Phonetic, phonemic, and phonological factors in cross-language
discrimination of phonotactic contrasts

Lisa Davidson
New York University

Lisa Davidson, Department of Linguistics, New York University.
The author thanks the NYU PEP Lab and audiences at the Linguistic Society of America and the Acoustical Society of America for feedback on this research, and Tuuli Adams for assistance with data collection. This research was supported by NSF CAREER grant BCS-0449560.

Correspondence concerning this article should be addressed to Lisa Davidson, Department of Linguistics, New York University, 10 Washington Place, New York, NY 10003. E-mail: lisa.davidson@nyu.edu
Abstract

Previous research indicates that multiple levels of linguistic information play a role in the perception and discrimination of non-native phonemes (Best, McRoberts, & Sithole, 1988; Polka, 1991, 1992; Werker & Logan, 1985). This study examines the interaction of phonetic, phonemic and phonological factors in the discrimination of non-native phonotactic contrasts. Listeners of Catalan, English, and Russian are presented with an initial #CC-#CaC contrast in a discrimination task. For the Catalan group, the phonemes and their phonetic implementation were native, but the #CC phonotactics were not. For Russian listeners, the phonemes and phonetic implementation were not native but Russian allows a large number of #CC sequences. For English listeners, none of the phonetics, phonemes, nor phonotactics are native. Two task variables, stimuli length and order of presentation, were also manipulated. Results showed that the Russian listeners were most accurate overall, suggesting that the presence of the phonotactic structure in the listeners’ native language may be more important than either phonemic or phonetic information. The interaction between the task manipulations and the linguistic variables is also addressed.

Keywords: phonotactics, cross-language speech perception, acoustics, phonology, discrimination
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

Cross-language speech perception research has often employed discrimination tasks as a method of understanding how experience with a particular native language affects the processing of non-native elements. Previous studies have shown that the ability of adult listeners to discriminate between non-native phonemes, or between a non-native phoneme and a native one is not the same for all non-native phonemes (e.g., Anderson, Morgan, & White, 2003; Best, McRoberts, & Goodell, 2001; Best et al., 1988; Bohn & Flege, 1990; Cebrian, 2003; Gilkerson, 2005; Harnsberger, 2000b; Hayes-Harb, 2007; MacKain, Best, & Strange, 1981; Mora & Fullana, 2007; Polka, 1991, 1992; Werker & Lalonde, 1988; Werker & Logan, 1985; Werker & Tees, 1984b). Differences have been shown to be due to a variety of factors, such as the particular phonemes being compared, the duration of the interval between stimulus presentation, or the structure of a training phase.

Several theories of cross-language speech perception have been developed to account for many different findings in the literature (Best, 1995; Best & Tyler, 2007; Flege, 1995; Kuhl et al., 2008; Kuhl & Iverson, 1995; Strange, 2006), but these theories have mostly focused on data from single phoneme discrimination. For example, English listeners may be presented with pairs like voiceless dental stop [ta] and retroflex [ʈa] (Werker & Tees, 1984b) or voiceless glottalized velar stop [k’ə] and uvular [q’ə] (Anderson et al., 2003; Werker & Tees, 1984b) and asked to distinguish between them in some kind of discrimination task. On the basis of results from pairings of different consonants or different vowels, researchers have developed hypotheses about why listeners are better at discriminating some sound pairs than others. It is beyond the scope of this paper to provide a comprehensive review of the dominant theories of cross-language speech perception, but we can highlight one theory for illustrative purposes. The
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

Perceptual Assimilation Model (PAM, e.g. Best, 1995; Best & Tyler, 2007) proposes that there are a number of ways in which listeners may perceptually categorize a sound that is unfamiliar to them. For example, Two Category assimilation occurs when two non-native sounds are sufficiently dissimilar from one another and can be assimilated to two separate native categories. Single Category assimilation refers to the situation when two non-native sounds are treated by a listener as equally good or poor instances of a single phoneme. Category Goodness assimilation occurs when a listener recognizes that one sound is a better approximation of a native phoneme than another sound, but they are both assigned to that same phoneme. Other assimilation types are developed for sounds that do not easily fit into any native phoneme categories. In a series of AXB discrimination tasks, Best, McRoberts and Goodell (2001) demonstrated that English speakers accurately discriminated some contrasts in Zulu 90% of the time or more (i.e. mean accuracy for velar aspirated [kʰa] versus ejective [k’a] = 89.4%, for voiceless lateral [ɬɛ] versus voiced lateral [ɮɛ] = 95%), but were considerably worse on another (i.e. mean accuracy for voiced bilabial stop [bu] versus implosive bilabial [ɓu] = 65.9%). Best et al. argued that English listeners treated the velar contrast as a Category Goodness difference, the laterals as a Two Category contrast, and the bilabials as a Single Category assimilation.

In addition to proposals about how listeners map non-native phoneme categories onto the ones already present in their native language, it has also been shown that phonetic implementation and low-level acoustic information also affect the discrimination of non-native phonemes (Best et al., 1988; Polka, 1991, 1992; Werker & Logan, 1985). In an updated exposition of PAM (called PAM-L2), Best and Tyler (2007) emphasize that phonetic goodness-of-fit is a critical component of the perception of non-native speech. For example, Polka (1992)
showed that both English and Farsi listeners showed similar difficulties in distinguishing a Salish glottalized velar stop [k’a] / uvular [q’a] contrast, even though Farsi contains the velar/uvular distinction and English does not. Polka argued that because the phonetic implementation of the glottalized items was different from Farsi, the Farsi speakers were as disadvantaged as the English speakers who do not have the phonemic contrast at all. However, comparing the effects of phonemic and phonetic factors to the contribution of the phonology has received less attention. One way to investigate how the phonology interacts with phonemic or phonetic factors is to examine the perception of sequential strings of sounds, or phonotactics. For example, even in cases where listeners might easily assimilate the phonemes in a string to two different phonemic categories, their perception might be affected by the fact that the string itself is prohibited in their native language (e.g., though Japanese contains both [b] and [z], it has been argued that they perceive the string [ebzo] as [ebuzo], Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999). Similarly, even if both the phonemes and the phonotactics are permissible in a language, it is possible that a non-native phonetic implementation of those sounds will hinder adequate perception of the strings.

Indeed, studies on the perception of non-native phonotactics have shown that even when listeners are tested on stimuli that contain native phonemes, they do not perceive them with perfect accuracy if the sequential ordering of the consonants is not permitted in their native language (Berent, Steriade, Lennertz, & Vaknin, 2007; Davidson, 2007; Dupoux et al., 1999; Dupoux, Pallier, Kakehi, & Mehler, 2001; Hallé, Dominguez, Cuetos, & Segui, 2008; Hallé, Segui, Frauenfelder, & Meunier, 1998; Kabak & Idsardi, 2007; Moreton, 2002; Pitt, 1998). For example, Dupoux et al. (1999) used ABX discrimination tasks to determine whether Japanese listeners could distinguish between phonotactically illegal strings like [ebzo] or [agmi] and their
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

phonotactically possible counterparts [ebuzo] or [agumi]. Results varied depending on experimental manipulations such as a change in talker and the number of stimuli used in the studies, but Dupoux et al. found that Japanese listeners were significantly worse than French listeners on distinguishing the members of the stimulus pairs. Kabak and Idsardi (2007) carried out a similar study with Korean listeners, but showed that accurate perception of the medial consonant sequence depended on whether or not the first consonant in the string was a possible syllable coda. Using a gating task, Hallé et al. (1998) showed that French listeners tended to hear word-initial [dl] and [tl] as [gl] and [kl], respectively. In all of these studies, the phonemes in the strings being tested were possible for the listeners, but decreased accuracy in perceiving the combinations suggests that the phonological system mediates the perception of the sequences.

