is followed by a conclusion in Section 5.

2. Background and predictions
Both acoustic studies (e.g., Stevens 1989) and auditory studies (e.g., Chistovich and Lublinskaya 1979, Delgutte and Kiang 1984) present evidence in favor of the claim of the standard feature theory that segmental units are represented through sets of distinctive features. Building on this evidence, Halle and Stevens (1991) develop a classification of binary features on the basis of the articulatory properties of a given feature. Based on Halle and Stevens (1991), Stevens (1994: 242) divides distinctive features into the following three classes:

(1) articulator-free features that indicate whether a narrow constriction is made in the vocal tract and, if so, whether or not a complete closure is formed and whether pressure is built up behind the constriction; (2) articulator-bound features indicating the primary articulator that is active in forming the constriction (whether it be a narrow consonantal constriction or a less severe vocalic constriction; and (3) articulator-bound features indicating active adjustments of secondary articulators (i.e. articulators other than the primary ones), such as larynx, the soft palate, and the tongue body (for cases in which the tongue body is not the primary articulator).

The classification above corresponds to the standard feature theory in the following way. Articulator-free features in (1) refer to major class features (such as [±consonantal], [±sonorant], [±continuant]), articulator-bound features in (2) cover place of articulation and voicing features, and a different type of articulator-
bound features in (3) encompasses secondary articulations, such as labialization, pharyngealization, laryngealization, and palatalization.\(^2\)

It has been proposed by Stevens et al. (1986) and Stevens and Keyser (1989) that frequently occurring feature combinations in languages of the world arise since such combinations maximize perceptual distinctiveness through the mechanisms of feature enhancement. The enhancement theory holds that certain features in a sound inventory are perceptually the most salient, or primary, as opposed to less salient, or secondary, features which are selected to enhance the strength of primary features. This notion of salience is further developed in Stevens (1994) to reflect the proposed featural organization. Stevens (1994: 242) states that “[t]his hierarchical organization can serve as a basis for ordering the identification of features from the acoustic signal, with the more context-independent features being identified first and the more context-dependent features identified later.”\(^3\)

One of the predictions which follows from the above interpretation of the hierarchical organization of features is that articulator-free features should be identified by listeners before articulator-bound features, and articulator-bound primary features should in turn be more quickly perceived by listeners than articulator-bound secondary features.\(^4\) In other words, articulator-free features are expected to be maximally salient, articulator-bound features produced with the primary articulator are predicted to be less salient, and secondary features produced with the secondary articulators are predicted to be the least salient. This paper is a report on the results and on the theoretical and methodological implications of an experiment designed to test this prediction.\(^5\)

3. Experiment

A perception experiment was designed in order to test the connection of the feature classification, proposed in Halle and Stevens (1991) and further developed in Stevens (1994), with perceptual salience. If we interpret the theory of perceptual salience correctly, it generates the following predictions with respect to the timing of the listener’s disambiguation of a certain feature in the signal: a) articulator-free features will be disambiguated first; b) the confusions in articulator-bound primary features will be disambiguated second; and c) the confusion in articulator-bound secondary features will clear last. The experiment described below uses several types of confusions made by native Russian listeners to test these predictions. In particular, the confusions between oral and nasal
segments are expected to be disambiguated first since they differ with respect to the articulator-free feature as \([\pm \text{sonorant}]\). The confusions with respect to the feature \([\pm \text{voice}]\) and the place of articulation features \([\text{LABIAL}]\) and \([\text{CORONAL}]\) are predicted to be cleared second. Finally, the confusion in secondary features (which in the case of Russian is palatalization) is expected to persist for the longest time. The following sections show that while the first prediction is borne out, the other two do not appear to hold for the Russian data tested.

