

Perception of gestural overlap and self-organizing phonological contrasts*

Alexei Kochetov
Haskins Laboratories/Simon Fraser University

0. Introduction

Languages that maintain distinctive secondary articulation contrasts tend to avoid multiple vowel contrasts, particularly rounding contrasts in front and back vowels. At the same time, languages with complex vowel inventories very rarely show distinctions in secondary consonant articulations, for example, in palatalization or labialization. These observations are based both on an analysis of the UPSID Database (Maddieson & Precoda 1990) and on an examination of inventories of a number of languages of Europe that exhibit at least one of the above mentioned contrasts. In this paper I provide an explanatory account of these co-occurrence restrictions on seemingly unrelated segments and derive the two mutually exclusive patterns through a learning simulation. I demonstrate that these markedness effects emerge naturally from low-level interactions between a speaker and a listener/learner as a result of limits on what can be successfully transmitted through the speech communication channel. The key factor in the process is the failure on the part of the listener to correctly process overlapped gestures that happen to share the same articulator.

* The research is supported by a Post-doctoral Fellowship from the Social Sciences and Humanities Research Council of Canada (SSHRC), #756-2001-0145, held at Haskins Laboratories, New Haven, CT. The initial survey of sound inventories was conducted as a part of the Project on Contrast in Phonology supported by SSHRC grants #410-99-1309 and #410-96-0842 to Elan Dresher and Keren Rice, University of Toronto. I am grateful to Louis Goldstein of Haskins Laboratories/Yale University for insightful comments on general theoretical issues that have influenced this work. *All errors are my own.*

1. Observations

In this paper I focus on two types of contrasts: the high vowel contrasts that involve front/back and rounded/unrounded dimensions (e.g., /i/ vs. /y/ and /u/ vs. /ʊ/), and contrasts secondary articulations in consonants, “plain” versus palatalized (C^j), velarized (C^v), or labialized (C^w). Both types of contrasts are known to be marked. For instance, the vowels /y/ and /ʊ/ are less common in world languages than /i/ and /u/; so are consonants with distinctive secondary articulation (Maddieson 1984: 124-125; 38). The presence of the marked segments (e.g., front rounded /y/ and palatalized dental/alveolar /t^j/) often implies the presence of the unmarked ones (front unrounded /i/ and non-palatalized dental/alveolar /t/). What is interesting, however, is that the two types of contrasts very rarely co-occur in language inventories.

1.1 UPSID

An analysis of the UPSID Database (Maddieson & Precoda 1990; 451 languages) shows that languages tend to maintain either distinctive secondary articulation contrasts in stops or rounding and backness contrasts in high vowels. At the same time, languages that contrast unrounded and rounded vowels of the same tongue position (e.g., /i/ vs. /y/ and /y/ and /ʊ/) very rarely show distinctions in secondary consonant articulations.

The database contains 134 languages that have at least one of the following segments: high vowels /y/ or /ʊ/ contrastive with their unmarked counterparts /i/ and /u/; palatalized, velarized, and labialized stops of any place of articulation. These languages are listed in (I) in the Appendix. Note that Maddieson & Precoda 1990 list only “true

palatalized” segments, that is those, characterized by a simple addition of a secondary palatal approximant-like constriction and no modification of the primary place, such as dental/alveolar /t^j d^j/ (Maddieson 1984: 166-167). Thus the palatal stops /c ɟ/ and post-alveolar affricates /tʃ dʒ tɕ dʑ/ are not listed there, even though in a given language they may pattern together phonologically with other palatalized consonants (e.g. /p^j/ or /k^j/).

Out of 135 languages, 81 (60%) have consonants with secondary articulation at least at one place of articulation, but allow neither /y/ nor /ɯ/ (in addition to /i/ and /u/). Another 47 languages (35%) have the vowels of interest but no secondary articulation distinction in the consonants. Only 6 languages (4%) have both types of marked contrasts. Thus Mari and Selkup are listed in UPSID as having palatalized stops (dentals/alveolars and/or labials) together with the high front rounded /y/. The other 4 languages (Highland Chinantec, Lue, Mbabaram, and Kawaiisu) have labialized stops of only one place of articulation together with high back unrounded /ɯ/. It should be noted, however, that the status of palatalized consonants in Mari and Selkup is not completely clear, since other sources do not mention these segments in inventories of these languages (V.V. Vinogradov et al. 1966b, V.A. Vinogradov et al. 2001).

Overall, the set of languages with contrastive secondary articulations hardly “overlaps” with the set of those that distinguish backness and rounding contrasts. Exceptions seem to be limited to the languages that exhibit less robust, marginal contrasts in secondary articulation.

1.2 Languages of Europe

In order to further test these observations I turn to languages of Europe since many of these are known to have a greater than average number of vowels (Maddieson 1984: 128) as well as complex consonant inventories. Many of these languages have plain-palatalized contrast in stops or/and front rounded vowels. The survey reported here is based primarily on the following sources: Ball & Fife 1993, Comrie & Corbett 1993, Harris & Vincent 1988, Iartseva 1993, Iartseva et al. 1997, König & Van der Auwera 1994, MacAulay 1992, V.V. Vinogradov et al. 1966ab, V.A. Vinogradov et al. 2001. A list of 46 languages that exhibit the above mentioned contrasts is given in (II) in the Appendix.ⁱ Where the status of palatalized labials and velars in a language is disputed, these consonants are listed in parentheses. In a number of languages, palatalized counterparts of plain dentals/alveolars are realized as alveolo-palatal /tç dz/ or palato-alveolar affricates /tʃ dʒ/ and, in some cases, as palatal stops /c ɟ/. These are also given in parentheses.

The results are very similar to our findings based on UPSID.ⁱⁱ 46% of the languages (22 languages) exhibit secondary palatalization but have no front rounded vowels. In almost half of these languages the contrast between plain and palatalized consonants is fairly robust, extending to two or three places of articulation: labial, coronal, and velar. Another 46% of the sample (22 languages) exhibit the opposite pattern: front rounded vowels /y/, /ø/, or /œ/ occurring “at the expense of” secondary palatal articulation.ⁱⁱⁱ Only in 8% of the languages (4 languages) do palatalized consonants and front rounded vowels co-occur: Estonian, Karelian, Veps, and Chuvash. The first three of these languages

belong to the Baltic group of Finno-Ugric family, and they all exhibit palatalized coronal stops in addition to front rounded vowels /y/ and /œ/. Veps and some dialects of Karelian are also reported to have palatalized labials and velars (Iartseva 1993, V.V. Vinogradov et al. 1966b); however, it is not clear from the sources whether these are phonemically contrastive (at least in Veps). There are also distributional restrictions on both palatalized coronals and front rounded vowels: for instance, in Estonian /t^j d^j/ do not contrast word-initially and /y/ and /œ/ do not occur in non-initial syllables. It should be noted that in many of the languages with front rounded vowels these segments often participate in the processes of palatal vowel harmony (e.g., Finnish or Tatar) or umlaut (e.g., German or Icelandic). Interestingly, vowel harmony in the languages with palatalized segments, Estonian and Veps, is no longer productive (Iartseva 1993).