**Linguistic and task variables in the current study**

The purpose of the current study is to gain a better understanding of how phonological factor interact with both phonemic and phonetic factors in cross-language speech perception. Specifically, we manipulate three linguistic dimensions that may affect the perception of word-initial obstruent-obstruent and obstruent-nasal sequences; these dimensions are enumerated in (1):

(1) a. **Phonological**: Whether or not the phonotactic sequences in question, word-initial (#CC) obstruent-obstruent and obstruent-nasal, exist in the listeners’ native language.

b. **Phonemic**: Whether or not the specific phonemes being tested exist in the listeners’ native inventory.

c. **Phonetic**: Whether or not the particular phonetic implementation details of the stimuli matches the listeners’ native language.
These issues are examined by testing the ability of Catalan, American English, and Russian listeners to distinguish between a #CC- and #CəC-initial contrast on the same set of stimuli containing obstruent-obstruent and obstruent-nasal sequences. Since the stimuli were recorded by a Catalan speaker, for whom these #CC sequences are not attested, the phonotactically contrasting #CC- and #CəC- stimuli were created by splicing the vowel out from the #CəC- stimuli. Although using stimuli spliced in this manner does not provide the most naturalistic tokens, it allows us to control the stimuli such that for each set of participants, there are characteristics that are either phonologically, phonemically and/or phonetically absent in the participants’ native language. For the Catalan participants, the phonemes and phonetic implementation of the stimuli are possible in Catalan since they were recorded by a Catalan speaker, but the initial #CC sequences are not. For the English participants, the #CC stimuli are likewise impossible, but in addition, some of the phonemes present in the stimuli are not in the English inventory and the phonetic realization of many sounds is different than it is in American English. Finally, for the Russian listeners, a large range of obstruent-obstruent and obstruent-nasal #CC sequences are possible, but some of the phonemes are not found in Russian and the phonetic implementation is not exactly the same as Catalan.

The investigation of the discrimination of non-native phonotactics among listeners from three different language backgrounds sheds light on the contribution of phonetic, phonemic, and phonotactic inventories to cross-language speech perception. Specifically, we are interested in studying how discrimination is affected when (1) only the phonotactics are being tested (but not any other phonetic or phonemic characteristic of the input, i.e. the Catalan listeners), (2) both phonotactics and phonetic realization of the phonemes and some of the phonemes themselves are
non-native for the listener (English listeners), and (3) the language generally contains the phonotactic contrast being tested, but other phonetic and phonemic characteristics differ from the language of the stimuli (Russian listeners). The comparison of these factors will help weigh the importance of phonetic, phonemic, and phonotactic levels in the processing of non-native sequences.

A few predictions can be made regarding these factors. If the phonetic implementation of a sequence is most important to the facilitation of accurate discrimination of #CC-#CəC pairs, then the Catalan listeners should demonstrate the most accurate performance. However, if phonotactic permissibility is the most important factor in perception of non-native sequences, then Russian listeners should have the highest accuracy even though some phonemes and the phonetic implementation of Russian differ from the stimuli. Furthermore, if the phonological contribution of phonotactic permissibility overrides the contributions of phonemic and phonetic factors, then we expect that Russian listeners will be just as accurate on the clusters containing phonemes that do not exist in the language as on those that do. English listeners are predicted have the worst accuracy under both cases, since both the phonotactic inventory and phonetic implementation of the stimuli are different from English.

The obstruent and nasal phoneme inventory and phonotactic combinations in Catalan, English, and Russian differ from one another in crucial ways. The obstruents and nasals of each language are shown in Table 1. In Catalan, no obstruent-obstruent or obstruent-nasal clusters are allowed in word-initial position (e.g. *tpane, *tmase). Catalan, like English, does allow some stop and fricative + liquid clusters, such as flor ‘flower’ or prat ‘meadow’, though [s]+obstruent and [s]+nasal sequences are not permitted. In English, [s]+obstruent (e.g. stop), [s]+nasal (e.g.
smile), and marginally \([f]+\)obstruent or nasal cluster (e.g. shtick, schnapps) occur, but no other fricatives or stops can be combined with obstruents or nasals to create onset clusters.

In Russian, a large number of obstruent-obstruent and obstruent-nasal word-initial clusters containing plain (not palatalized) consonants are permitted (e.g. Davidson & Roon, 2008). However, not all of the sequences in this study are possible in Russian. As discussed further in the Materials section, intervocalic voiced stops in Catalan are realized as fricatives. Thus, because the stimuli are created by a Catalan speaker producing a non-word like \([zəðənu]\) and then splicing out the schwa, both the \#CəC- and \#CC- stimuli contain some phonemes and consonant sequences (e.g. \([zð-]\)) that do not exist in Russian. The sequences that are attested versus those that are unattested are shown in Table 2. These types of stimuli were specifically chosen so that we can analyze the difference between the clusters that exist in Russian, and those that contain a non-native phoneme. It should also be noted that in English, singleton voiceless stops are typically aspirated, and voiced stops often have only partial voicing during the closure, whereas Catalan and Russian voiceless stops are unaspirated and voiced stops are usually fully voiced (Petrova, Plapp, Ringen, & Szentgyörgi, 2006).

**TABLE 1**

**TABLE 2**

In addition to the main examination of phonetic, phonemic and phonological factors, three other factors are also manipulated: one is linguistic, and the other two are task variables. First, we examine the effect of the manner of articulation of the phonemes in the cluster on discrimination ability. Berent et al. (2008; 2009; 2007) argue that perception of \#CC words that
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

are not possible in English and discrimination of #CC-#CəC pairs are mediated by universal considerations of sonority sequencing. Sonority is a property of speech sounds that has been difficult to define from a phonetic perspective, but at a gross level, it may pertain to the relative intensity of phonemes to one another (Ladefoged, 2005; Parker, 2002). Thus, stops are often taken to be the least sonorous segments, followed by fricatives, nasals, liquids, glides and vowels, in order of increasing sonority (e.g. Blevins, 1995; Clements, 1990; Foley, 1972; Selkirk, 1984). Sonority sequencing in the stimuli used in this study refers to whether the relative sonority of sequential consonants in syllable onset position rises, falls, or plateaus. Berent et al. showed that discrimination is more accurate when the stimuli have rising sonority onsets (e.g. [bwɪf]/[bəwɪf]) than when sonority is either a plateau (e.g. [ptak]/[pətak]) or falling (e.g. [lbrɪf]/[ləbrɪf]). However, one shortcoming of Berent et al.’s (2008; 2009; 2007) research is that it does not include stimuli that have only subtle changes in sonority sequencing (e.g., stop-fricative clusters). In a production study using fricative-obstruent and fricative-nasal sequences, Davidson (2006a) provides some evidence that difficulties in producing non-native phonotactics may be more sensitive to specific manner of articulation combinations than to sonority sequencing considerations. In this study, five different combinations of obstruent-obstruent and obstruent-nasal sequences are examined to determine whether sonority sequencing considerations hold when more fine-grained combinations of consonants are perceived.

