Journal of Phonetics 40 (2012) 234–248

Contents lists available at SciVerse ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Sources of illusion in consonant cluster perception

Lisa Davidson [a],*, Jason A. Shaw [b]

[a] New York University, Department of Linguistics, 10 Washington Place, New York, NY 10012, USA
[b] MARCS Auditory Laboratories/School of Humanities and Languages, University of Western Sydney, Locked Bag 1797, Penrith South DC, NSW 2751, Australia

ARTICLE INFO

Article history:
Received 4 November 2010
Received in revised form
11 November 2011
Accepted 18 November 2011
Available online 15 December 2011

ABSTRACT

Previous studies have shown that listeners have difficulty discriminating between non-native CC sequences and licit alternatives (e.g. Japanese [ebzo]-[ebuzo], English [bnif]-[bənif]) (Berent et al., 2007; Dupoux et al., 1999). Some have argued that the difficulty in distinguishing these illicit–licit pairs is due to a "perceptual illusion" caused by the phonological system, which prevents listeners from accurately perceiving a phonotactically unattested consonant cluster. In this study, we explore this and other sources of perceptual illusion by presenting English listeners with non-native word-initial clusters paired with various modifications, including epenthesis, deletion, C_1 change, and prothesis, in both AX and ABX discrimination tasks (e.g. [zmatu]-[zəmatu], [matu], [smatu], or [əzmatu]). For English listeners, fricative–initial sequences are most often confused with prothesis, stop–nasal sequences with deletion or change of the first consonant, and stop–stop sequences with vowel insertion. The pattern of results across tasks indicates that in addition to interference from the phonological system, sources of perceptual illusion include language-specific phonetic knowledge, the acoustic similarity of the stimulus items, the task itself, and the number of modifications to illicit sequences used in the experiment.

1. Introduction

Perceptual illusions in speech are often said to occur when listeners encounter a sequence that is phonotactically impossible in their own language. For example, it has been claimed that Japanese listeners tend to perceive illusory vowels between the consonants in VC_1C_2V stimuli when the C_1C_2 string is not possible in Japanese (Dehaene-Lambertz, Dupoux, & Gout, 2000; Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Dupoux, Pallier, Kakehi, & Mehler, 2001; Dupoux, Parlato, Frota, Hirose, & Peperkamp, 2011). Further investigations of perceptual illusions have argued that they are more prevalent as the phonotactic structure becomes more phonologically complex or phonetically difficult along some dimension that is relevant to the particular language being studied. For example, Kabak and Idsardi (2007) claimed that for Korean listeners, it is not just the linear string of consonants that induces an illusion, but rather whether or not the consonant in the C_1 position in VC_1C_2V can appear as a syllable coda in the language. In that study, illicit codas were more likely to induce perceptual illusions. For English and Korean speakers listening to Russian-possible onset sequences, Berent and colleagues (Berent, Lennertz, Jun, Moreno, & Smolensky, 2008; Berent, Lennertz, Smolensky, & Vaknin-Nusbaum, 2009; Berent, Steriade, Lennertz, & Vaknin, 2007) have claimed that illusions are more likely for onset consonant clusters that have a plateau or falling sonority contour (e.g. [bdif]) as compared to those with a larger sonority increase from C_1 to C_2 (e.g. [bnif]) (where /n/ has a higher sonority value than /d/, e.g. Selkirk, 1984). Using fine-grained place and manner distinctions among onset consonant clusters, Davidson (2011b) showed that different combinations of manner of articulation allow for a more refined account of poor discrimination between non-native CC and native CəC sequences. To explain improved accuracy in discriminating clusters with smaller sonority distance like [fsa]-[fəsá] than clusters with a larger distance like [fna]-[fəná], Davidson speculated that phonetic properties of the clusters are most likely to account for the observed perceptual patterns (see also Dupoux et al., 2011).

One question left largely unaddressed by these past studies is the following: which perceptual illusion is most likely for a given consonant cluster in a given language? Past studies have typically involved tasks comparing phonotactically illicit sequences to one (of many) possible modification(s) of that sequence. For example, in one experiment in Dupoux et al. (1999), Japanese listeners were presented with VCV_xCV words, where the duration of V_x was on a 6-step continuum from 0 to greater than 58 ms. For each step on the continuum, listeners were asked to report whether or not they heard a vowel. Even for the tokens at 0 ms, Japanese listeners

* Corresponding author: Tel.: +1 1 212 992 8761.
E-mail address: lisa.davidson@nyu.edu (L. Davidson).