Two Turkic languages, Karaim and Gagauz, are listed both with “front rounded vowel languages” and with “palatalization languages.” This is done because different dialects of these languages exhibit one of the two types of contrasts: either complex vowel contrasts or secondary palatal articulation contrasts (Comrie 1984: 63-64, Iartseva et al. 1997, V.A. Vinogradov et al. 2001). It is interesting that palatal vowel harmony in some dialects corresponds to consonant secondary articulation harmony (e.g., Karaim *køzymde* ~ *k^joz^jum^jd^ja* ‘in my eye’: Comrie 1984: 63). Yiddish presents another interesting case: the lack of front rounded vowels sets it apart from other Germanic languages. The absence of these segments correlates with the phonemic status of palatalized sonorants /n^j/ and /l^j/, and in some dialects with palatalized dentals /t^j d^j/ (König & Van der Auwera 1994; V.V. Vinogradov et al. 1966a).

To summarize, there is a strong tendency for languages to avoid having both distinctive secondary articulation contrasts and multiple distinctions in rounding/backness and for languages with multiple vowel contrasts to avoid distinctions in secondary consonant articulations. The main question is: why are these two seemingly unrelated contrasts incompatible? In the rest of the paper I will provide an explanatory account of this phenomenon.

2. Sources of explanation

The traditional approach to phonological universals assumes that all markedness effects are innate, pre-specified in Universal Grammar. Thus the facts of incompatibility of the two marked contrasts have to be built into either harmonic rankings of constraints (*Optimality Theory*; Prince & Smolensky 1993) or universal phonological representations (e.g., Clements 1985). In this paper I consider an alternative approach that argues that these markedness effects arise due to lower-level factors – primarily due to the limitations of speech production and perception.

This view builds in part on existing work investigating the role of low-level articulatory and perceptual factors in shaping phonological structure (Ohala 1981, Kawasaki 1982, Browman & Goldstein 1986, 1999, Hume & Johnson 2001, Pierrehumbert, Beckman, & Ladd 2001, among others). It also crucially relies on the concept of *self-organization*, or spontaneous emergence of order that is characteristic of various natural dynamic systems (see, for example, Kauffman 1995). Some recent work in the fields of Artificial Intelligence and Artificial Life has demonstrated that complex structures and high-level ontologies can emerge due to low-level sensory-motor

interactions of simple autonomous entities – robots or simulated agents. Significantly, this is achieved without any prior specification for this higher level knowledge (see Pfeifer & Scheier 2001 for a review; see also Brooks 1991, Langton 1995, and Steels 1995).

The self-organization approach, extended to phonology, can be stated as follows: high-level phonological structure — phonological markedness effects — can result from low-level speaker-listener interactions without being directly specified in Universal Grammar. A simplified version of these interactions can be seen as production and perception of lexical items (sensory-motor coordination) and certain kinds of higher-level processing of the perceived input (categorization and generalization). In this approach markedness effects take on a different meaning. “Phonologically unmarked” can be understood as stable with respect to either production, or perception, or higher-level processing, or, in dynamic terms, an *equilibrium position*. “Phonologically marked” would mean unstable with respect to either production, or perception, or higher-level processing, or a *non-equilibrium position*. Note that the notion of marked or unmarked may reflect a combined effect or interaction of these kinds of factors. Over time languages tend to retain stable, unmarked, phonological structures and discard the structures that are less stable, or marked.

Returning to the problem in question, how can we explain the apparent incompatibility of complex vowel contrasts and secondary articulation contrasts? It is hypothesized here that a grammar that allows both types of contrasts is highly unstable with respect to production and perception. That is, either the speaker’s articulation of

these contrasts or the listener's perception of them, or both these activities have an error rate high enough to affect the transmission of this grammar from the speaker to the listener/learner. Given these natural limitations, the system will easily give way to more stable patterns with either of the two marked types of contrasts.

It is crucial for our analysis that high vowels and secondary articulations in consonants are phonetically related. Both segments involve the same articulators: *tongue body*, which is either fronted (as, e.g., for /i/ and /p^j/) or backed (as, e.g., for /u/ and /p^y/); and *lips*, which are rounded (as, e.g., for /y/ and /t^w/) or unrounded (as, e.g., for /u/ and /t^v/). As a consequence of this articulatory similarity the corresponding high vowels and secondary articulations are also similar acoustically and perceptually. These factors are built into the simulation discussed below.

3. Simulation

The hypothesis outlined above can be investigated using a computer simulation of speaker-listener/learner interactions where the “speaker” and the “listener” are “agents” or simple autonomous entities. Agent-based programming has been used recently to investigate various emergent phonological phenomena (e.g., gestural phasing: Browman & Goldstein 1999; vowel inventories: de Boer 2000; word pronunciations: Lieberman 2002, vowel harmony: Harrison, Dras & Kapicioglu, to appear). The simulation presented in this paper is far less elaborate than some of those in the works mentioned above; however, it appears to be adequate to handle the problem at hand.

3.1 A hypothetical language

In order to test whether unmarked patterns can emerge through speaker-listener interactions I intentionally chose a hypothetical language with excessively marked inventories of consonants and vowels. This language employs 4 consonants that share their primary place and differ in their secondary articulation: palatalized, labio-palatalized, velarized, and labialized (1a). It has 4 high vowels that are differentiated along the front-back and rounded-unrounded dimensions (1b), thus corresponding to the secondary articulations. Lexical items in this language are limited to the shape C_1VC_2 (where $C_1 = C_2$), giving the total of 16 items (2a). Each of the items has a distinct meaning; however, the details of phonological-semantic mapping are not important here. Note that in this lexicon all four consonants and all four vowels are fully contrastive, that is, they occur in all logically possible environments.