### 3.1. Materials

The main reason that the study was based on Russian is that the presence of phonemic palatalization in Russian is non-controversial, and almost all non-palatalized segments in the consonant inventory of Russian have palatalized counterparts. The following partial sound inventory was used for the purposes of the experiment:

(1)  

<table>
<thead>
<tr>
<th></th>
<th>LABIAL</th>
<th>CORONAL</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>STOP</td>
<td>voiceless</td>
<td>p pj</td>
<td>t tj</td>
</tr>
<tr>
<td></td>
<td>voiced</td>
<td>b bj</td>
<td>d dj</td>
</tr>
<tr>
<td>NASAL</td>
<td></td>
<td>m mj</td>
<td>n nj</td>
</tr>
</tbody>
</table>

b. Vowels:

\[u\]  
\[e\]  
\[a\]

A list of 36 words was constructed on the basis of this inventory. All words began with a single consonant, followed by a stressed vowel. Even though all phonemically contrastive Russian vowels \([i e a o u i]\) can bear stress, only three vowels were recorded for the experiment. A high front vowel \([i]\) and a high mid vowel \([i]\) were not considered since they are in complementary distribution in Russian: \([i]\) occurs only after the palatalized consonants, while \([i]\) surfaces only after the plain ones, which makes it impossible to have minimal pairs with respect to palatalization. Only one back vowel \([u]\) was considered, in order to keep the size of the experiment manageable.
The recorded words were no longer than two syllables: whenever monosyllabic words were used, they were of CV or CVC syllable structure, and disyllabic words used consisted of open syllables only, with the stress on the first syllable. Most tokens were real words, but it was impossible to balance the list for the frequency of occurrence for the following reasons. First, Russian native vocabulary lacks sequences of non-palatalized consonants followed by [e] almost entirely. To solve this problem, the names of the letters [pe], [be], [de] and [te] were used. These are less frequent than most words in the list, but nonetheless letter names are often pronounced when the alphabet is recited. For the nasal followed by [e] sequences, a borrowed word for ‘major’ and an acronym were used which are also quite infrequent. Second, Russian lacks stressed syllables like [p’u] and [m’u] so nonsense words were used in these cases.

(2) shows the full word list used in the experiment:

(2)  ‘pe ‘p’ena ‘foam’
  ‘papa ‘p’at’ ‘five’
  ‘puti ‘p’u nonce-word
  ‘be ‘b’edi ‘troubles’
  ‘baba ‘b’aka ‘unpleasantness’
  ‘budu ‘b’u nonce-word
  ‘te ‘t’ema ‘topic’
  ‘tak ‘t’apka ‘hoe’
  ‘tuk ‘t’uk ‘bundle’
  ‘de ‘d’edi ‘grandfathers’
  ‘data ‘d’ad’a ‘uncle’
  ‘duma ‘d’una ‘dune’
  ‘mer ‘m’ed’ ‘copper’
  ‘mama ‘m’ata ‘mint’
  ‘muka ‘m’u word-word
  ‘nep acronym
  ‘nado ‘n’et ‘no’
  ‘nada ‘n’an’a ‘nanny’
  ‘nu ‘n’ura name
The word list was recorded as read by a male speaker of Russian from Moscow. The recording was made on a DAT recorder in a double-walled sound booth. After the recording, individual tokens were saved as audio files (.wav, the sampling rate 20 kHz) and subsequently transferred to the Waves program. A gating program (part of the Waves program) was used to truncate words from the releases of consonants at the first gate of 30 ms (and then 60, 90, 120 ms) of any given word. Stop releases were marked at the first indication of the burst, and nasal releases were labeled at the place of amplitude and/or spectral discontinuity in nasals. The remaining part of a word was replaced by gaussian noise\(^1\). After truncation, only the first consonant and the first part of the following vowel remained. No information about transitions to postvocalic consonants which could provide additional undesired cues for word recognition was included in the signal.\(^2\)

### 3.2. Procedure

The SoundEdit program was used to design the part of the experiment responsible for the program-listener interaction. The stimuli were organized into a 144-token list (12 consonants x 3 vowels x 4 gates) in random order. 10 native Russian listeners participated in the experiment. For each token played, a card with 12 fixed response options was offered to a listener on a computer screen. For example, if a listener heard a truncated word such as [papa] ‘father’, a card with twelve possible choices of a CV sequence labeled in Cyrillic was shown. (3) shows a sample card with fixed responses for all CV sequences with the vowel [a]: [pa], [p\(\ddot{a}\)], [ba], [b\(\ddot{a}\)], [ta], [t\(\ddot{a}\)], [da], [d\(\ddot{a}\)], [ma], [m\(\ddot{a}\)], [na], and [n\(\ddot{a}\)].\(^3\)