0095-4470/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.wocn.2011.11.005

reported the presence of a vowel more than 60% of the time, whereas French listeners did not report a vowel for these stimuli. Accuracy on this and other tasks comparing illicit consonant clusters with one type of alternative has often been attributed to native language phonological restrictions (Berent et al., 2007, 2008, 2009; Matthews & Brown, 2004). However, the phonotactics of a language shape listener experience both at the level of discrete segments and at the level of continuous phonetic variables. Each of these levels may constitute a source of perceptual confusion, but past studies have not been designed to tease them apart. In the present study, we investigate how English listeners may be hearing non-native phonotactics by pairing unattested consonant clusters with a variety of sequences that are possible confusions, and we use this design to investigate different sources of perceptual illusion.

1.1. Consonant cluster modifications in perception and production

A survey of the literature indicates that the type of perceptual illusions elicited in a speech experiment depends on the task and the stimuli included in that experiment. In studies relying on the illusion of a consonant change, researchers examined English listeners' perception of unattested sequences such as [dl], [tl], and [sr] by developing stimuli in which [d], [t], and [s] were followed by an [r]-[l] continuum (Massaro & Cohen, 1983; Pitt, 1998, see also Moreton, 2002). In experiment 1 of his study, Pitt (1998) found that compared to their perceptions of liquids after [ɡ] or [b], where both [r] and [l] are legal, American English listeners were biased to report hearing [l] following [s] and [r] following [t]/[d]. In experiment 3, Pitt examined a different perceptual illusion – presence of a vowel – by lengthening the duration of the liquid. The results showed that listeners were more likely to label words with longer liquids in [tl] and [sr] as being disyllabic than words containing [tr] and [sl]. The comparison of experiments 1 and 3 in Pitt (1998) shows that the particular perceptual illusion that is induced is task-dependent; that is, on the same sequences, listeners perceive a change in a consonant's identity when phonetic properties distinguishing two consonants are manipulated, or the presence versus absence of a vowel if that is the demand of the task. In a task designed to involve lexical influence, Hallé et al. (2008) used a visual masked priming paradigm with Spanish speakers to demonstrate that illusory vowels can influence reaction times in a lexical decision task. When target words like *especial* are primed with the written pseudoword *special*, reaction times are as fast as in identity priming trials. Hallé et al. conclude that perceptual illusions – *special* processed as *especial* – affect real lexical items of a language and that they can be induced through orthography alone.

Results from production show that while #CəC is a frequent repair for many #CC words (Davidson, 2010), it is not the only possible way that English speakers produce these consonant clusters. In addition to inserting a vowel between the two consonants, listeners also may change the first consonant, insert a vowel before the CC sequence (prothesis), or delete the first consonant (see also Berent et al., 2007, 2009; Brown & Hildum, 1956; Davidson, 2007 for similar types of repairs English speakers use in written tasks). It should also be noted that when speakers change the first consonant, they do not always do so in a way that makes the sequence phonotactically legal. For example, participants asked to produce a stop–nasal sequence as in [pnaka] may produce [tnaka] or [bnaka]. One possibility is that they have misperceived the cluster and are producing what they believe they heard.

This brings us back to our central question: what are the sources of illusion in consonant cluster perception? One attempt at addressing this was an open-choice experiment in Hallé et al. (1998), in which French listeners were presented with words beginning with [dl] and [tl] (e.g. [tlabdo]) and were asked to write down what they heard. Over 85% of the responses included a velar stop – [kl] and [ɡl] – instead of a coronal one. While this is suggestive of how French listeners perceive [dl] and [tl] sequences, the results may also reflect an orthographic bias to write a sequence of letters that occurs in the language. To avoid this potential bias, Hallé et al. further examined the perception of [tl] and [dl] sequences by asking listeners to respond only whether the word began with a labial, coronal, or velar consonant (presented to them as letters). Results showed that the participants chose a velar consonant 63% of the time for [t] and 48% of the time for [d], which is a less robust proportion than the first experiment, but still revealing (see also Hallé & Best, 2007 for discrimination studies demonstrating a coronal-to-velar shift in /Cl/ sequences for English and French listeners). While this task avoids the bias toward writing unattested orthographic sequences, the task also restricts the space of perceptual modifications to consonant change.

1.2. Current study

This study has two main aims. The first is to investigate the most likely perceptual confusions of unattested $\#C_1C_2$ sequences for English listeners by pairing them with several alternatives, including vowel insertion ($\#C_1əC_2$), C_1 change ($\#C_3C_2$), prothesis ($\#əC_1C_2$), and C_1 deletion ($\#C_2$). Understanding which modifications are confusable for a given cluster may be highly informative in revealing contributions of phonological factors in non-native perception. We focus on word-initial obstruent–obstruent and obstruent–nasal sequences that are possible in some languages but not in English, such as those in [zmatu] or [ktase]. For a non-word like [zmatu], if English listeners do not perceive it accurately, are they most likely to hear [zəmatu], [ezmatu], or perhaps [smatu]?