- (1) a. Consonant inventory: $\{C^j C^q C^y C^w\}$
 b. Vowel inventory: $\{i y u \text{ } \text{ } \}$

(2)

$C^j i C^j$	$C^j y C^j$	$C^j u C^j$	$C^j \text{ } C^j$
$C^q i C^q$	$C^q y C^q$	$C^q u C^q$	$C^q \text{ } C^q$
$C^y i C^y$	$C^y y C^y$	$C^y u C^y$	$C^y \text{ } C^y$
$C^w i C^w$	$C^w y C^w$	$C^w u C^w$	$C^w \text{ } C^w$

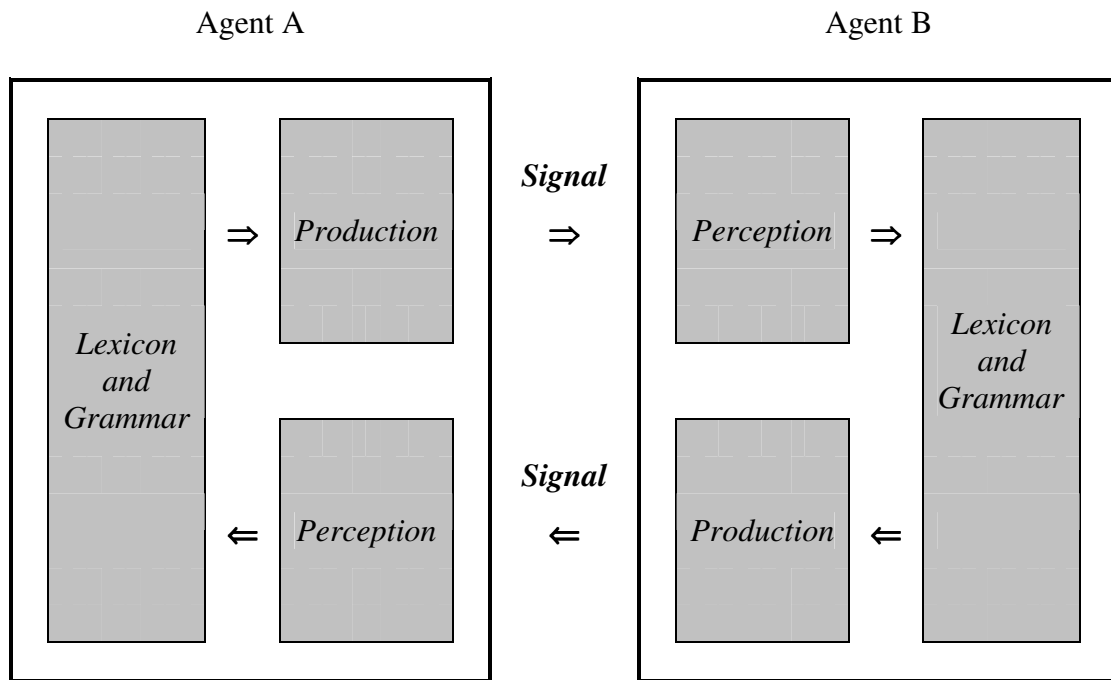
Note that “C” in our analysis represents a stop of any place of articulation, since our focus is on the properties (phonetic and phonological) of the secondary articulations rather than differences in the primary place. In addition, omitting the primary place of

articulation substantially simplifies modeling of the articulation and perception of the consonants of interest.

3.2 Agent interactions

The interactions involve two agents: an adult agent, Agent A, and the learning agent, Agent B (3). Each of the agents consists of the following components, or modules: production, perception, and lexicon/grammar.

(3) Speaker-listener interactions



In brief, an interaction between the agents proceeds as follows: Agent A picks up a lexical item from the lexicon and produces it as sequences of overlapping articulatory targets (as described below). The acoustic signal resulting from the production is presented to the listener/learner, Agent B. Whether correctly recovered or misidentified,

the items are stored in the learner’s “lexicon.” Further generalizations across the recovered items and the inventory in general are also assumed, but are not implemented in the current simulation. Subsequently, Agent B produces an item from his/her lexicon and adjusts the item’s representation based on the communicative success (see de Boer 2000) and additional tokens of this item. The initial part of the interaction, Agent A’s production and Agent B’s perception, is of particular interest to us, since it is here where most errors are likely to occur. (This part of the interaction is currently implemented using *Matlab*.)

The details of production and perception modules draw heavily from de Boer’s agent-based simulation of emergent vowel inventories (2000). These modules, together with the lexicon/grammar component, are described in more detail below.

3.3 Production module

The production module models targets of vowels and secondary articulations — backness, height, and rounding — based on articulatory synthesizer by Maeda (1989). Each articulatory target is assumed to be [0] or [1], where [1] denotes the targets “front”, “high”, or “rounded”, and [0] denotes the opposite specifications: “back,” “low,” and “unrounded” (4). An articulation of each segment is modeled as a vector of these numbers. The segments of interest have the same height but contrast in backness and rounding. Vowels and the corresponding secondary articulations are specified the same way.

(4) Vowels and Consonants (secondary articulation)

		Backness	Height	Rounding
/i/, /C ^j /	=	[0	1	0]
/y/, /C ^ɥ /	=	[0	1	1]
/ɯ/, /C ^ɰ /	=	[1	1	0]
/u/, /C ^w /	=	[1	1	1]

Words are modeled as vectors of articulatory targets for each segment (cf., Liberman 2002), that is, as matrices of the digits 0 and 1. Thus the lexical item /C^juC^j/ is represented as in (5).

$$(5) \quad \begin{array}{ccc} C^j & u & C^j \\ \left[\begin{array}{ccc} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{array} \right] \end{array}$$

It is important to note that representing articulatory targets with discrete values (0 or 1) does not mean that their realization is also discrete. First, achievement of articulatory targets in humans is never perfect, and this fact is captured in the simulation by adding “articulatory noise,” random fluctuations within the range of ± 0.25 . In other words, the backness target for /u/, which is specified for [1], can be realized during the production as any value between [0.75] and [1]; similarly, the same parameter for /i/, which is specified for [0], can be realized as any value between [0] and [0.75]. Second, articulatory gestures involved in the production of lexical items are subject to overlap, or co-production (Browman & Goldstein 1986), which tends to result in an “undershoot” of their targets (Lindblom 1963, 1988). This is particularly true when two almost simultaneously activated gestures have conflicting targets, such as tongue body backing for /u/ ([1]) and

tongue body fronting ([0]) for /C^j/ in C^juC^j. Thus the main cause of undershoot is purely dynamic: there are physical limits on how well articulators can attain their targets.