(3) A sample card for Ca tokens

<table>
<thead>
<tr>
<th>па</th>
<th>пя</th>
<th>ба</th>
<th>бя</th>
</tr>
</thead>
<tbody>
<tr>
<td>та</td>
<td>тя</td>
<td>да</td>
<td>дя</td>
</tr>
<tr>
<td>ма</td>
<td>мя</td>
<td>на</td>
<td>ня</td>
</tr>
</tbody>
</table>

The listeners were instructed to press a button corresponding to the sequence they heard. Each token was played only once, and a listener had
unlimited time to click the button of their choice. As soon as the response was selected, the next token was played.

3.3. Analysis
After the data collection, confusion matrices\(^{14}\) based on listeners’ responses were constructed for each listener at four gates, for all listeners together at four gates, and for all responses as a whole (see Appendix 1). The results were analyzed by running repeated-measures ANOVA (Statistical Package for the Social Sciences (SPSS) was used for the analysis). Statistical analysis performed on the listeners’ responses (dependent variable: total number of errors) showed that the effect of the time (gate) as an independent factor was significant at \( p < 0.007 \) and the effect of the vowel and of the vowel-gate interaction was not significant. The further analyses were using the following independent factors:

- (1) sonority with two levels (nasal and oral);
- (2) obstruent voicing with two levels (voiced and voiceless);
- (3) place of articulation with two levels (labial and dental);
- (4) palatalization with two levels (plain and palatalized).

The percentage of confusions was calculated as the percentage of erroneous responses given the total possible number of errors for a given feature. The analysis reported below is based on the results including the three vowels used in the study and all ten subjects who participated in the listening experiment.

4. Results and discussion
The following sections present the results of the experiment in the form of graphs showing the percentage of confusion with respect to the following parameters:

I. **Articulator-free “primary” feature [±sonorant]**
   - sonorants (nasals) heard as obstruents (stops) and vice versa

II. **Articulator-bound “primary” features**
   - labials heard as dentals and vice versa [Place of Articulation]
   - voiced segments heard as voiceless and vice versa [±voice]

III. **Articulator-bound “secondary” features**
   - palatalized segments heard as non-palatalized and vice versa
The percentage of confusion is shown separately for each of the four gates. In the statistical analysis, the dependent variable is always defined as the percentage of errors, and time (gate) is treated as an independent factor along with other parameters.

4.1. Primary features
4.1.1. Sonority
First, we consider the direction of confusion in primary features, such as sonority and voicing. Figure 1 illustrates the direction of confusion for nasals and stops. The three graphs show nasals misperceived as oral sounds (N-O), and the confusion of obstruents as nasals is broken down further to voiced stops misheard as nasals (V-N) and voiceless stops misheard as nasals (Vl-N).

Figure 1. Sonority confusions: oral vs. nasal

Statistical analysis was performed with gate and sonority as dependent factors:

- gate: \( F(3,27) = 5.03, p < 0.01 \)
- sonority: \( F(2,18) = 3.25, p < 0.1 \)
- gate * sonority: \( F(6,54) = 2.42, p < 0.05 \)

There are two noticeable asymmetries in these data. First, as predicted by Stevens (1994), the confusion of sonorants with obstruents is the smallest at all four gates. Statistical analysis shows that the effect of gate is significant, but sonority is not (marginally significant at \( p < 0.1 \)). This result holds independent of the direction of confusion and provides support to the prediction that the feature
[±sonorant] (or arguably [nasal] in this case) should be identified first in the sound signal as the most salient.

It can be hypothesized that the asymmetry in the nasal/stop confusions could be connected with the privative nature of the feature [NASAL] which corresponds to velum lowering. Privativity of [NASAL] was argued for by Steriade (1993a, 1993b), Trigo (1993), and in subsequent work. Steriade (1995: 149) states that “[t]here is virtually no evidence left suggesting that orality is represented phonologically, in any language.” The facts in Figure 1 follow naturally from the privativity of [NASAL]. Once the privative feature is present in the signal, it is unambiguously heard as such. Once it is absent, there is no information with respect to the nasalization present, so the listeners can interpret the signal as either nasal or oral.