What listeners are actually hearing informs the type of process that is engaged in perception. If native language phonotactics influence perception after categorization, then we expect percepts to be warped in the direction of phonological expectations. This will tend to result in perceptual illusions that improve on markedness, e.g., a more marked cluster will be perceived as a less marked or even as a permissible alternative. On the other hand, if language-specific phonetic knowledge is dictating patterns of perception, then we expect listeners to rely on phonetic cues that are important for contrasts in their own language, but to downplay cues that are not important in their language (but may be relied on by speakers of other languages). These two sources of perceptual confusion, language-specific phonetic knowledge and top-down phonological knowledge, can be distinguished by analyzing the pairings of unattested consonant clusters with multiple legal and illegal alternatives.

The second main aim of the paper is to make a methodological contribution by comparing two different perceptual paradigms, the AX task and the ABX task, on word-length stimuli. Although there is a canon of foundational research providing extensive discussion of the cognitive processes and representations involved in AX and ABX discrimination paradigms (e.g., Carney, Widin, & Viemeister, 1977; Durlach & Braida, 1969; Pisoni, 1973; Polka, 1991; Strange & Shafer, 2008; Werker & Logan, 1985), much of this discussion is based on results involving short speech stimuli constituting sub-minimal words (on word minimality see Demuth, 1996; McCarthy & Prince, 1996; Peperkamp, 1999). In recent years, the AX and ABX perceptual paradigms have been extended to address increasingly sophisticated questions about the relation between phonology and perception (e.g., Berent et al., 2007; Best, McRoberts, & Goodell, 2001; Davidson, 2011b; Dupoux et al., 1999; Gallagher, 2010; Kabak & Idsardi, 2007; Matthews & Brown, 2004).

While many of these questions have been investigated using longer stimuli, the assumptions about the cognitive processes engaged by the tasks have not been reevaluated. Nevertheless, such reevaluation may be necessary. Davidson (2011b) demonstrated that AX discrimination performance is degraded when target contrasts are included in longer, more word-like strings (e.g. [zɡamo]) as compared to shorter monosyllabic strings (e.g. [zɡa]). In that study, the phonotactically unattested stimuli consisted of both long and short non-words beginning with fricative–fricative, fricative–nasal, fricative–stop, stop–fricative, and stop–nasal consonant clusters (e.g. [ɡzá]/[ɡəzá] and [ɡzábo]/[ɡəzábo], [pná]/[pəná] or [pnáka]/[pənáka]). Results showed that American English listeners were significantly below chance on all pairings for the longer words, regardless of the combination of consonants. For the shorter words, accuracy on fricative–fricative, stop–fricative, and stop–nasal clusters was significantly above chance, but participants did not reach greater than 70% accuracy for any of the sequences.

The results of Davidson (2011b) call into question whether conclusions about the AX task drawn on the basis of short stimuli can be generalized to word-like stimuli. For short stimuli, there is widespread agreement that AX discrimination (1) does not impose a heavy memory load and (2) encourages use of fine acoustic details to discriminate two tokens (e.g., Pisoni, 1973; Polka, 1991; Strange & Shafer, 2008). On the basis of the Davidson (2011b) results, we contest the commonly held assumption about the memory load imposed by AX. Scanning pairs of quasi-continuous representations of fine acoustic detail for low-level differences appears to be a highly difficult task for word-length stimuli that are mostly identical. Furthermore, as Gerrits and Schouten (2004: 364) argue, listeners may be very conservative in AX tasks and "may decide to respond 'different' only if they are very sure of their decision."