This view of gestural overlap is consistent with phonetic accounts of languages with secondary articulations: an achievement of the secondary articulation targets leads to a remarkably different quality of adjacent vowels (e.g., Russian: Bolla 1981, Kochetov 2002; Marshallese: Choi 1992; Irish: Ó Dochartaigh 1992). Thus, /u/ is substantially fronted after palatalized consonants and /i/ is backed after velarized segments. The reverse is often observed in languages with multiple front-back vowel contrasts: an attainment of vowel targets results in allophonic velarization or palatalization of adjacent consonants (e.g., Turkic languages: Comrie 1981: 63).

In this model it is assumed that when two gestures with conflicting articulatory targets overlap, only one of them achieves the target completely (an undershoot rate $u = 0$), while the target of the second gesture is always undershot.^{iv} The degree of undershoot is set up to be 0, 0.25, and 0.50. The first one ($u = 0$) is highly unlikely to be observed in natural speech; it is used in the simulation as a starting point. The other two degrees of overlap are likely to be more typical, at least of casual and fast speech.^v

A 25% undershoot of the vowel target of /C^juC^j/ is shown in (6a). Thus backness and rounding parameters for /u/ are reduced from [1] to [0.75], since the near-simultaneous secondary articulation targets are specified for [0] (the first and third rows). There is no reduction in height, since all three targets are specified for [1]. The same degree of undershoot of the consonant secondary articulation targets is shown in (6b), where we can see a 25% shift to a more back and rounded articulation of /C^j/.^{vi}

$$\begin{array}{l}
 (6) \quad \text{a.} \quad \begin{array}{ccc} C^j & u & C^j \\ \left[\begin{array}{ccc} 0 & 0.75 & 0 \\ 1 & 1 & 1 \\ 0 & 0.75 & 0 \end{array} \right] \end{array} \\
 \\
 \text{b.} \quad \begin{array}{ccc} C^j & u & C^j \\ \left[\begin{array}{ccc} 0.25 & 1 & 0.25 \\ 1 & 1 & 1 \\ 0.25 & 1 & 0.25 \end{array} \right] \end{array}
 \end{array}$$

3.4 Signal

An acoustic signal resulting from the production is calculated based on Vallée 1994 (as reported in de Boer 2000). Only first and second formants are used in the analysis (F1 and F2, in Hertz). These formants for our vowels and secondary articulations are shown in (7). In order to ensure acceptable perceptual quality all lexical items were synthesized using *Synthworks*, an acoustic synthesizer.

(7) Formants of vowels and secondary articulations (Hertz)

		F1	F2
/i/, /C ^j /	=	[252	2202]
/y/, /C ^y /	=	[250	1878]
/ɯ/, /C ^y /	=	[305	1099]
/u/, /C ^w /	=	[276	740]

Acoustic noise, as random fluctuations of formants within certain ranges (± 100 Hz for F1 and ± 200 Hz for F2), is added to the signal. Adding noise is expected to make the learning situation closer to real-life acquisition, where lexical items are hardly ever acquired in complete silence.

3.5 Perception module

The resulting signal is presented to Agent B, the listener/learner. The listener's recovery of items from the signal involves extracting formants at 3 points in time (the onset, midpoint, and offset of the vowel), converting them to an auditory scale (in Barks; see de Boer 2000 for details), and matching the output to the available vowel and consonant categories, shown in (8). This is achieved by calculating a Euclidean distance from each of the categories. For example, if a part of the signal is identified as having the values $F1 = 3.08$ Barks and $F2 = 9.98$ Barks, it is labeled as /ʉ/ (or /Cʏ/), since this category is the closest to the recovered signal (a distance of 0.48 Barks; compare to 1.54 Barks for /y/ or /Cʏ/, 1.76 Barks for /u/ or /Cʷ/, and 2.09 Barks for /i/ or /Cʲ/).

(8) Formants of vowels and secondary articulations (Barks)

		F1	F2
/i/, /Cʲ/	=	[2.52	13.65]
/y/, /Cʏ/	=	[2.50	12.59]
/ʉ/, /Cʏ/	=	[3.08	9.10]
/u/, /Cʷ/	=	[2.78	6.82]

Obviously, the categories are not perceptually equidistant. While /i/ and /ʉ/ are fairly close to their corresponding rounded vowels, /y/ (0.58 Barks) and /u/ (1.29 Barks), the distance between, for example, /y/ and /ʉ/ or /u/, is substantially higher (2.00 and 3.17 Barks respectively). For simplicity the shape of lexical items CVC, where $C1 = C2$, is assumed to be known to the learner.

3.6 *Lexicon and grammar*

The limitations of articulation — overlap of gestures, with additional articulatory and acoustic noise — have important consequences for perception and, ultimately, for the lexicon and the grammar. Lexical items produced by the speaker with undershot targets may not always be perfectly perceived by the listener/learner. As a result of perceptual confusion, some items will end up being represented in the lexicon of Agent B differently from those of Agent A. As discussed in the next section, an instance of $C^j u C^j$ with a vowel undershoot can be interpreted as $/C^j u C^j/$, $/C^j u C^j/$, etc.; while the same item with a consonant undershoot is likely to be perceived as $/C^j u C^j/$, $/C^q u C^q/$, etc.

All tokens of a particular lexical item are temporarily stored in the lexicon and are used in the calculation of its abstract representation. This representation consists of the segments most frequently occurring in the stored tokens. Suppose that tokens for a particular lexical item vary in the quality of the vowel: 65 out of 100 tokens have /u/, 25 have /y/, and 10 have /u/. The most common vowel among these, /u/ will be the one used in the lexical representation. It is assumed here that the agent's grammar, or rankings of constraints, is constructed based on the acquired lexical items. The mechanism of this ranking is not explored in the simulation (but see Kochetov 2002).

4. Results

In this section I describe the results of the simulation by examining the results under the conditions of three different degrees of undershoot. The first case involves no overlap of vocalic targets and thus no undershoot. As already mentioned, this is an unrealistic

situation, but it serves as a baseline for the other cases. The second case involves a certain degree of overlap of targets, and as a consequence, an undershoot of one of them by 25%. The third case presents a substantial overlap of the targets which leads to a 50% undershoot of one of them. This degree of undershoot is likely to be typical of fast casual speech. A sample run, perception of the item /C^juC^j/ based on 100 produced tokens, is presented in (III) in the Appendix.

In each case, undershoot of vowels and secondary articulations is considered separately. Recall that in each case the goal of the learner, Agent B, is to build a lexicon based on perceived tokens. This lexicon may or may not turn out to be identical to the lexicon of Agent A.