Second, we can see that at the first two gates voiced obstruents are confused with nasals significantly more often than voiceless obstruents. Bonferroni tests indicate that the percent of confusions for voiced stops differed both from voiceless stops and nasals which did not differ from each other. I suggest that this asymmetry is to be expected. In Russian, phonologically voiced obstruents (both plain and palatalized) are fully voiced phonetically (Matusevich 1976, Bolla 1981), so the vocal cord vibration continues throughout the closure interval (Stevens 1998). Such “prevoicing” in stops can be easily interpreted as nasalization, especially syllable-initially, as in the case of our data.

4.1.2. Voicing
The direction of confusion with respect to voicing is asymmetrical as well, as shown in Figure 2: voiceless segments are heard as voiced consistently more often than voiced segments are heard as voiceless (nasals were not included in this comparison).
The dependent variables for ANOVA were gate and voicing of the initial consonant. The analysis shows that voiceless segments are heard as voiced significantly more often than voiced segments are confused with voiceless at all four gates (no significant interaction between gate and voicing was obtained):

gate: \( F(3,27) = 5.24, p < 0.01 \)
voicing: \( F(1,9) = 10.99, p < 0.01 \)

gate * voicing: \( F(3,27) = 1.17, \text{n.s.} \)

It is tempting to analyze voicing confusion asymmetry in the same way we analyzed the directionality of confusion in nasals vs. orals. If we assume that \([\text{VOICE}]\) is privative, as we did for \([\text{NASAL}]\) in discussing the graphs in Figure 1, we can hypothesize that this feature is disambiguated by the listeners as soon as it is present in the signal; if the feature \([\text{VOICE}]\) is missing, as in the case of voiceless stops, the listeners could be guessing either its presence or its absence. Although this is an appealing solution, the privative nature of voicing is considerably more controversial than the privativity of \([\text{NASAL}]\). Researchers such as Mester and Ito (1989) and Lombardi (1991, 1995) argued for the privativity of \([\text{VOICE}]\) (and some phonologists, for example, van Rooy and Wissing 2001, assume voicing to be privative without further discussion, but Rubach (1996) provides a convincing argument against this view.\(^1\)

### 4.2. Secondary features: palatalization

We now turn to the main topic of this paper, palatalization as a factor in consonant confusion. The patterns of confusion with respect to place of

---

\(^1\)
articulation and palatalization appear to go against the predictions of Stevens (1994). Recall that it is predicted that place of articulation features would be more robust and more readily (quickly) heard than secondary articulation features, such as palatalization.

The graphs in Figure 3 show that that the percentage of confusion of palatalized segments with non-palatalized is slightly higher than of non-palatalized segments with palatalized (where X stands for all non-palatalized tokens while X’ signifies all palatalized tokens). However, the statistical analysis has shown that this tendency is not statistically significant (p < 0.3).

![Figure 3](image)

Figure 3. Palatalization confusion

Even though the difference in the direction of confusion with respect to palatalization is not significant, patterns emerge with the further analysis of the palatalization data, as soon as such factors as sonority, voicing, and place of articulation of the palatalized vs. plain consonants in question are taken into consideration. These patterns are presented in the following sections.

4.2.1. Palatalization and sonority
The graph in Figure 4 shows confusions of palatalization in orals (any plain oral segment heard as palatalized and vice versa) and in nasals (any plain nasal segment heard as palatalized and vice versa). It is evident that palatalization is hardly ever confused in nasals, as opposed to orals, especially in the first two gates.
Both independent factors (gate and sonority) are statistically significant:

- gate: \( F(3,27) = 27.57, \ p < 0.001 \)
- sonority: \( F(2,18) = 71.91, \ p < 0.001 \)
- gate * sonority: \( F(6,54) = 11.9, \ p < 0.001 \)

The asymmetry in Figure 4 appears to have an explanation in acoustics. Even though in nasals and orals the phasing of the palatalization gesture with respect to other gestures is similar (Louis Goldstein, p.c.), in the case of stops no acoustic information is available during the first two gates (30 and 60 ms) to disambiguate palatalization. The cue for palatalization is mainly the transition to the following vowel which is lacking in the first two gates.\(^2\) In the case of palatalized nasals, however, the difference in acoustics is present in nasal closure (Bolla 1981), which prevents confusion of palatalized vs. non-palatalized nasals.\(^2\)