In the current study we will compare results on an AX task with results on an ABX categorial discrimination task. In the ABX task, listeners are presented with three physically different tokens: the cluster, some modification of the cluster (these are counterbalanced for order), and the last token, which is a different utterance either containing the same cluster or the same modification. Listeners have to determine whether the last stimulus was the same as the first or the second. To adequately perform this task, listeners must label the A stimulus and the B stimulus so that they can determine whether X is an instance of A or B (Gerrits & Schouten, 2004). Since the task requires holding two labels in memory, it is widely assumed that ABX has a higher cognitive load than AX, and that, consequently, performance on ABX will be degraded relative to AX (Carney et al., 1977; Pisoni, 1973; Strange & Shafer, 2008). Again, however, these conclusions are usually based on comparisons of very simple strings, e.g. retroflex [ʈa] versus dental [ta] as in Werker and Logan (1985). Since the ABX task does not require that a representation of fine acoustic information from both the A and B stimuli persist throughout the trial, for long stimuli it may be easier to do the ABX task than the AX task. This prediction runs counter to the understanding of these tasks established in past work. However, the results of Davidson (2011b) already suggest that this understanding as applied to word-length stimuli may be incomplete. The current study offers a comparison of these two discrimination tasks using word-like stimuli.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Listeners included 38 participants recruited primarily through New York University classes and a posting on the website Craigslist (http://newyork.craigslist.com, a website for free online classified advertisements) in New York. They ranged in age from 19 to 42. No participants reported any history of speech or hearing disorders, or any knowledge of Slavic or Semitic languages, since these languages allow the clusters being tested in this study. The listeners were paid $10 for their participation. The data from one participant were discarded because he failed to respond to more than half of the trials.

2.1.2. Materials

The target materials for the AX discrimination trials consisted of non-words containing initial obstruent–obstruent and obstruent–nasal sequences and matching non-words that represented the kinds of modifications that English listeners make in production (Davidson, 2006a, 2010).[1] In Davidson (2010), English speakers were asked to produce a wide range of non-native consonant clusters. Over 90% of the productions errors in that study fell into four categories: insertion [fm-]/[fəm-], C_1 deletion [zb-]/[b-], prothesis [dm-]/[ədm-], and C_1 change [ɡd-]/[bd-]. Following Davidson (2010), which found that the combination of manner of articulation in the clusters was most informative for explaining patterns of production, 16 onset clusters constituting four different manner combinations were selected for the current study: fricative–nasal (FN: [fm], [sm], [zm], [vm]), fricative–stop (FS: [fp], [sp], [zb], [vb]), stop–stop (SS: [dɡ], [ɡd], [tk], [kt]), and stop–nasal (SN: [bm], [dm], [pm], [tm]). Each of these combinations was used to make cluster stimuli of the form CCáCV. The diacritic over the [a] denotes stress. In the discrimination trials, each word containing a consonant cluster was paired with matching words for each of the following modifications: insertion (e.g., [tmáfa]/[təmáfa]), C_1 deletion (e.g., [tmáfa]/[máfa]), prothesis (e.g., [tmáfa]/[ətmáfa]), and C_1 change.

For the C_1 change modification, there was more than one trial, since there are multiple ways in which a consonant can be changed (e.g. [tmafa]/[dmafa], [tmafa]/[pmafa]). The particular modification types that were chosen for comparison, including the specific C_1 change modifications that are paired with the cluster, are based on the errors that speakers made in Davidson (2010). There were a total of 74 cluster-modification trials. A full list of all 'different' trials is given in Appendix A. The 'same' trials consisted of each word paired with itself, both for cluster and modification stimuli.

The stimuli were recorded by a bilingual English/Russian speaker. In order to ensure that listeners were responding 'different' on the basis of the onset of the word only, the –áCV portion from one stimulus in the paradigm was spliced onto the rest of the stimuli. For example, all of the words that [tmafa] was paired with, including [mafa], [təmafa], [ətmafa], [dmafa], [pmafa], [bmafa], all shared the physically same [-afa]. However, the choice of –áCV was chosen randomly for each paradigm so that the same initial sequence was not always the source of the remaining –áCV. The –áCV portion was always cut from zero crossings to avoid acoustic artifacts. Splicing was carried out straightforwardly by looking for obvious landmarks such as the end of a stop burst in C_1 position or the beginning of middle formant attenuation for the nasals in C_2 position. Acoustic characteristics of the cluster and the insertion, prothetic, C_1 deletion, and C_1 change modifications are shown in Appendix B.

The recordings were done in a sound-treated room using a Marantz PMD-670 solid state recorder and a Shure Beta 58A microphone. The stimuli were recorded as wave files onto a compact flash card at 22.050 kHz.

[1] Readers interested in a comparison of the production and perception results for the non-native #CC clusters are referred to Shaw and Davidson (2011).

In addition to the test stimuli, five tokens with stop–liquid or /s/-stop onset clusters were recorded by an American English speaker for the practice phase. These were also paired with C_1 change, C_1 deletion, prothesis, or epenthesis modifications. None of these were the same words used in the test phase.