4.1 No undershoot

Running the simulation with no undershoot of targets results in a very high degree of success on the part of Agent B in replicating the target lexicon (see (2)). There is a very high probability that all the lexical items are perceived and stored correctly. We can see this from the sample run for /C^juC^j/ in (III) in the Appendix. In a few instances the listener confuses perceptually similar vowels (/i/ and /y/; /ʉ/ and /u/) and similar secondary articulations (/C^j/ and /C^q/; /C^y/ and /C^w/). Yet given a high number of presented tokens per each words (100), the errors are unlikely to influence the learner's choice of the underlying form. Given these results of perception, Agent B will posit the underlying form /C^juC^j/, which is identical to that of Agent A.^{vii}

Overall, the “perfect” production ensures the near-perfect transmission of the lexicon from Agent A to Agent B. I now turn to a more realistic production that involves overlap of gestures and undershoot of targets.

4.2 Undershoot of 25%

The results show that a 25% undershoot of all vowel targets has important consequences for perception. I first consider the situation when the vowel target is undershot, while the consonant secondary articulation target is fully achieved. Under these circumstances, the back rounded /u/ between secondary front articulations, /C^j/ and /C^q/, is perceived by Agent B as either /u/ or the back unrounded /u̯/. The second variant, however, is more frequent (see (III) in the Appendix). Similarly the front unrounded /i/ between secondary back articulations is perceived most often as the front rounded /y/, rather than as /i/. Given this tendency, the most likely lexicon of Agent B (9) will fail to distinguish between /i/ and /y/ and between /u̯/ and /u/ in certain environments, leading to the virtual reduction of the vowel contrasts from 4 to 3. At the same time the contrast in secondary articulation remains intact.

(9)

C ^j iC ^j	C ^j yC ^j	C ^j u̯C ^j	*C ^j uC ^j
C ^q iC ^q	C ^q yC ^q	C ^q u̯C ^q	*C ^q uC ^q
*C ^y iC ^y	C ^y yC ^y	C ^y u̯C ^y	C ^y uC ^y
*C ^w iC ^w	C ^w yC ^w	C ^w u̯C ^w	C ^w uC ^w

The second situation involves a full achievement of vowel target while the consonant secondary articulation target is undershot by 25%. In this case Agent B fails to correctly identify secondary palatal articulation in the environment of back vowels: /C^j/ is commonly perceived as /C^q/. The same applies to the secondary labial articulation in the environment of front vowels. The resulting lexicon (10) will distinguish between 4 vowels and will fail to distinguish between /C^j/ and /C^q/ and between /C^w/ and /C^y/ in certain contexts, thus leading to the reduction of consonant contrasts from 4 to 3.

(10)

C ^j iC ^j	C ^j yC ^j	*C ^j uɯC ^j	*C ^j uC ^j
C ^q iC ^q	C ^q yC ^q	C ^q uɯC ^q	C ^q uC ^q
C ^y iC ^y	C ^y yC ^y	C ^y uɯC ^y	C ^y uC ^y
*C ^w iC ^w	*C ^w yC ^w	C ^w uɯC ^w	C ^w uC ^w

4.3 Undershoot of 50%

Now we will see how increasing the degree of overlap and the degree of undershoot of targets may further affect perception and the resulting lexicon.

The results show that a 50% undershoot of the vowel targets leads to a higher perceptual error rate than in the previous case and thus to a lexicon dramatically different from the original one. The most likely outcome is shown in (11). The front vowels /i/ and /y/ are in complementary distribution, with /i/ occurring only between palatalized consonants. The back vowels /uɯ/ and /u/ are also in complementary distribution, with /u/ restricted to the environment between labialized consonants. Interestingly, /u/ between palatalized consonants is often considered perceptually similar to /y/ (see (III) in the

Appendix).^{viii} Similarly, we find frequent perception of /i/ between labialized or velarized consonants as /ɯ/.

Overall, the contrasts in vowels are reduced to the distinction between [front] and [back]. All four consonants are found in the lexicon, although with certain positional restrictions.

(11)

$C^j i C^j$	$*C^j y C^j$	$C^j \text{ɯ} C^j$	$*C^j u C^j$
$*C^q i C^q$	$C^q y C^q$	$C^q \text{ɯ} C^q$	$*C^q u C^q$
$*C^y i C^y$	$C^y y C^y$	$C^y \text{ɯ} C^y$	$*C^y u C^y$
$*C^w i C^w$	$C^w y C^w$	$*C^w \text{ɯ} C^w$	$C^w u C^w$

Ultimately, the grammar based on these lexical forms would maintain the contrast between multiple secondary articulations $\{C^j C^q C^y C^w\}$ (although limited positionally) and differentiate vowels only based on the front/back dimension, $\{y \text{ɯ}\}$ or $\{i u\}$.

The same degree of undershoot of secondary articulation targets results in the contrasts in vowels being fairly well maintained, while the contrast between the consonants becomes highly restricted (12). There is no contrast between secondary rounded and unrounded articulations for both front ($/C^j/$ vs. $/C^q/$) and back ($/C^y/$ vs. $/C^w/$) tongue positions. The quality in terms of rounding/unrounding of a consonant is predictable from the neighboring vowel environment: $/C^j/$ occurs only in the context of /i/ and $/C^q/$ is found elsewhere. Similarly, $/C^w/$ occurs in the environment of /u/ and $/C^y/$ is found in all the other vowel environments.

(12)

$C^j i C^j$	$*C^j y C^j$	$*C^j u C^j$	$*C^j u C^j$
$*C^q i C^q$	$C^q y C^q$	$C^q u C^q$	$C^q u C^q$
$C^v i C^v$	$C^v y C^v$	$C^v u C^v$	$*C^v u C^v$
$*C^w i C^w$	$*C^w y C^w$	$*C^w u C^w$	$C^w u C^w$

Thus the grammar constructed based on this lexicon would differentiate a full range of vowel contrasts {i y u u} (although restricted positionally), and distinguish consonants by their front or back secondary articulation, {C^q C^v} or {C^j C^w}.

4.4 Summary

The main result of the simulation is that a grammar such as the target grammar in (13a) that allows multiple contrasts in backness and rounding both in vowels and secondary articulations, is highly unstable because it cannot be well replicated by the learner. Recall that perceptual confusion of vowels and secondary articulations distinguished solely by lip rounding is not uncommon even when their targets are fully achieved. This confusion increases substantially in more natural speech, when gestures overlap in time and their targets are undershot. As we saw, there are certain attractors, default states, at which the grammar naturally arrives. The first one, default grammar 1 (13b), allows multiple secondary articulation contrasts at the expense of vowel distinctions. The second grammar, default grammar 2 (13c), limits secondary articulation contrasts, while maintaining multiple vowel distinctions. These more stable grammars are exhibited by the majority of the language in our typological survey: languages tend to have either contrastive secondary articulations or front/back rounded/unrounded contrasts in vowels.