4.2.2. Palatalization and place of articulation
The next factor to be considered in the palatalization confusion is place of articulation. Figure 5 shows that palatalized labials are confused with their plain counterparts significantly more often than palatalized dentals are confused with plain dentals.
Both independent factors (gate and POA) are statistically significant while the interaction is not:

- gate: $F(3,27) = 15.93, p < 0.001$
- POA: $F(1,9) = 33.75, p < 0.001$
- gate * POA: $F(3,27) = 1.82, n.s.$

The asymmetry present in Figure 5 is expected. Kochetov (2002: 116) predicts that “the palatalized labial /p/ should be more likely to be perceived as plain than the palatalized coronal /t/...”. The reason for this prediction lies in the articulatory properties of palatalized labials and coronals. In both palatalized labials and coronals two articulators are involved, but there is nevertheless the difference in production. According to Ladefoged and Maddieson (1996: 364), “bilabial sounds do not require any specific position of the tongue for their articulation, the tongue body can assume an i-like position during their production without any conflict with the demands of the primary articulation.” However, in the case of the palatalized coronal the primary articulator is shifting its position. Ladefoged and Maddieson (1996: 365) note that palatalization of a coronal “consists of a displacement of the surface of the tongue front from the position that it would assume in the non-palatalized counterpart, when its role is to support the movement of the tongue tip or blade,” so that palatalized articulation of coronals “can be viewed as the summation of two movements, with the displacement of the tongue front often producing a slightly different primary constriction location.” These differences between the interaction of labiality and coronality of a segment with the palatalization gesture can explain the asymmetry.
in Figure 5. Acoustically, the palatalization gesture affects the coronal burst more than the labial burst, so more information about palatalization is present earlier in the signal. In the case of a labial, palatalization is more an off-glide effect detected later.

Since the palatalization gesture is similar to the coronal gesture we predict that palatalized labials which have coronal cues will be confused with coronals [t] and [d] more often than non-palatalized labials which do not have such cues. Figure 6 shows graphs for the confusions of plain and palatalized labials with dentals.

![Figure 6](image)

**Figure 6.** Palatalized and plain labials confused with dentals

While there is a clear tendency for palatalized labials to be confused with dentals more often than for the plain ones, the difference is only marginally statistically significant:

- gate: \( F(3,27) = 20.97, p < 0.001 \)
- palatalization: \( F(1,9) = 7.92, p < 0.02 \)
- gate * palatalization \( F(3,27) < 1, \text{n.s.} \)

4.2.3. **Palatalization and voicing**

Finally, Figure 7 addresses the direction of palatalization confusion in voiced vs. voiceless stops. The graphs in Figure 7 illustrate four types of confusions:

1) voiced plain stops heard as voiced palatalized stops (V – V');
2) voiced palatalized stops heard as voiced plain stops (V' – V);
3) voiceless plain stops heard as voiceless palatalized stops (VL – VL');
4) voiceless palatalized stops heard as voiceless plain stops (VL' – VL).
**Figure 7.** Palatalization confusion in voiced vs. voiceless (by direction, both labial and dental)

The statistical analysis shows that plain voiced stops are significantly more likely to be misperceived as palatalized:

- **gate:** $F(3,27) = 13.27, p < 0.001$
- **voicing:** $F(1,9) = 11.25, p < 0.001$
- **gate * voicing:** $F(3,27) = 4.94, p < 0.007$
- **palatalization:** $F(1,9) = 2.35, \text{n.s.}$

I suggest that the explanation of this asymmetric pattern is crucially connected with aerodynamic constraints that govern the generation of turbulence in air flow, specifically, with the effect of a following high vowel, glide, or palatalization *per se* on the burst of a stop. Due to the high turbulence noise caused by the changes of aperture, palatalized voiceless stops frequently have longer and more fricated releases which sometimes results in their phonological spirantization (Ohala 1989, see also Kavitskaya 1997). The presence of high frequency noise and a longer duration of noise upon the release can help listeners to distinguish voiceless palatalized stops from their non-palatalized counterparts early in the signal.