The second variable that is also manipulated is stimulus length. Most cross-language speech perception studies of phoneme discrimination tend to focus on stimuli composed of CV syllables. As a result, while non-native listeners demonstrate less accurate performance than native speakers in the studies that included both populations (e.g., Werker & Tees, 1984a), they are nevertheless often much above chance. For example, in the classic Werker and Tees (1984b)
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

study, English speakers’ ability to distinguish between the Hindi dental [ta] and retroflex [ʈa] and Thompson velar ejective [k’a] and uvular ejective [q’a] contrast was rather good in an AX discrimination task, with average A-prime scores of 0.867 and 0.877, respectively. One reason for relatively high accuracy on these discrimination tasks may be that there is little phonetic material between the contrast being tested. Listeners do not have to retain very much information in short-term memory in order to compare the two stimuli in the pair (cf. Harnsberger, 2000a). While this may assist listeners in performing discrimination tasks, it does not necessarily follow that their performance on these stimuli will reflect how they process input that is longer and more complex than a simple CV syllable.

Studies of the perception of non-native phonotactics usually have longer stimuli, both because of the extra consonant necessary to form a sequence, and because they tend to use multisyllabic words or nonwords (e.g. [tlabdo], Hallé et al., 1998). As a consequence, performance is often decreased compared to the discrimination of single phonemes. However, neither type of study of cross-linguistic perception—phonemes or phonotactics—has varied stimulus length as a factor, despite the fact that these kinds of stimulus variables have a significant impact on how listeners perform on these tasks (Jenkins, 1979; Strange & Shafer, 2008). In the case of single phoneme studies that typically use only CV syllables, it is unclear whether listeners’ accuracy would suffer further if more language-like, multi-syllabic stimuli were used. For stimuli that include sequences of consonants, as in this study, it is unknown whether listeners’ discrimination improves on shorter stimuli relative to longer stimuli, and whether stimulus length interacts with the linguistic variables. If, for example, listeners of all languages are less accurate on longer words than on shorter ones, it would suggest that the material intervening between the sequence of interest prevents listeners from accessing the
acoustic information necessary to distinguish between them, and that this is true regardless of phonological, phonemic, or phonetic factors. Alternatively, listeners could show “floor” or “ceiling” effects, and the influence of the linguistic factors might be evidenced only on short or long stimuli.

The last manipulation examined in this study is the order of presentation of the stimuli. Previous research with adults has shown that accuracy in distinguishing between two stimuli related on some continuum (e.g. [i]~[I], [ba]~[va]) depends on the order in which the stimuli are presented (e.g. Cowan & Morse, 1986; Polka & Bohn, 1996; Repp & Crowder, 1990). In a study of order effects of the perception of [ba] and [va] by Japanese speakers, Tsushima et al. (2003) demonstrated that before any training, Japanese listeners were more accurate in an AX discrimination task when [va] preceded [ba] than the other way around. Tsushima et al. speculate that when a relatively native-like stimulus (in this case, [ba]) is presented first, it activates an established category that then subsumes the next incoming input that is a poor example of the native category. The results that Tsushima et al. (2003) present for [ba] versus [va] suggest that for consonantal contrasts, language-specific considerations factor into the asymmetrical perception.

However, it has been argued that for vowels, universal factors may better account for perceptual asymmetries in discrimination tasks. Polka and Bohn (2003) review a number of infant studies that indicate that discrimination is better when a vowel that is less peripheral in acoustic and articulatory space (higher F1 or lower F2) is presented first followed by a more peripheral vowel (lower F1 or higher F2). They argue that in the reverse order, the more peripheral vowel is a strong attractor that prevents accurate discrimination of the two stimuli. Polka and Bohn suggest that more peripheral vowels are universally more salient, and may be
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

more stable perceptual targets that can guide infants in their acquisition of speech. However, they also note that since they are summarizing work done on infants and there are fewer studies on adults, it is unclear whether language-specific effects would become more influential once the listeners are more proficient users of their native language. In the current study, following Tsushima et al.’s (2003) findings, it is predicted that in the perception of consonant sequences, language-specific phonotactic restrictions make it more likely that listeners will be more accurate when the cluster (the non-native item) is presented first.

Method

Participants

The Catalan participants included 42 psychology students recruited at the University of Barcelona. The participants ranged in age from 19-27. All of the participants were bilingual speakers of Spanish. Thirty-two participants reported speaking Catalan with their parents, and ten participants reported speaking Spanish at home. However, all of the participants have lived in Catalonia all of their lives and those who speak Spanish at home reported learning Catalan at age 3 or younger. All participants received a bilingual education from primary school through university and stated that they were equally comfortable in both languages. None of them reported any history of speech or hearing disorders, or any knowledge of a Slavic language or a language with the consonant clusters used in this study. The participants at the University of Barcelona were given course credit for their participation.

The American English listeners included 30 participants recruited primarily through a posting on Craigslist in New York. They ranged in age from 18-61. Twenty-one participants were from New York or New Jersey, with the remainder reporting Maryland, Arizona,
Massachusetts, California and Nevada as their home states. No participants reported learning another language before 12 years old. No participants reported any history of speech or hearing disorders, or any knowledge of a Slavic language. The American English listeners were paid $10 for their participation.

Thirteen Russian speaking participants were also recruited through Craigslist in New York. They ranged in age from 22-55. Six of the participants were born in Russia, three in Ukraine, one in Latvia, one in Uzbekistan, and two in New York. All of the participants born outside of the United States reported Russian as their first language, with one reporting being bilingual in Ukrainian and another in Latvian. The participant from Uzbekistan moved to New York at eight years old, but no other participant moved to New York before the age of 12. The two participants born in New York are bilingual Russian/English speakers, and both report living in Russian-speaking communities in Brooklyn and speaking Russian with their families and friends. Several participants reported learning languages such as Spanish, French or German after 12 years of age. No participant reported a history of speech or hearing disorders. The Russian listeners were paid $15 for their participation.

**Materials**

The materials consisted of non-words containing 27 different initial consonant sequences and matching consonant-schwa-consonant sequences (e.g. [fm-]/[fəm-], [gz-]/[gəz-], [zð-]/[zəð-]). Following research showing that the combination of manner of articulation affects the accuracy of production of non-native sequences (Davidson, 2006a), it is also possible that such differences extend into perception. Thus, these sequences can be divided into five groups based on the combination of the manners of their consonant sequences: fricative-fricative (FF: [f\textipa{ʃ}], [z\textipa{β}], [z\textipa{o}],
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

[zy]), fricative-nasal (FN: [fm], [fn], [zm], [zn]), fricative-stop (FS: [fk], [ft]), stop-fricative (SF: [d̪], [g̪], [gz], [kf], [ks], [pf], [tf]), and stop-nasal (SN: [bm], [dn], [gm], [gn], [km], [kn], [pm], [pn], [tm], [tn]). These clusters represent all of the initial obstruent-obstruent and obstruent-nasal sequences that can be created from #CəC sequences that do not form words in Catalan. These stimuli were created because we wanted to be able to compare as many types of manner combinations as possible to examine their effects in perception. No stop-stop sequences could be obtained either because voiced stops become fricatives in intervocalic position, or because potential combinations were a word in the short stimuli. The categories are otherwise imbalanced because we did not want to limit our investigation to a small subset of the language, and so we created stimuli from all available combinations of consonants. The imbalance will be addressed in the statistical analysis by using a Type III Sums of Squares in the ANOVAs, which is appropriate when cell sizes are unequal (Searle, 2006).