Note that a grammar that limits both secondary articulations (e.g., only “plain” consonants) and vowel contrasts (e.g., only front vs. back distinction) (13d) is likely to be even more stable in terms of production and perception. As we know, this is the state of affairs characteristic for most of the world’s languages: 70% of the UPSID languages have neither (surface) secondary articulation contrasts, nor rounding contrasts in front or back vowels.

- (13) a. Target grammar
- Multiple secondary articulation contrasts
 - Multiple vowel contrasts
- b. Default grammar 1
- Multiple secondary articulation contrasts
 - Limited vowel contrasts
- c. Default grammar 2
- Limited secondary articulation contrasts
 - Multiple vowel contrasts
- d. Default grammar 3
- Limited secondary articulation contrasts
 - Limited vowel contrasts

This simulation has allowed me to explain one of many phonological markedness phenomena observed in language. A number of questions related to the results require further consideration. First, the lexicon discussed here is a result of initial processing, based on 100 tokens. The learner is likely to restructure this lexicon based on subsequent communication as well as by making certain generalizations over segments and environments. The pressure to avoid homophony, which is as high as 50% in our case, is also likely to affect the process. Second, it is likely that the choice of segments in variable cases (e.g., /i/ or /y/ and /u/ or /u/) is influenced by other factors, namely, the general

preference for /i/ and /u/ over /y/ and /ɯ/, which appear to result from more complex long-term interactions (see de Boer 2000) and possibly from other factors. Third, positing an a priori set of phonemic categories made the simulation more manageable by restricting the choices of the learner. More realistically, it would not be surprising if the vowel and secondary articulation categories constructed by the learner based on the highly variable input were not identical to those of the speaker^{ix} (see Liberman 2002 on modeling of word pronunciation in populations of agents). Finally, further work should aim to rely on a more complex interactions and a more realistic model of human speech production and perception. It should use a wider range of lexical items and give more attention to higher-level processing of the perceived input.

5. Conclusion

In this paper I have attempted to demonstrate that an investigation of low-level speaker-listener interactions provides insight into the causes of phonological markedness (cf. (Ohala 1981, Kawasaki 1982, de Boer 2000, among others). Apparent restrictions on co-occurrence of certain vowel and secondary articulation contrasts in language inventories can be generated in a simulated environment with no *a priori* knowledge of markedness. No “innate” restrictions against having both types of contrasts in inventories need to be assumed, since such a system is highly unstable due to limitations on articulation and perception. A language having this system will inevitably “self-organize” by shifting to a more stable pattern: with either rounding contrasts in the vowels, or secondary articulation contrasts in the consonants, or none of these marked contrasts.

References

- Avanesov, R.I. 1972. *Russkoe literaturnoe proiznoshenie*. Moscow: Prosveshchenie.
- Ball, M.J., and J. Fife (eds.). 1993. *The Celtic languages*. New York, N.Y.: Routledge.
- Bolla, K. 1981. *A conspectus of Russian speech sounds*. Cologne: Böla.
- Brooks, R. A. 1991. Intelligence without representation. *Artificial Intelligence Journal* (47). 139–159.
- Browman, C. P., and Goldstein, L. 1986. Towards an articulatory phonology. *Phonology Yearbook*, 3. 219-252.
- Browman, C. P., and Goldstein, L. 1990. Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, 18. 299-320.
- Browman, C. P., and Goldstein, L. 1999. Competing constraints on intergestural coordination and self-organization of phonological structures. In: *Bulletin de la Communication Parlée. Grenoble*.
- Byrd, D. 1996. Influences on articulatory timing in consonant sequences. *Journal of Phonetics* 24(2). 209-244.
- Choi, J.D. 1992. Phonetic underspecification and target-interpolation: An acoustic study of Marshallese vowel allophony. *UCLA Working Papers in Phonetics* 82.
- Clements, G.N. 1985. The geometry of phonological features. *Phonology Yearbook* 2. 223-252.
- Comrie, B. 1981. *The languages of the Soviet Union*. Cambridge: Cambridge University Press.
- Comrie, B., and G.G. Corbett (eds.). 1993. *The Slavonic languages*. London, New York, N.Y.: Routledge.
- de Boer, B. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28. 441-465.
- Harris, M. and N. Vincent (eds.). 1988. *The Romance languages*. London: Croom Helm.
- Harrison, K. D., M. Dras, and B. Kapicioglu. (to appear). Agent-based modeling of the evolution of vowel harmony. *Proceedings of NELS* 32, New York, N.Y., October 2001.
- Hume, E., and K. Johnson. 2001a. A model of the interplay of speech perception and phonology. In: *The role of speech perception in phonology*, eds. E. Hume and K. Johnson. New York, N.Y.: Academic Press. 3-26.
- Iartseva, V.N. 1993. *Iazyki mira: Ural'skie iazyki*. Moscow: Nauka.
- Iartseva, V.N. et al. (eds.). 1997. *Iazyki Rossiiskoi federatsii i sosednix gosudarstv*. Enciklopediia. Vol. 1. eds. Moscow: Nauka.
- Jakobson, R. 1971. K kharakteristike evraziiskogo yazykovogo soyuza. In: *Roman Jakobson: Selected Writings*, vol. 1: Phonological Studies. The Hague: Mouton. 144-201.
- Kauffman, S. 1995. *At home in the universe: The search for the laws of self-organization and complexity*. Oxford: Oxford University Press.
- Kawasaki, H. 1982. An acoustic basis for universal constraints on sound sequences. Doctoral dissertation, University of California, Berkeley.
- Kochetov, A. 2002. *Production, perception, and emergent phonotactic patterns: A case of contrastive palatalization*. New York, London: Routledge.
- König, E. and J. Van der Auwera (eds.). 1994. *The Germanic languages*. London: Routledge.
- Ladefoged, P., and I. Maddieson. 1996. *The sounds of the world's languages*. Cambridge, MA: Blackwell.
- Langton, C. 1995. *Artificial Life: An overview*. MIT Press, Cambridge, MA.
- Liberman, M. 2002. Simple models for emergence of a shared vocabulary. Paper presented at the Eighth Conference on Laboratory Phonology (LabPhon VIII), June 27-30, 2002, Yale University and Haskins Laboratories.
- Lindblom, B. 1963. Spectrographic study of vowel reduction. *JASA* 35. 1773-1781.
- Lindblom, B. 1988. Phonetic invariance and the adaptive nature of speech. In: *Working models of human perception*. London: Academic Press.
- MacAulay, D. (ed.) 1992. *The Celtic languages*. New York, N.Y.: Cambridge University Press.
- Maddieson I. and Precoda K. 1990. Updating UPSID. *UCLA Working Papers in Phonetics* 74. 104-120.
- Maeda, S. 1989. Compensatory articulation during speech: Evidence from the analysis and synthesis of tract shapes using an articulatory model. In: *Speech production and speech modeling*, eds. Hardcastle, W.J. and A. Marchal. Dordrecht: Kluwer. 131-149.