**5. Summary and conclusions**

The listening experiment described above provided an opportunity to test the relative salience of sonority, place of articulation, voicing and palatalization. The results of the experiment partially supported the predictions of Stevens (1994). Indeed, nasals were identified by the listeners more readily and quickly than orals.
However, the direction of confusion was asymmetrical: orals were identified as nasals more often than vice versa. We proposed that the privative nature of the feature [NASAL] was responsible for this asymmetry.

We have also seen an asymmetry in confusions with respect to voicing. First, voiced stops were confused with nasals more often than voiceless stops. We have proposed that this asymmetry was connected with acoustic properties of stops and nasals, and that prevoicing in voiced stops could be mistaken for nasalization and responsible for the higher rate of confusion. Second, voiceless stops were confused with voiced stops more often than vice versa. We tentatively suggested that the privativity of [VOICE] might be responsible for this asymmetry.

Contrary to the expectations, the difference in confusion between place of articulation and palatalization was not statistically significant. However, a more subcategorized analysis revealed further asymmetries in the direction of confusion. First, palatalized nasals were identified better than palatalized orals. We proposed that this is due to the fact that palatalization can be detected very early in the nasal closure. Second, palatalization cues were missed in the case of palatalized labials significantly more often than in palatalized coronals. Since acoustically the palatalization gesture affects the coronals more than labials, this asymmetry received a phonetic explanation as well.

The results of this experiment had shown that cues for palatalization are at least as perceptually salient for the speakers of Russian, as cues for voicing or place of articulation. This suggests that the distinction between the primary and secondary features is not correlated with the notion of perceptual salience. This study provides evidence that features are not extracted from a sound signal in some particular order. Rather, the listeners disambiguate the featural information whenever it is present, and if the interaction of acoustic and articulatory properties makes it possible to detect certain features early in the acoustic signal, it is done so by the listeners.
Appendix 1: Perceptual confusions among consonants

<table>
<thead>
<tr>
<th>Played</th>
<th>p</th>
<th>p'</th>
<th>t</th>
<th>t'</th>
<th>b</th>
<th>b'</th>
<th>d</th>
<th>d'</th>
<th>m</th>
<th>m'</th>
<th>n</th>
<th>n'</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>21</td>
<td>4</td>
<td>2</td>
<td>14</td>
<td>3</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>p'</td>
<td>12</td>
<td>4</td>
<td>3</td>
<td>22</td>
<td>3</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>t</td>
<td>21</td>
<td>4</td>
<td>2</td>
<td>14</td>
<td>3</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>t'</td>
<td>16</td>
<td>1</td>
<td>2</td>
<td>22</td>
<td>1</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>b'</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>22</td>
<td>1</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>d</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>d'</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>22</td>
<td>1</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Total perception; Gate = 30 ms

Perceived

<table>
<thead>
<tr>
<th>Played</th>
<th>p</th>
<th>p'</th>
<th>t</th>
<th>t'</th>
<th>b</th>
<th>b'</th>
<th>d</th>
<th>d'</th>
<th>m</th>
<th>m'</th>
<th>n</th>
<th>n'</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>21</td>
<td>4</td>
<td>2</td>
<td>14</td>
<td>3</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>p'</td>
<td>12</td>
<td>4</td>
<td>3</td>
<td>22</td>
<td>3</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>t</td>
<td>21</td>
<td>4</td>
<td>2</td>
<td>14</td>
<td>3</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>t'</td>
<td>16</td>
<td>1</td>
<td>2</td>
<td>22</td>
<td>1</td>
<td>14</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>18</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>1</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>b'</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>22</td>
<td>1</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>d</td>
<td>7</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>d'</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>22</td>
<td>1</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

Total perception; Gate = 60 ms

Perceived
### Total perception; Gate = 90 ms

<table>
<thead>
<tr>
<th>Perceived</th>
<th>Played</th>
<th>Perceived</th>
<th>Played</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p'</td>
<td>t</td>
<td>t'</td>
</tr>
<tr>
<td>p</td>
<td>30</td>
<td>p'</td>
<td>23</td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n'</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Total perception; Gate = 120 ms

<table>
<thead>
<tr>
<th>Perceived</th>
<th>Played</th>
<th>Perceived</th>
<th>Played</th>
</tr>
</thead>
<tbody>
<tr>
<td>p</td>
<td>p'</td>
<td>t</td>
<td>t'</td>
</tr>
<tr>
<td>p</td>
<td>30</td>
<td>p'</td>
<td>15</td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n'</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
References
Kochetov, Alexei. This volume. Syllable position effects and gestural organization: Articulatory evidence from Russian.