Each of the 27 consonant combinations were used to make both long and short stimuli. The relevant contrasts for the long words consisted of pairs of stimuli of the form CCaCV/CəCaCV, such as [fmaˈtu]/[fəmaˈtu], [gzaˈbo]/[gəzɑˈbo], [zəˈðɑnu]/[zəˈðɑnu]. The diacritic over the [a] marks stress. The short words were 27 CCa/CəCa pairs such as [fma]/[fəma], [gza]/[gəzɑ], [zəˈdɑ]/[zəˈdɑ]. The CəC- forms of both long and short stimuli were recorded by a phonetically trained female Catalan speaker who read the non-words, which had been transcribed into Catalan orthography and randomized into a list. In Catalan, orthographic “a” is pronounced as a schwa when the syllable is unstressed, and in the absence of orthographic stress marking,

1 The stimuli included 4 other sequences: /sf/, /sm/, /sp/ and /st/. However, these are excluded in any further analyses since they form legal clusters in English and may adversely affect interpretation of the phonological, phonemic and phonetic contributions to the discrimination results.
stressed falls on the penultimate syllable of a three-syllable word. Thus, the orthographic string
\textit{famatu} was read by the Catalan speaker as [f\text{"\text{\textregistered}}\text{\textup{m}}\text{\textup{\text{"u}})]. For the short stimuli, the speaker was told
that stress should fall on the last syllable, ensuring that the first vowel in a string like \textit{fama} would
be pronounced as a schwa. This is a naturalistic stress pattern in Catalan, since infinitive verbs in
the first conjugation class have exactly this form (e.g. [t\text{"\text{\textup{k}}\text{\textup{\text{"a}}}] ‘to stain’, [p\text{"\text{\textup{\text{\textup{l}}}] ‘to peel’}). All of the
stimuli are listed in the appendix.

The CC forms of each pair were created by using Praat to splice out the vocalic portion
between the consonants. Excised schwas were always cut from zero crossings to avoid acoustic
artifacts. Splicing the schwa out was carried out straightforwardly by looking for obvious
landmarks such as the beginning or end of aperiodic frication, the end of a stop burst or the
beginning of silence, or the beginning or end of middle formant attenuation for the nasals. Thus,
the resulting CCaCV/\textit{CaCaCV} and CCa/\textit{CaCa} pairs consisted of exactly the same acoustic
material except for the presence or absence of the schwa. An analysis of the duration of the
schwa in each length condition is shown in Table 2.

\begin{table}[h]
\centering
\caption{Table 3 About Here}
\end{table}

The recordings were done in a sound-treated room using an M-Audio MicroTrack 24/96
and the electret microphone that comes with the MicroTrack. The stimuli were recorded as wave
files onto a compact flash card at 44.1kHz and were resampled at 22kHz.

In addition to the test stimuli, five obstruent-initial CaCV tokens similar in form to the
long stimuli were recorded by an American English speaker for the practice phase. None of these
were the same words used in the test phase. The CCaCV tokens were created by splicing the schwa out using the same method as described for the test stimuli.

**Procedure**

Using the 27 short words and the 27 long words, the AX discrimination task contained 54 same trials and 54 different trials for each length. The same trials consisted of 27 cluster pairs (e.g. [fmátu]/[fmátu] or [fmá]/[fmá]) and 27 schwa pairs (e.g. [fəmátu]/[fəmátu] or [fəmá]/[fəmá]) for each length. The different trials contained both orders of presentation, so each participant heard 27 cluster-schwa pairs (e.g. [fmátu]/[fəmátu] or [fmá]/[fəmá]) and 27 schwa-cluster pairs (e.g. [fəmátu]/[fmátu] or [fəmá]/[fmá]) for each length. There were 108 trials for each of the short and long stimuli, for a total of 216 trials for each participant. The stimuli were blocked into a short condition and a long condition, and these were counterbalanced so half of the participants heard the short words first and half heard the long words first. The practice stimuli consisted of five different trials (three cluster-schwa, two schwa-cluster pairs) and five same trials (three schwa-schwa, two cluster-cluster). The experiment was run using ePrime.

Catalan and English participants were given the following instructions in their native language, and the Russian listeners were given the instructions in English since they were all proficient in reading English: “In the following task, you will hear sound files presented as pairs of words. In some of the pairs, the sound files that you hear will be slightly different from one another, and others will be the same. Your task is to decide whether or not the sound files that are played to you are exactly the same. After hearing the second word, decide whether the two sound files are the same or different.” Using the E-Prime button box, participants were told to
press a button labeled “S” (or “I” in Catalan for “identical”) with the index finger if they thought the sound files were the same, and “D” with the middle finger if they thought the sound files were different. Participants were encouraged to answer quickly, and were told to choose either “S”/”I” or “D” even if they were not sure of the answer.

The procedure for each trial was as follows. A crosshair appeared on the screen to alert the participant to the start of the trial, accompanied by the simultaneous presentation of the first sound file of the trial. At the end of the inherent duration of each sound file (i.e. the length of each word), there was a 250ms pause, and then another fixation cross along with the second sound file. Participants could make a response any time after the start of the second sound file. As soon as the participant made a response, there was a 500ms pause and the next trial started. The interstimulus interval (ISI) of 250ms was used to stay consistent with Davidson (2007), who had specifically used a low ISI in order to encourage acoustic-level judgments (Pisoni, 1973; Werker & Logan, 1985).

At the University of Barcelona, up to two Catalan speakers participated at a time in a quiet testing room using PC computers. Sounds were played over Sennheiser headphones. At New York University, American English and Russian participants were seated in individual small, quiet rooms containing PC computers and Sennheiser headphones. Participants were first given 10 practice trials to familiarize themselves with the task; there was no feedback in the practice. In the experimental session, there was a break after the first block of trials.

Results

Accuracy for language, length, order of presentation
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

A mixed model analysis of variance (ANOVA) was conducted to initially determine whether there were significant differences in accuracy for language background, length, and order of presentation. The between-subjects dependent variable was language background (Catalan, English, Russian), and within subjects dependent variables were length (long, short) and order of presentation (cluster-schwa, schwa-cluster). Following previous studies such as Best et al. (2001) and Dupoux et al. (1999), the dependent variable was the mean proportion correct for different trials, which was arcsine transformed. However, results for d-prime, which were very similar to the accuracy results, are reported in the Appendix for each language.