2.1.3. Procedure

The experiment was implemented in ePrime (version 1.1) (Schneider, Eschman, & Zuccolotto, 2002). The AX discrimination task contained 74 different trials and 57 same trials, for a total of 121 trials. The number of same and different trials was different since words were paired with more than one modification in the different trials. The mismatch in the number of same and different trials is warranted; since previous research suggests that listeners will respond ‘same’ to a substantial portion of the different trials in studies of phonotactic illegality, the mismatch may help prevent the listener from developing a strategy of deciding that ‘different’ trials are rare compared to ‘same’ ones. The different trials contained both orders of presentation, so each participant heard half of the stimuli with the cluster word first (e.g. [tmáfa]/[máfa]) and half with the modification word first (e.g. [ədɡáse]/[dɡáse]). The order for any particular trial was chosen at random by ePrime, but each participant heard half of the different trials in cluster-modification order and the other half in modification-cluster order. In addition, the interstimulus interval (ISI) was a variable in this experiment. Each participant heard all of the stimuli with ISIs of both 250 ms and 1500 ms, for a total of 300 trials for each participant. The stimuli were blocked for ISI, and the ISI conditions were counterbalanced so half of the participants heard the short ISI first and half heard the long ISI first.

The participants were given the following instructions: “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” if they thought the sound files were the same, and “D” if they thought the sound files were different. Participants were encouraged to answer quickly, and were told to choose either “S” 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 duration of each sound file (i.e. the length of each word), there was a 250 ms or 1500 ms pause, and then another fixation cross along with the second sound file. Participants could make a response anytime after the start of the second sound file. As soon as the participant made a response, there was a 2500 ms pause and the next trial started. The whole study lasted approximately 15 min.

Participants were seated in individual small, quiet rooms containing PCs and Sennheiser headphones. Participants were first given 8 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 half of the trials.

2.2. Results

2.2.1. Accuracy

An analysis of variance was conducted to examine participants’ performance on the AX discrimination trials. The within-subjects independent variables were interstimulus interval (ISI: 250 ms, 1500 ms), sequence type (FN, FS, SN, SS), and modification (insertion, prothesis, C_1 change, C_1 deletion). Subjects were included as a random factor. The dependent variable was d' (d prime), a sensitivity measure that computes how easily a listener can detect whether or not the signal is present (Green & Swets, 1966; Macmillan & Creelman, 2005). In the context of this study, the signal is present on the different trials. When d' is 0, listeners are effectively answering at chance. The d' score can also be negative, which would demonstrate that listeners are not discriminating between the stimuli. The maximum d' value is 4.65, which indicates that listeners are scoring at ceiling on both different and same trials. Results for d' are shown in Fig. 1 and results for accuracy (proportion correct on different trials) is shown in Table 1.

Results for the ANOVA showed a main effect of sequence [$F(3, 108)=30.05$, $p<0.001$] and modification [$F(3, 108)=28.05$,

Table 1
Accuracy (proportion of hits) to different trials for each sequence type and modification.

	FN	FS	SN	SS
C_1 change	0.85	0.87	0.64	0.67
C_1 deletion	0.95	0.93	0.64	0.72
Prothesis	0.66	0.41	0.72	0.57
Insertion	0.80	0.88	0.86	0.51

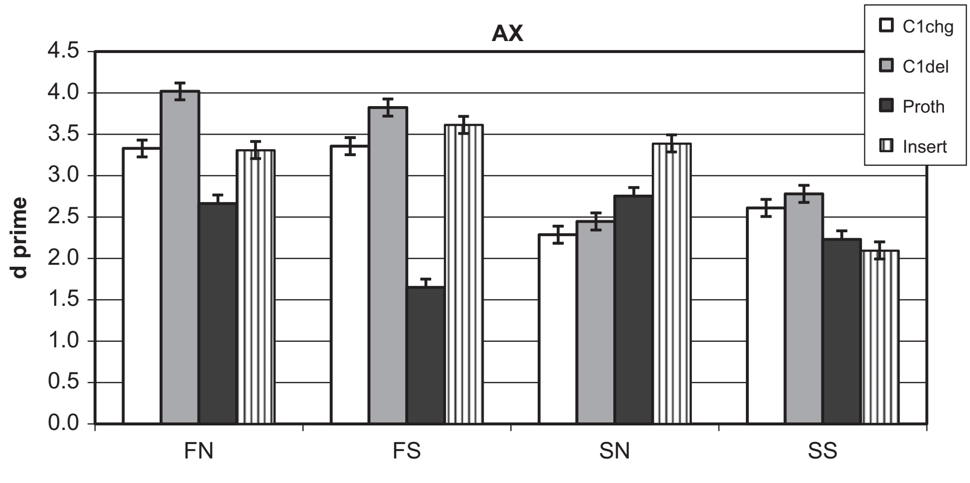

Fig. 1. d' scores for each type of modification for the AX discrimination task. In this and following figures, FN=fricative-nasal; FS=fricative-stop; SN=stop–nasal, SS=stop–stop. C_1 chg=C_1 change, C_1 del=C_1 deletion, Proth=prothesis, Insert=insertion. Error bars indicate standard error.