- Ó Dochartaigh, C. 1992. Irish. In: *The Celtic languages*, ed. D. MacAulay. New York, N.Y.: Cambridge University Press. 11-99.
- Ohala, J.J. 1981. The listener as a source of sound change. In: *Papers from Chicago Linguistic Society parasession on language and behavior*, C.S. Masek, R.A. Hendrick, and M.F. Miller. Chicago, IL: CLS. 178-203.
- Perrier, P., H. Loevenbruck, and Y. Payan. 1996. Control of tongue movements in speech: The Equilibrium Point Hypothesis perspective. *Journal of Phonetics*. 24. 53-75.
- Pfeifer, R., and C. Scheier. 2001. *Understanding intelligence*. Cambridge, MA: MIT Press.
- Pierrehumbert, J., M. E. Beckman, and D. R. Ladd. 2001. Conceptual foundations of phonology as a laboratory science. In: *Phonological knowledge*, eds. N. Burton-Roberts, P. Carr, and G. Docherty, Oxford: Oxford University Press. 273-304.
- Prince, A., and P. Smolensky 1993. Optimality Theory: Constraint interaction in generative grammar. Ms., Rutgers University.
- Steels, L. 1995. Intelligence - dynamics and representations. In: Steels, L. (ed.), *The biology and technology of intelligent autonomous agents*. Berlin: Springer-Verlag.
- Vallée, N. 1994. Systèmes vocaliques: de la typologie aux prédictions. Thèse préparée au sein de l'Institute Communication Parlée (Grenoble-URA C.N.R.S. no. 368).
- Vinogradov, V.A. et al. (eds.). 2001. *Iazyki Rossiiskoi federatsii i sosednix gosudarstv*. Enciklopediia. Vol. 2. eds. Moscow: Nauka.
- Vinogradov, V.V. et al. (eds.). 1966a. *Iazyki narodov SSSR. Vol. 1. Indoeuropeiskie iazyki*. Moscow: Nauka.
- Vinogradov, V.V. et al. (eds.). 1966b. *Iazyki narodov SSSR. Vol. 3. Finno-Ugorskie i samodiiskie iazyki*. Moscow: Nauka.

Appendix

(I) UPSID (Maddieson & Precoda 1990) languages that have consonants with secondary articulations and high vowels /y/ and /ɥ/ (in addition to /i/ and /u/)

a. Languages with stops with secondary articulation and no contrastive rounding in high vowels (81):

- palatalized and labialized/velarized stops: Irish, Lakkia, Kam, Lai, Kabardian, Igbo, Hausa, Tera, Amuzgo, Tsimshian, Nambakaengo;
- velarized and labialized stops: Chipewyan*;
- palatalized stops only: Lithuanian, Russian, Bulgarian, Saami, Nenets, Resigaro*, Ocaina*;
- labialized stops only: Sui, Lenakel, Pohnpeian, Kwaio, Taishan, Lak, Rutul, Archi, Kpelle, Kohumono, Konyagi, Kolokuma Ijo, Amharic, Awiya, Iraqw, Beja, Ngizim, Dahalo, Hadza, Haida, Tlingit, Navajo, Huave*, Mixtec, Tseshah, Kwakw'ala, Quileute, Lushootseed, Luiseno, Hopi, Picuris, Diegueno, Zuni, Tonkawa, Wiyot, Wichita, Nahuatl, Bella Coola, Upper Chehalis, Caddo, Huasteco, Shuswap, Southern Nambiquara, Yupik, Kwoma, Guarani, Ticuna, Siona, Iranxe, Tarascan, Warao, Paya, Cuna, Movima, Saliba, Guambiano, Yupik, Kwoma, Dani, Wantoat, Yessan Mayo.

* languages with /ɥ/ instead of /u/ (i.e., no contrast in rounding).

b. Languages with contrastive rounding in high vowels and no secondary articulation in stops (47):

- /y/ and /ʉ/: Turkish, Chuvash, Yakut, Korean, Naxi;
- /y/ (or /ʏ/) only: Breton, German, Norwegian, French, Albanian, Finnish, Hungarian, Nganasan, Azerbaijani, Kirghiz, Bashkir, Tuva, Dagur, Iai, Mandarin, Changzhou, Fuzhou, Ejagham, Tzeltal, Huari;
- /ʉ/ (or /UU/) only: Khanty, Komi, Vietnamese, Khmer, Parauk, Sre, Nikobarese, Nyah Kur, Bruu, Yay, Lungchow, Bai, Dafla, Ao, Tulu, Aizi, FeʔFeʔ, Karib, Apinaye, Jivaro, Araucanian, Panare;

c. Languages with both stops with secondary articulation and contrastive rounding in high vowels (6):

- labialized stops, /y/ and /ʉ/: Highland Chinantec (/k^w/);
- labialized stops and /ʉ/: Lue (/k^w/), Kawaiisu (/k^w/), Mbabaram (/g^w n g^w /);
- palatalized stops and /y/: Mari (/p^j b^j t^j/), Selkup (/t^j/);

(II) Languages of Europe that have consonants with secondary palatal articulation and/or high or mid front rounded vowels; sounds in parentheses = status disputed, marginal, or not realized as “true” palatalized consonants”; short/lax vs. long/tense distinction is ignored