For the accuracy ANOVA, only different trials are included since speakers of all languages were consistently above 90% accurate on same trials. Results showed a main effect of language (F (2, 82) = 13.86, p < .001), length (F (1, 82) = 37.04, p < .001), and order of presentation (F (1, 82) = 19.53, p < .001). The interaction between language and length (F (1, 82) = 6.49, p < .001) and language and order of presentation were significant (F (1, 82) = 7.28, p < .001). The three-way interaction was also significant (F (2, 82) = 8.30, p < .001). Planned comparisons of the main effect of language, which is the most important aspect of this ANOVA, show that Russian listeners were most accurate (M = 75%, s.d. = 2%), followed by Catalan listeners (M = 68%, s.d. = 11%), and then by English listeners (M = 45%, s.d. = 10%) (all comparisons are significant, p < .001).

Further analyses of the effects of length and order of presentation were examined with separate ANOVAs for each language group. Subjects were treated as a random variable. For Catalan listeners, there was a main effect of order of presentation (F (1, 41) = 20.5 p < .001) and a main effect of length (F (1, 41) = 45.02, p < .001). The interaction between length and order of presentation was not significant (F (1, 41) < 1). Accuracy rates for Catalan listeners are shown in
Figure 1a. The results for all conditions, including the same trials, are given in Table 3 for all languages. The main effects indicate that listeners were significantly more accurate on pairs presented in cluster-schwa order and that they were more accurate on short words.

For English listeners, there were main effects of length ($F (1, 27) = 33.55, p < .001$) and order of presentation ($F (1, 27) = 15.38, p < .001$) and a marginal interaction between length and order of presentation ($F (1, 26) = 3.99, p = .056$). Results for English listeners are shown in Figure 1b. Planned comparisons indicate that there is a significant difference for order of presentation within both long word pairs ($p = .029$), and short word pairs ($p < .001$). For length, listeners are significantly more accurate on short word pairs for both the cluster-schwa order and the schwa-cluster order ($p < .001$).

Russian listeners show an effect of length ($F (1, 12) = 7.75, p = .016$), but no effect for order of presentation ($F < 1$) and no interaction of length and order of presentation ($F (1, 12) = 1.51 p = .25$). Results for Russian listeners are shown in Figure 1c. The effect of length shows that Russian listeners are significantly more accurate on short words than on long ones.

In summary, all listeners were significantly more accurate on pairs of short words. Catalan and English listeners were also significantly more accurate on pairs in cluster-schwa order, and Russian listeners showed no significant difference for order. Overall, Russian listeners were the most accurate on all types of trials, followed by Catalan and then by English listeners.

Cluster type accuracy

FIGURE 1 ABOUT HERE

TABLE 4 ABOUT HERE
For each language, an ANOVA was performed including cluster type in the analysis. The independent variables were cluster type (SN = stop-nasal, FN = fricative-nasal, FF = fricative-fricative, FS = fricative-stop, SF = stop-fricative) and length. Order of presentation was omitted from this analysis in order to focus on the variables of greater interest. The dependent variable was accuracy, which was arcsine transformed. Subjects were treated as a random variable. In each case, the effects for length were the same as in the previous analyses, so only the results for cluster type and its interactions are reported here.

For Catalan, there was a significant main effect of cluster type ($F (4, 168) = 10.46, p < .001$), and a significant interaction between cluster type and length ($F (4, 168) = 13.21, p < .001$). A Student-Newman-Keuls post-hoc test for the main effect of cluster type provides the following groupings, in order of accuracy (where cluster types in a group are not significantly different from one another): (1) FN, SN, FF (2) FS, SF. The interaction effects are due to the fact FS and SF are the lowest in accuracy for the long and short conditions respectively. The post-hoc test for the long stimuli shows that the grouping is (1) SN, SF, FN, FF (2) FS. For the short stimuli, the grouping is (1) FF, FN, FS, SN (2) SF. Cluster type results for Catalan listeners for the long and short words are shown in Figure 2a.

English listeners showed a main effect of cluster type ($F (4, 108) = 9.83, p < .001$), and a significant interaction for cluster type and length ($F (4, 108) = 6.68, p < .001$). A Student-Newman-Keuls post-hoc test for the main effect of cluster type provides the following groupings: (1) SN (2) SF, FF (3) FN, FS ($p < .05$). The interaction between cluster type and length is due to different patterns of groupings among the short and long words. Though the pattern for the short words followed the general pattern of cluster type, the post-hoc test for the
long words showed a less detailed grouping, where FN was not significantly different from either FF, SF, or SN (p < .05). Results for English listeners are shown in Figure 2b.

Results for Russian listeners show a main effect of cluster type ($F(4, 49) = 8.08, p < .001$) and a significant interaction between length and cluster type ($F(4, 49) = 7.35, p < .001$). A Student-Newman-Keuls post-hoc test for the main effect of cluster type indicates the following grouping: (1) SN (2) SF, FS (3) FN (4) FF (p < .05). Whereas the long words pattern like the general pattern for Russian, there are no significant differences for the short words for any cluster types. Results for Russian listeners are shown in Figure 2c.

The results for cluster type demonstrate that there are language-specific differences for each language group, which will be discussed further in the general discussion.

**FIGURE 2 ABOUT HERE**

**General Discussion**

*The role of phonetic, phonemic and phonological factors*

Collapsing across both length and order of presentation, Russian listeners were most accurate, followed by Catalan listeners, and then by English listeners. An examination of the relationship between the phonetic implementation and phonemes of each language and the stimuli sheds light on the accuracy level of each group. First, for Catalan listeners, all of the phonetic and phonemic characteristics of the stimuli are native except for the unattested onsets in #CC tokens. Catalan listeners showed an effect of both length and order of presentation, but their performance was above chance for each combination of length and order of presentation (using
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

the binomial test, p < .001 for all conditions, with chance taken to be 50%). The mean accuracy collapsing across length and order of presentation was 68%. The results for the Catalan listeners suggests that while the native implementation of the individual sounds helps them discriminate the pairs with greater than chance accuracy, the absence of the #CC-#CəC contrast in the phonological system of the language causes a greater decrement in performance relative to Russian listeners who do have the contrast.

Like the Catalan listeners, Russian listeners display above chance performance (p < .001) on all of the experimental sequences despite the fact that some phonemes and the phonetic implementation of the sounds in the stimuli are non-native for Russian listeners. The mean accuracy for Russian listeners is 75%, which was shown in the results section to be significantly higher than for the Catalan listeners. This is consistent with the hypothesis that having the phonological contrast between #CC-#CəC is the most influential factor in distinguishing these sequences in perception. In order to bolster that claim, a further analysis of the Russian listeners was carried out to determine whether they performed equally well on clusters that are actually attested in Russian, versus those that are not (including those containing phonemes that are not possible in Russian). An ANOVA including length, order of presentation, and legality in Russian as independent variables was carried out. Results showed that there was no main effect of legality (F < 1), and no significant interactions of legality with any of the other independent variables. Thus, because the Russian phonological grammar generally allows obstruent-obstruent and obstruent-nasal sequences, Russian speakers perform equally well on the contrast in this study whether or not the phonemes exist in Russian.