$p<0.001$], but no effect of ISI [$F(1, 36)<1$]. There was also an interaction between sequence and modification [$F(9, 324)=26.05$, $p<0.001$], but no two- or three-way interactions with ISI were significant. A Tukey HSD post-hoc test shows that the main effect of sequence was due to significant differences among all of the cluster types (> indicates significantly higher d' scores, $p<0.05$): FN (mean $d'=3.33$) > FS (3.11) > SN (2.72) > SS (2.43). The main effect of modification also showed significant differences among the four types: C_1 deletion (3.27) > insertion (3.10) > C_1 change (2.89) > prothesis (2.32).

To investigate the interaction between sequence and modification (see Fig. 1), separate ANOVAs were conducted for each sequence with modification as the independent variable. For FN sequences, there was a significant effect of modification [$F(3, 108)=23.42$, $p<0.001$]. A Tukey HSD post-hoc test shows the following pattern of d' results: C_1 deletion > C_1 change = insertion > prothesis. For FS sequences, the effect of modification was also significant [$F(3, 108)=68.97$, $p<0.001$], and the d' results were C_1 deletion = insertion > C_1 change > prothesis. For SN, there was a significant effect of modification [$F(3, 108)=11.31$, $p<0.001$], with the post-hoc test showing the pattern insertion > prothesis > C_1 deletion = C_1 change. Finally, there was a significant effect of modification for SS [$F(3, 108)=8.17$, $p<0.001$] with the pattern C_1 deletion = C_1 change > prothesis = insertion.

2.2.2. Reaction time

An analysis of variance was also carried out on the reaction time in milliseconds (which was log transformed), using the same independent variables as in the accuracy analysis. Reaction times were recorded beginning at the end of the *X* stimulus, and were analyzed only for different trials that participants had responded to correctly. Results are shown in Fig. 2.

Results for the ANOVA showed a significant effect of modification [$F(3, 99)=6.14$, $p<0.001$], but no main effect of sequence [$F(3, 96)<1$]. There was no interaction between sequence and modification [$F(9, 317)<1$]. Post-hoc Tukey HSD tests (using adjustments for multiple comparisons) showed that the effect of modification was due to significantly higher reaction time for C_1 deletion as compared to insertion ($p=0.007$) and C_1 change ($p=0.05$). No other comparisons were significant.

2.2.3. Place of articulation

In order to easily compare the results of these studies both to the production results in Davidson (2010) and the perception results of Davidson (2011b), these experiments were structured primarily to test manner combinations. However, it is of interest to compare performance on different repairs as a function of place of articulation combination as well. The means for d' measures for modification and place of articulation combinations are in Table 2. An ANOVA with place combination (Coronal–Labial, Labial–Labial, Coronal–Velar, Velar-Coronal) and modification (C_1 change, C_1 deletion, insertion, prothesis) as independent variables, and d' as the dependent variable was carried out. Subjects were treated as a random variable. Results showed no significant result of place combination [$F(5, 157)=6.58$, $p=0.07$], but there was a significant effect of modification [$F(3, 90)=21.04$, $p<0.001$] and a significant interaction between modification and place combination [$F(15, 420)=4.38$, $p<0.001$]. Tukey's HSD post-hoc tests within place combination for each modification type were also carried out (' > ' indicates significantly higher d' score, $p<0.05$; Coronal–Labial: C_1 deletion > insertion = C_1 change > prothesis; Labial–Labial: insertion = C_1 deletion > C_1 change > prothesis; Coronal–Velar: C_1 change = prothesis > C_1 deletion = insertion; Velar–Coronal: C_1 deletion > C_1 change = insertion = prothesis.)

2.3. Discussion

The results of the AX discrimination task indicate that the most likely confusion for each sequence depends on the combination of consonants that form the unattested cluster. This result can be contrasted with past experiments where findings suggested that CVC is very confusable with CC for a variety of sequence types (Berent et al., 2007, 2009; Davidson, 2011b; Dupoux et al., 1999, 2011). In this experiment, stop–stop clusters were highly confusable with the insertion modification, though these clusters were also confusable with the prothetic modification. Other clusters showed different patterns of perceptual confusion. For both fricative–nasal and fricative–stop sequences, listeners were most likely to confuse the prothetic modification with the cluster. Stop–nasal sequences were confused with C_1 deletion and C_1 change most (and equally) often. These results suggest that there is not a singular top-down phonological repair that affects the perception of non-native sequences. As will be

Table 2
Means for d' measures for modification and place of articulation combinations (coronal–labial, labial–labial, coronal–velar, velar–coronal). Standard error is in parentheses.