	Language	Group, family	Palatalized consonants			Front vowels	
			labial	coronal	velar	high	mid
1	Belorussian	Slavic, IE	p ^j b ^j	(ts ^j dz ^j)	k ^j g ^j		
2	Bulgarian	Slavic, IE	p ^j b ^j	t ^j d ^j	k ^j g ^j		
3	Irish	Celtic, IE	p ^j b ^j	(tʃ dʒ/ tʃ dʒ)	k ^j g ^j		
4	Lithuanian	Baltic, IE	p ^j b ^j	t ^j d ^j (or tʃ dʒ)	k ^j g ^j		
5	Roma	Indo-Aryan, IE	p ^j b ^j	t ^j d ^j t ^h j	k ^j g ^j k ^h j		
6	Saami (Eastern)	Finno-Ugric, Uralic	p ^j b ^j	t ^j d ^j t ^h j d ^h j	k ^j g ^j		
7	Russian	Slavic, IE	p ^j b ^j	t ^j d ^j	k ^j g ^j		
8	Scots G.	Celtic, IE	(p ^j b ^j)	(tʃ dʒ/ tʃ dʒ)	(c ʃ)		
9	Manx	Celtic, IE		(tʃ dʒ)	k ^j g ^j		
10	Nenets	Samoyed, Uralic	p ^j b ^j	t ^j d ^j			
11	Polish	Slavic, IE	(p ^j b ^j)	(tʃ dʒ)	(k ^j g ^j)		
12	Upper Sorbian	Slavic, IE	p ^j b ^j	(tʃ dʒ)			
13	Lower Sorbian	Slavic, IE	p ^j b ^j	(ʧ ʒ)			
14	Liv	Finno-Ugric, Uralic		t ^j d ^j			
15	Erzya Mordva	Finno-Ugric, Uralic		t ^j d ^j			
16	Moksha Mordva	Finno-Ugric, Uralic		t ^j d ^j			
17	Ukrainian	Slavic, IE		t ^j d ^j			
18	Yiddish	Germanic, IE		(t ^j d ^j)			
19	Czech	Slavic, IE		(c ʃ)			
20	Slovak	Slavic, IE		(c ʃ)			

21	Karaim I	Turkic, Altaic	p ^j b ^j	t ^j d ^j	k ^j		
22	Gagauz I	Turkic, Altaic	p ^j b ^j	t ^j d ^j	k ^j g ^j		
23	Karaim II	Turkic, Altaic				y	ø
24	Gagauz II	Turkic, Altaic				y	ø
25	Albanian	Albanian, IE				y	
26	Occitan	Romance, IE				y	
27	Bashkir	Turkic, Altaic				y	ø
28	Danish	Germanic, IE					ø hi
29	Faroese	Germanic, IE				y	ø lo
30	Finnish	Finno-Ugric, Uralic				y	ø (œ)
31	Frisian	Germanic, IE				y	ø
33	Gorno-Mari	Finno-Ugric, Uralic				y	œ
33	Hungarian	Finno-Ugric, Uralic				y	ø
34	Icelandic	Germanic, IE				Y	œ
35	Izhora	Finno-Ugric, Uralic				y	œ
36	Mari	Finno-Ugric, Uralic				y	œ
37	Norwegian	Germanic, IE				y	ø
38	Swedish	Germanic, IE				y ₁ (y ₂)	ø
39	Tatar	Turkic, Altaic				y	ø
40	Vod'	Finno-Ugric, Uralic				y	œ
41	French	Romance, IE				y	œ ø
42	Breton	Celtic, IE				y	œ ø
43	Dutch	Germanic, IE				y	œ ø
44	German	Germanic, IE				y Y	œ ø
45	Karelian	Finno-Ugric, Uralic	(p ^j b ^j)	t ^j d ^j	(k ^j g ^j)	y	œ
46	Veps	Finno-Ugric, Uralic	(p ^j)	t ^j d ^j	(k ^j g ^j)	y	œ
47	Estonian	Finno-Ugric, Uralic		t ^j d ^j		y	œ
48	Chuvash	Turkic, Altaic		t ^j		y	ø

Additional characteristics:

- Vowel harmony: Finnish, Karelian, Estonian (not productive), Veps (not productive), Izhora, Vod', Mari, Gorno-Mari, Chuvash, Karaim II, Gagauz II, Bashkir, Tatar.
- Umlaut: Icelandic, Norwegian, Swedish, German, ... ?

(III) Perception of the item /C^juC^j/, a sample run based on 100 tokens; the numbers indicate Agent B's "responses" separately for vowels and consonants (C1 = C2); the highest numbers are given in bold

a. No vowel undershoot; no consonant undershoot

i	y	u	u
0	0	24	76
C ^j	C ^u	C ^y	C ^w
73	27	0	0

b. Vowel undershoot of 25%; no consonant undershoot

i	y	ʊ	u
0	0	87	13
C^j	C^ɥ	C^ʏ	C^w
77	29	0	0

c. Vowel undershoot of 50%; no consonant undershoot

i	y	ʊ	u
0	39	61	0
C^j	C^ɥ	C^ʏ	C^w
81	19	0	0

d. Consonant undershoot of 25%; no vowel undershoot

i	y	ʊ	u
0	0	26	74
C^j	C^ɥ	C^ʏ	C^w
7	89	4	0

e. Consonant undershoot of 50%; no vowel undershoot

i	y	ʊ	u
0	0	29	71
C^j	C^ɥ	C^ʏ	C^w
7	46	54	0

ⁱ 15 of these languages are also listed in the UPSID database.

ⁱⁱ While it is clear that many of the language characteristics described below are results of contact-induced changes (see, for example, Jakobson 1971 on the “Eurasian language union”), certain linguistic constraints are apparently at play and determine what types of contrasts can freely co-exist in a language inventory.

ⁱⁱⁱ These languages seem to have a higher number of basic vowels and diphthongs in general.

^{iv} Although this “either ... or” interpretation of undershoot involves a certain oversimplification, it seems to be a plausible approximation given the language data

reported in the previous paragraph. A more realistic model should allow for degrees of undershoot of both targets, or “blending” of gestures (as in GEST, a computational gestural model: Browman & Goldstein 1990). It is still an empirical question, however, which targets are undershot more than which others in a given language.

^v It is well established that gestures tend to show more overlap and thus more undershoot in fast speech (e.g., Byrd 1992, Lindblom 1988, Perrier et al. 1996).

^{vi} Note that the numbers actually generated by the production module will not be the same due to the addition of articulatory noise.

^{vii} Note that different outcomes (e.g., /C^ju^jC^j/, /C^uu^uC^u/, or /C^uu^jC^u/) are also technically possible but only when the lexical form is based on a very small number of tokens.

^{viii} This perceptual similarity explains the fact that when borrowing words from French or German, Russian renders the original sequences /C/+/y/ as /C^j+/u/ (Avanesov 1972). See also the example from Karaim in Section 1.2.

^{ix} Thus, a vowel of /C^ju^jC^j/ can be represented not only as invariable /u/, /u̯/, or /y/, but also as any intermediate values, such as high central vowels /ɨ/ or /i̯/, or even mid central /ə/ or /e̯/. The exact quality of the secondary articulation of /C^j/ can also vary, with a “plain” consonant a possible outcome.

Alexei Kochetov
Haskins Laboratories and
Simon Fraser University
270 Crown St.
New Haven, CT 06511-6695
U.S.A.
Email: kochetov@haskins.yale.edu