The numerical results for both Catalan and Russian listeners indicate that if listeners can benefit from the phonotactic legality of the sequence (Russian listeners) or from a combination
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

of phonetic and phonemic characteristics (Catalan listeners), then there is sufficient information for them to perform the task at above-chance accuracy. However, English listeners suffer from phonetic, phonemic and phonotactic differences between their native language and the stimuli: some phonemes in Catalan do not exist in English, and for those phonemes that are considered the same, their physical implementation is different (e.g. unaspirated stops, monophthongal vowels), and #CC sequences are not allowed in English either. For English listeners, only short pairs in cluster-schwa order are above chance (p < .001; short schwa-cluster pairs not different from chance, p = .30, and both long pairs are below chance p < .001).

Despite the advantage of native phonemes and phonetic implementation, Catalan listeners’ performance was compromised by a phonological prohibition on #CC clusters. However, they performed better than English speakers, who were often not above chance. Russian learners were the most accurate of all, but considering that they too were not near 100% accuracy, they were likely globally affected by non-native phonemes and phonetic implementation. This suggests that while phonotactics are most important in predicting discrimination ability in this study, differences in phonemes and phonetic implementation may compound discrimination difficulties, and in the case of English listeners, particularly when the phonotactics are unattested in the listener’s native language. These effects are further amplified when the length manipulation is considered, as discussed in the next section.

The implications of length and order of presentation

Results for length showed that for all three languages, listeners were reliably more accurate on pairs containing short words than long words. These findings are consistent with Werker and Logan (1985), who claim that when task conditions place more memory demands on
the listener, participants are more likely to process the input (single phonemes for Werker and Logan, phonotactic strings in this study) using the phonetic and phonological categories available to them in their own language. This processing strategy would predict that longer stimuli, because they contain more material and possibly because they are more similar to realistic word-like input, should lead to lower accuracy. The fact that the length manipulation affects all three languages suggests that the detrimental effects of all three linguistic factors—non-native phonetic implementation, phonemes, and phonotactics—are exacerbated under greater memory load. Interestingly, the amount of the decrement is similar for all listeners, who show a consistent decrease in accuracy of about 20% for both orders of presentation (see Figure 1). In this case, it does not seem that the memory load interacts particularly with some factors and less so with others; rather, it causes an across-the-board decline in performance for all listeners.

The other main task manipulation in this study involved order of presentation. Following Tsushima et al. (2003), who found that Japanese listeners were less accurate on AX discrimination when the native [ba] preceded the non-native [va] than the other way around, it was predicted that Catalan and English listeners would be less accurate on schwa-cluster (native/non-native) order than on cluster-schwa (non-native/native) order. These results were obtained. The Russian listeners, for whom both #CC and #CəC is phonotactically permitted, did not show an effect of order of presentation.

One explanation for the findings for Catalan and English listeners may be reduced stability of short-term representations for non-native phonological elements. When a native phonotactic sequence, or possible word (such as [fətake]) is perceived, it should be recognized as a possible though non-existent word of English or Catalan. Because #CəC is phonologically
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

attested, the listener can establish a representation of the word that accurately reflects the incoming utterance and is not ruled out by the phonological grammar. However, when a non-native phonotactic sequence (such as [ftake]) is perceived, listeners may only be able to establish a degraded memory representation of the word since it does not contact a legal phonotactic frame of their language. Thus, when a non-native utterance precedes one that is possible in the native language, the closest native phonotactic category (e.g. #CəC) only receives partial activation. When the next token, which is a better exemplar of a native category, is presented, the phonotactic category is then fully activated. The increase in activation between the two stimuli may help the listener to label the pair as different. Conversely, when a possible native token is uttered first as in schwa-cluster (SC) trials, the phonotactic category is fully activated and consequently subsumes the incoming non-native token. In other words, when a possible native representation has already been established, a following non-native input is more likely to be treated as a variant of the possible native word because a possible representation is currently activated. However, if the memory trace of the non-native word is too weak, it will not similarly subsume the native-possible word.

Support for the hypothesis that listeners do not establish an adequate representation for non-native phonotactic structures comes from multiple sources. First, studies of non-word recall have shown that both child and adult participants have much better recall of non-words containing common phonological structures as compared to non-words that have very low phonotactic probability (Gathercole, 1995; Gathercole, Willis, Emslie, & Baddeley, 1991; van Bon & van der Pijl, 1997; Vitevich & Luce, 1999, 2005). Though recall is a different task than discrimination, these studies provide important evidence that the phonotactic characteristics of a non-word contribute to the robustness of the encoding of the stimuli, and help define the
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

relationship between the short-term memory trace and stored phonological information. Most of these authors agree that the influence of phonotactic probability on non-words has a sub-lexical locus, which can be thought of as smaller units such as syllables, disyllables, segments, or features. Vitevich and Luce (1998; 1999) couch their findings in terms of Adaptive Resonance Theory (ART) (e.g. Grossberg, 1986; Grossberg, Boardman, & Cohen, 1997), which posits that resonances are established between items in short term and working memory. In this theory, resonances are a result of both the bottom-up activation caused by the items in working memory and excitatory signals then sent back from short-term to working memory. In the non-word case, it is predicted that there is more resonance between high probability items and the appropriate corresponding sub-lexical chunks than for low probability items, where resonances between low probability items and stored information the memory system will be much weaker.

Though the research on non-word recall typically does not focus on elements that are phonologically prohibited, sequences like /#ft/ or /#gd/ have a probability of zero, which can be treated as the extreme end of the phonotactic probability spectrum. Thus, if we were to frame the results of this study within a theory like ART, we might hypothesize that non-existent items like /#ft/ or /#gd/ cannot activate any perfectly matched sub-lexical chunks, whereas items like /#fət/ or /#gəd/ can activate units that are associated with some amount of resonance. With these units activated when a schwa word is presented first, it will be harder to reject a following cluster stimulus since the information in the cluster stimulus is largely overlapping with the signals sent back from short-term memory. However, if a cluster word does not strongly activate any sub-lexical information, then it will be more easily distinguished from a following schwa word, which does activate sub-lexical information. By extension, this explanation has implications for how phonetic detail is processed in lexical item selection. If a contrast such as /#fət~/#/ft/ exists
in a language, listeners would have to closely attend the acoustic information corresponding to the schwa. However, if a language only allows one of these possibilities, then the production of the other sequence may be treated as a less optimal but potential variant of the phonotactics that do exist.