	CorLab	LabLab	CorVel	VelCor
C_1 chg	3.041 (0.104)	2.914 (0.108)	3.116 (0.153)	2.238 (0.189)
C_1 del	3.564 (0.128)	3.316 (0.131)	2.298 (0.177)	3.599 (0.222)
Insertion	3.192 (0.130)	3.693 (0.107)	2.258 (0.178)	1.820 (0.247)
Prothesis	2.446 (0.134)	2.297 (0.145)	2.882 (0.234)	1.696 (0.200)

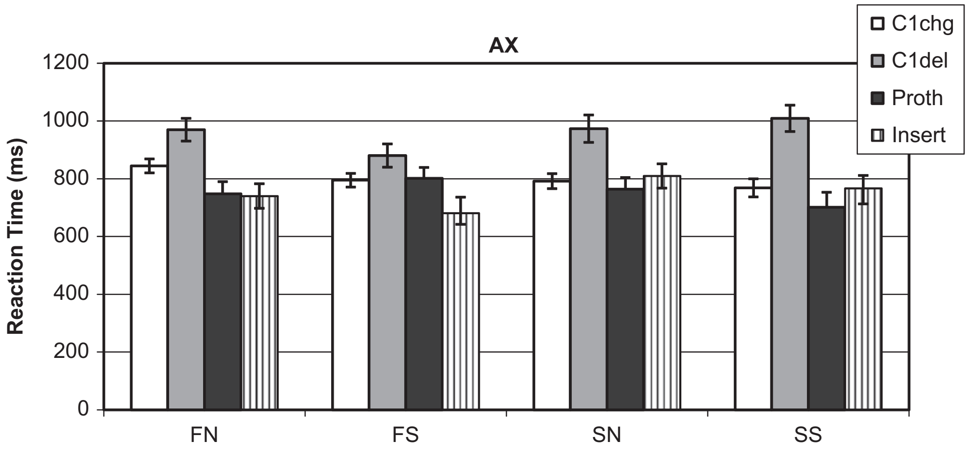

Fig. 2. Reaction time in milliseconds for each type of modification for the AX discrimination task. Error bars indicate standard error.

discussed further in Section 4, the most common confusions shown in this study can be straightforwardly explained by appealing to their phonetic characteristics.

Reaction time results show that there were no differences for sequence and very few differences for modification; the only significant result was that listeners were significantly slower on C_1 deletion comparisons than any other. A possible reason for this result is that since the –aCV part of the stimuli was exactly the same for each word in the trial, C_1 deletion cases (which were the shortest stimuli) had the highest proportion of material that was acoustically identical between both words in the trial. Listeners may have been slower to respond in these cases because there were fewer differences among the stimuli (though this apparently did not affect accuracy except for SN stimuli). This interpretation is consistent with the view that, in an AX task, listeners are scanning quasi-continuous representations for differences. This task takes longer when the proportion of different material in the two stimuli is decreased.

A closer look at the results for place of articulation suggests that they are mostly attributable to the manner combinations of the sequences being discriminated. In the AX task, the place combinations coronal–labial and labial–labial are comprised of fricative–nasal, fricative–stop, and stop–nasal sequences. The manner combination results showed that confusion with the prothetic modification was most common for fricative–initial sequences (which make up 2/3 of these place combinations), and this is also true for the coronal–labial and labial–labial sequence. After prothesis, the next lowest d' score is for C_1 change for both of these sequences, which is attributable to stop–nasal sequences (1/3 of these place combinations).

The place combinations coronal–velar and velar–coronal pertain to the stop–stop sequences, and for both of these, as expected, sensitivity to the insertion repair is low. However, there are also two asymmetries in the coronal–velar and velar–coronal sequences. First, d' of the C_1 deletion modification is significantly higher than all other modifications in velar–coronal combinations, but not in coronal–velar. Further analysis of the data reveals that this is almost entirely due to poor performance on the /dɡ/ cluster for C_1 deletion (27% accuracy, compared to 89% for /tk/, 92% for /ɡd/, and 84% for /kt/). A comparison of the burst characteristics in Appendix B shows that the burst of the /d/ in /dɡ/ is only 13 ms, less than half the duration of the bursts of all the other C_1s in the relevant comparisons. As we will further discuss for SN sequences in Section 4, clusters with weak stop bursts following C_1 are easily confused with the deletion modification. Second, planned comparisons show a difference in sensitivity to the prothesis modification between velar–coronal ($d'=1.696$) and coronal–velar ($d'=2.882$) place combinations ($p<.05$). Further analysis of the data shows that prothesis confusions occur mostly for voiced /ɡd/ sequences, and less so for /kt/ (accuracy for /ɡd/: 20%, for /kt/: 66%). While both initial /ɡ/ and /d/ in the stimuli were produced with voicing during the stop closure, listeners were apparently more likely to confuse the voicing for /ɡ/ with prothesis than they were for the coronal. Notably, duration for voicing for /ɡ/ was 109 ms, versus 81 ms for /d/. From this data it is evident that the low sensitivity to the prothesis modification in stop–stop sequences is entirely attributable to velar–coronal sequences.