Endnotes:

* I would like to thank Louis Goldstein and John Ohala for their help and encouragement and Doug Whalen and Matthew Richardson for help with statistics. I owe much gratitude to two reviewers of the manuscript. I am also indebted to Jonathan Barnes for insightful comments and helpful discussions.

1 For example, Lisker (1986) lists the following perceptual cues which can be relevant for the listener in categorizing a sound as voiced or voiceless: pre-closure cues (vowel duration, duration of $F_1$ transition, $F_1$ offset frequency, $F_1$ transition offset time, time of voice offset, $F_0$ contour, decay time signal); closure cues (closure duration, duration of glottal pulsing, intensity of glottal signal), and post-closure cues (release burst intensity, VOT, onset of $F_1$ transition, $F_1$ onset frequency, $F_1$ transition duration, $F_0$ contour).

2 Voicing belongs to group (2); even though larynx is mentioned in (3), it is only to refer to laryngealization as a secondary articulation.

3 However, see Whalen (1984) study for an argument that certain secondary cues (e.g. transitions of a fricative vowel) can be taken into account even when the “primary” cues (fricative noise) are unambiguous.

4 In this paper, I will treat the prediction that certain features should be “identified first” as roughly equivalent to “identified best”. I thank an anonymous reviewer for pointing this out.

5 Note that we are not questioning other premises of the enhancement theory, such as the claim that secondary features will be included in languages’ inventories only given the selections of primary ones.

6 Note that [t] and [d] in Russian have dental, rather than alveolar, articulation.

7 Padgett (2001) argues that the contrast between consonants before /i/ and / ī/ is that of palatalized vs. velarized ones.

8 Such sequences occur in recent unassimilated borrowings, e.g. [temp] ‘tempo’, [po tensija] ‘potency’, etc.

9 Nonsense words did seem not present any problems to the test group. According to the listeners, these tokens with did not sound any more foreign or unnatural than the others.

10 Unstressed vowels undergo reduction in Russian which is not shown since it is irrelevant for the purposes of this work.

11 The noise portions were uniform in amplitude and duration, not modulated to mimic the missing part of the word.

12 It was also important to control for the third consonant in a word used as a token. The words were chosen so this consonant would not be a fricative, in order not to provide listeners with additional cues which would be present in a vowel.

13 In Cyrillic alphabet, the distinction between plain and palatalized consonants is marked on the following vowel. Thus, [pa] is written as ‘na’, while [p'a] is written with the same sign for the consonant but a different one for the vowel, ‘на’.

14 See Miller and Nicely (1955) on the procedure of interpreting confusion matrices.

15 Steriade (1995: 170) mentions that according to Cohn (1990) there is evidence that “oral segments may possess specified articulatory targets for a raised velum position.” However, since
the presence of a phonetic target does not necessarily imply the presence of a corresponding phonological feature, we do not consider it to be counterevidence to the privativity of [NASAL].

16 This analysis constitute a possible counterexample to John Ohala’s theory of asymmetrical confusion. According to Ohala (2001), it is easier to mishear a feature which is present than to “hallucinate” an absent feature.

17 I suspect that different results would follow if syllable-final postvocalic consonants were tested in the same set-up. For syllable position effects see Kochetov (this volume).

18 As was mentioned in Footnote 1, voicing has many acoustic cues. It is also unclear if the feature [VOICE] can be associated with one gesture. This feature can be interpreted as larynx lowering, as well as vocal fold vibration. Jessen (2001) notes that while the feature [VOICE] is generally interpreted as vocal fold vibration, such vibration is not a single phenomenon but several gestures that need to be coordinated.

19 Note that the direction of confusion of palatalization in nasals (that is, palatalized nasal as plain vs. plain nasal as palatalized) is not statistically significant.

20 But see discussion below for the differences between palatalized labials and coronals.

21 The same is true of the formant structure of laterals which were not tested here. I predict that the same effects will be observed for liquids tested in a similar experiment.