**Effects of cluster composition**

From the findings of previous research (Berent et al., 2008; 2009; 2007), it would be expected that the results for English and Catalan listeners, whose native phonologies do not contain the onset clusters being tested, should show effects of sonority sequencing on accuracy in discrimination. In the stimuli used in this study, there are sonority rises in stop-nasal and fricative-nasal sequences and a sonority plateau in fricative-fricative sequences. Using a sonority scale that distinguishes between stops and fricatives, stop-fricative is a sonority rise and fricative-stop is a sonority fall (Blevins, 1995; Selkirk, 1982). Thus, if English and Catalan listeners’ performance is subject to the phonological principle of sonority sequencing, they should be most accurate on the stop-nasal (SN), fricative-nasal (FN), and stop-fricative (SF) sequences which are sonority rises, followed by fricative-fricative (FF), and then by fricative-stop (FS). However, the results for clusters which contain much smaller differences in sonority than the ones examined by Berent et al. cannot be explained by sonority factors. Catalan listeners showed a general accuracy pattern of (1) FN, SN, FF (2) FS, SF and the English listeners demonstrated a pattern of (1) SN (2) SF, FF (3) FN, FS. Even the Russian listeners, despite their overall accurate performance, do show some distinctions between the clusters for the long stimuli, but it is not attributable to sonority: (1) SN (2) SF, FS (3) FN (4) FF.
While the insights from the sonority sequencing hierarchy may be consistent with the findings for stimuli that have relatively large distances for sonority rises and falls, the findings from this study raise the possibility that there may be a different explanation for previous results. As it has already been argued that phonetic and phonemic factors affect the perception of non-native sequences, it is plausible that they may also account for specific disparities among the different manner combinations. Another aspect of this research that calls into question the top-down effects of sonority sequencing—presumably a universal phonological principle—in perception is that the listeners of each language group show a different pattern of accuracy. The data from this study alone do not provide enough evidence to conclusively explain the language-specific patterns of the Catalan, English and Russian listeners; to get at this question, a study that systematically examines the phonemic and acoustic properties that the listeners of each language group are sensitive to would be necessary. This is an issue for future research.

**Conclusion**

The main purpose of this study was to gain a better understanding of the relative contributions of phonetic, phonemic, and phonotactic elements in cross-language speech perception. This aim was addressed by examining the perception of listeners from three different language backgrounds—Catalan, English, and Russian—on the same stimuli. Although the consonant cluster stimuli were created from utterances produced by a Catalan speaker, it was the Russian listeners who were the most accurate. This suggests that the presence of the relevant phonological structure in one’s native language is perhaps the most important predictor of discrimination ability, even if the phonetic implementation and some phonemes are not native for the listener. Since Catalan and English do not have initial obstruent-obstruent and obstruent-
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

nasal sequences (with the exception of /s/-initial sequences for English listeners), participants showed a significant decrement in performance compared to the Russian listeners, though Catalan listeners may have benefited from the stimuli being recorded by a native speaker of their language. The comparison of English and Catalan listeners suggests that we might expect differences in perception and the processing of phonetic detail between second language acquisition and the type of loanword adaptation where a small group of initial borrowers passes down non-native phonemes and phonotactics to the larger monolingual community (see Davidson, 2007; Haugen, 1950).

The results of this study also show that the main linguistic factors also interact with task manipulations in interesting ways. As Werker and Tees first observed in 1984, laboratory tasks may not always adequately reflect the perceptual processing that occurs during more naturalistic language learning and language recognition situations. In this study, it was shown that length of the word and order of presentation also interact with discrimination. Perhaps not surprisingly, listeners performed less accurately on long words, likely due to increased memory demands, but this is relevant because listeners are not always going to encounter CV or other monosyllabic stimuli during language acquisition or loanword adaptation. This finding also potentially affects development of theories of cross-language speech perception, since understanding the perceptual categories that listeners use to process non-native speech cannot be independent from the nature of the material containing the non-native element. The effects of order of presentation also underscore that experimental manipulations must be taken into consideration, but more speculatively, this finding may also bear on how learners establish a phonological contrast between a native and a new, non-native element. If lexical items containing non-native phonemes or sequences are acquired first, it may help listeners to more accurately establish a
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

phonological contrast between two phonemes or sequences (e.g. #CC versus #CəC). However, if lexical items containing a native-possible sequence are learned initially, it may block the adequate formation of a phonological contrast and a representation for the non-native counterpart. This is a potential area for future research.

Appendix

Stimuli

Long and short stimuli used in the study. All words were paired with themselves for the same trials and in both orders of presentation for the different trials.

<table>
<thead>
<tr>
<th>Cluster type</th>
<th>Long CəC stimuli</th>
<th>Long CC stimuli</th>
<th>Short CəC stimuli</th>
<th>Short CC stimuli</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>fəsaga</td>
<td>fsaga</td>
<td>fəsa</td>
<td>fsa</td>
</tr>
<tr>
<td></td>
<td>zəβatu</td>
<td>zβatu</td>
<td>zəβa</td>
<td>zβa</td>
</tr>
<tr>
<td></td>
<td>zəɣano</td>
<td>zəɣano</td>
<td>zəɣa</td>
<td>zɣa</td>
</tr>
<tr>
<td></td>
<td>zəðanu</td>
<td>zðanu</td>
<td>zəða</td>
<td>zða</td>
</tr>
<tr>
<td>FN</td>
<td>fənada</td>
<td>fnada</td>
<td>fəna</td>
<td>fna</td>
</tr>
<tr>
<td></td>
<td>fəmatu</td>
<td>fmatu</td>
<td>fəma</td>
<td>fma</td>
</tr>
<tr>
<td></td>
<td>zəmafo</td>
<td>zmafo</td>
<td>zəma</td>
<td>zma</td>
</tr>
<tr>
<td></td>
<td>zənagi</td>
<td>znagi</td>
<td>zəna</td>
<td>zna</td>
</tr>
<tr>
<td>FS</td>
<td>fəpami</td>
<td>fpami</td>
<td>fəpa</td>
<td>fpa</td>
</tr>
</tbody>
</table>
## FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

<table>
<thead>
<tr>
<th>SF</th>
<th>dɔβaka</th>
<th>dβaka</th>
<th>dœba</th>
<th>dβa</th>
</tr>
</thead>
<tbody>
<tr>
<td>gɔdasi</td>
<td>gɔasi</td>
<td>gœda</td>
<td>gœa</td>
<td></td>
</tr>
<tr>
<td>gɔzabo</td>
<td>gzabo</td>
<td>gæza</td>
<td>gza</td>
<td></td>
</tr>
<tr>
<td>kɔfagi</td>
<td>kfagi</td>
<td>kœsa</td>
<td>ksa</td>
<td></td>
</tr>
<tr>
<td>kœsami</td>
<td>ksami</td>
<td>kœfa</td>
<td>kfa</td>
<td></td>
</tr>
<tr>
<td>pœfala</td>
<td>pfala</td>
<td>pœfa</td>
<td>pfa</td>
<td></td>
</tr>
<tr>
<td>tœfano</td>
<td>tfano</td>
<td>tœfa</td>
<td>tfa</td>
<td></td>
</tr>
<tr>
<td>SN</td>
<td>bœmada</td>
<td>bmanda</td>
<td>bœma</td>
<td>bma</td>
</tr>
<tr>
<td>dœnapa</td>
<td>dnapa</td>
<td>dœna</td>
<td>dna</td>
<td></td>
</tr>
<tr>
<td>gœnasu</td>
<td>gnasu</td>
<td>gœma</td>
<td>gma</td>
<td></td>
</tr>
<tr>
<td>gœmalu</td>
<td>gmalu</td>
<td>gœna</td>
<td>gna</td>
<td></td>
</tr>
<tr>
<td>kœnabo</td>
<td>knabo</td>
<td>kœna</td>
<td>kna</td>
<td></td>
</tr>
<tr>
<td>kœmapi</td>
<td>kmapi</td>
<td>kœma</td>
<td>kma</td>
<td></td>
</tr>
<tr>
<td>pœmava</td>
<td>pmava</td>
<td>pœma</td>
<td>pma</td>
<td></td>
</tr>
<tr>
<td>pœnaka</td>
<td>pnaka</td>
<td>pœna</td>
<td>pna</td>
<td></td>
</tr>
<tr>
<td>tœnase</td>
<td>tnase</td>
<td>tœna</td>
<td>tna</td>
<td></td>
</tr>
<tr>
<td>tœmapa</td>
<td>tmapa</td>
<td>tœma</td>
<td>tma</td>
<td></td>
</tr>
</tbody>
</table>
**Results for d-prime**

Mean d-prime values. CS = Cluster-Schwa, SC = Schwa-Cluster. Standard error is in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Catalan</th>
<th>English</th>
<th>Russian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long</td>
<td>CS</td>
<td>2.08 (0.26)</td>
<td>1.39 (0.14)</td>
</tr>
<tr>
<td></td>
<td>SC</td>
<td>2.15 (0.11)</td>
<td>0.92 (0.14)</td>
</tr>
<tr>
<td>Short</td>
<td>CS</td>
<td>3.89 (0.11)</td>
<td>2.45 (0.14)</td>
</tr>
<tr>
<td></td>
<td>SC</td>
<td>3.45 (0.11)</td>
<td>1.85 (0.14)</td>
</tr>
</tbody>
</table>

Statistical results for d-prime for each language. Subjects were treated as a random variable.

<table>
<thead>
<tr>
<th>Language</th>
<th>Effects</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>Length</td>
<td>F(1, 47) = 40.02</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Order of Presentation</td>
<td>F(1, 47) = 5.01</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Length*Order</td>
<td>F &lt; 1</td>
<td></td>
</tr>
<tr>
<td>English</td>
<td>Length</td>
<td>F(1, 27) = 20.27</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td></td>
<td>Order of Presentation</td>
<td>F(1, 27) = 8.80</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>Length*Order</td>
<td>F &lt; 1</td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td>Length</td>
<td>F(1, 14) = 10.95</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>Order of Presentation</td>
<td>F &lt; 1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Length*Order</td>
<td>F &lt; 1</td>
<td></td>
</tr>
</tbody>
</table>
References


FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS


FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS


FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS


FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS


FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS


FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS


FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS


FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS


FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

Table 1. The obstruents and nasals of the three languages used in this study. Parentheses around the Catalan fricatives indicate that they are allophones of stops that are realized in intervocalic position. For Russian, only the plain (not palatalized) consonants are shown, since only these are relevant to this study. Also, the status of /v/ as a fricative or approximant continues to be debated, but it is listed here to reflect the possibility that it is phonologically and phonetically an obstruent in the types of clusters used in this study.

<table>
<thead>
<tr>
<th></th>
<th>Catalan</th>
<th>English</th>
<th>Russian</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stops</td>
<td>p t k b d g</td>
<td>p t k b d g</td>
<td>p t k b d g</td>
</tr>
<tr>
<td>Fricatives</td>
<td>(β) f (ð) s z ʃ ʒ (ɣ)</td>
<td>f v s z ʃ ʒ</td>
<td>f v s z ʃ ʒ x</td>
</tr>
<tr>
<td>Nasals</td>
<td>m n ɲ ŋ ŋ</td>
<td>m n ɲ ŋ</td>
<td>m n ɲ ŋ</td>
</tr>
</tbody>
</table>
Table 2. Word-initial clusters used in the experiment that are attested or unattested in Russian.

Note that some of these clusters are attested primarily through morphological concatenation of the preposition /v-/ ‘in’, /s-/ ‘with’, or /k-/ ‘to, toward’. Parentheses around a cluster means that it may be attested in a loanword (e.g. ‘pfenig’)

<table>
<thead>
<tr>
<th>Attested</th>
<th>Unattested</th>
</tr>
</thead>
<tbody>
<tr>
<td>FF</td>
<td>fs-</td>
</tr>
<tr>
<td>FN</td>
<td>zm- zn-</td>
</tr>
<tr>
<td>FS</td>
<td>fp- fk- ft-</td>
</tr>
<tr>
<td>SF</td>
<td>kf- ks- gz-</td>
</tr>
<tr>
<td>SN</td>
<td>dn- gn- kn- km- pn- tm-</td>
</tr>
</tbody>
</table>
Table 3. Duration, standard deviation, and range of schwa in milliseconds

<table>
<thead>
<tr>
<th></th>
<th>Mean duration</th>
<th>Standard deviation</th>
<th>Min duration</th>
<th>Max duration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long stimuli</td>
<td>64.59</td>
<td>11.57</td>
<td>46</td>
<td>90</td>
</tr>
<tr>
<td>Short stimuli</td>
<td>70.19</td>
<td>12.44</td>
<td>44</td>
<td>90</td>
</tr>
</tbody>
</table>
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

Table 4. Accuracy results for each language group, including both same trials (CC, SS) and different trials (CS, SC). CC = Cluster-Cluster, SS = Schwa-Schwa, CS = Cluster-Schwa, SC = Schwa-Cluster. Standard error is in parentheses.

<table>
<thead>
<tr>
<th>Language</th>
<th>Long CC</th>
<th>(Standard Error)</th>
<th>Long SS</th>
<th>(Standard Error)</th>
<th>Long CS</th>
<th>(Standard Error)</th>
<th>Long SC</th>
<th>(Standard Error)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalan</td>
<td>0.91</td>
<td>(0.009)</td>
<td>0.94</td>
<td>(0.009)</td>
<td>0.99</td>
<td>(0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>English</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Russian</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short CC</td>
<td>0.94</td>
<td>(0.008)</td>
<td>0.93</td>
<td>(0.013)</td>
<td>0.97</td>
<td>(0.008)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short SS</td>
<td>0.94</td>
<td>(0.008)</td>
<td>0.96</td>
<td>(0.008)</td>
<td>0.97</td>
<td>(0.009)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short CS</td>
<td>0.82</td>
<td>(0.016)</td>
<td>0.61</td>
<td>(0.025)</td>
<td>0.82</td>
<td>(0.030)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short SC</td>
<td>0.75</td>
<td>(0.019)</td>
<td>0.48</td>
<td>(0.027)</td>
<td>0.88</td>
<td>(0.029)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure captions

Figure 1. Accuracy (hits) proportions for different trials, divided by length of stimuli (long or short) and order of presentation (cluster-schwa or schwa-cluster). For this and following figures, error bars show standard error of the mean. (a) Catalan (b) English (c) Russian.

Figure 2. Accuracy (hits) proportions for long and short different trials (collapsed for order of presentation). (a) Catalan (b) English (c) Russian.
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

Catalan

Accuracy

Consonant sequence type

cluster-schwa (CS) schwa-cluster (SC)

English

Accuracy

Consonant sequence type

cluster-schwa (CS) schwa-cluster (SC)
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

Russian

![Graph showing accuracy levels for different consonant sequence types in Russian. The graph compares long and short sequences, with error bars indicating variability.](image-url)
FACTORS IN DISCRIMINATION OF PHONOTACTIC CONTRASTS

Russian

Consonant sequence type

Accuracy

FS  FF  FN  SF  SN

Long
Short