The d' and accuracy scores indicate that when listeners confuse the #CC sequence with the closest perceptual modification, they are not performing close to a ceiling level—for example, for the manner combination results, accuracy of hits on different trials is below 70% for each sequence type (e.g. 66% and 41% accuracy for the prothesis modification for FN and FS, respectively, 64% for both C_1 change and C_1 deletion for SN, and 51% and 57% for SS for insertion and prothesis, respectively). This suggests that even in the purportedly simplest paradigm for making acoustic comparisons, listeners still have difficulty distinguishing between #CC sequences and the most similar modification. As suggested by Gerrits and Schouten (2004), it is possible that listeners are instead using some criteria to deem them "close enough" and report that they are the same. If this is true, a different task – namely, one that requires some kind of categorization – may further direct listener attention to the relevant acoustic cues in order to appropriately label the comparison stimuli. Decisions based on labels instead of on more continuous representations of the acoustic signal may encourage listeners to be less conservative in their responses. This hypothesis is tested in Experiment 2, which uses an ABX categorial discrimination paradigm. If categorization of the stimuli requires listeners to make use of phonetic detail to label the A and B stimuli, then they are expected to choose the correct match for *X* more often in the ABX than in the AX task. However, if they cannot adequately use the phonetic differences between the A and B stimuli to accurately label them, then participants should be at chance for the modification that is most easily confused with the cluster.

3. Experiment 2

3.1. Method

3.1.1. Participants

There were 36 participants recruited primarily through New York University classes and a posting on the website Craigslist in New York. All of these participants met the criteria in Experiment 1, though none of these subjects also participated in the first experiment. They ranged in age from 18 to 52. The listeners were paid $10 for their participation.

3.1.2. Materials

The materials for the ABX categorial discrimination task were of the same type as those used in Experiment 1. The manner combination sequences are the same in this study as in the previous one, except that [Cn] sequences were added to the stop–nasal category to ensure that generalizations about stop–nasal clusters held across different places of articulation of C_2; whereas the other categories all varied in their place of articulation of C_2 in Experiment 1, SN clusters did not. Thus, the target clusters for the SN stimuli consist of [bm], [dm], [pm], [tm], [bn], [dn], [pn], [tn].

The stimuli for Experiment 2 were recorded by the same Russian/English bilingual speaker as in Experiment 1. For this study the speaker recorded several tokens of each word. For each ABX trial, two physically different stimuli of the same non-word were chosen. Thus, for a trial such as [dmafa] [tmafa] [dmafa], the first and third stimuli were different utterances, but the correct answer would be to respond that the third stimulus is the same as the first. Acoustic measurements for the elements of the stimuli preceding –VCV are given in Appendix B.

The recordings were done in a sound-treated room using a Marantz PMD-670 solid state recorder and a Shure Beta 58A microphone. The stimuli were recorded as wave files onto a compact flash card at 22.050 kHz. The stimuli for the five practice trials were similar to those described in Experiment 1.

3.1.3. Procedure

The ABX discrimination task contained 364 trials. Each cluster/modification pair is presented in four stimulus orders – ABA, ABB, BAB, and BAA – where the A stimuli are clusters and the B stimuli are modifications. Since ISI was not significant in the previous experiment, only one ISI of 500 ms was used. The experiment was implemented in ePrime.

The participants were given the following instructions: "In the following task, you will hear three words presented in a row. In some of the triplets, the last word will be the same as the first. In other triplets, the last word will be the same as the second. Your task is to decide whether the LAST word is the same as the first or second. If you think the last word is the same as the first, press the button labeled "1" on the response box. If you think the last word is the same as the middle word, press the button labeled "2" on the response box. Please respond as quickly and as accurately as possible AFTER you hear the last word. Even if you don't think you know the answer, please choose either "1" or "2"." Using the ePrime button box, participants pressed a button labeled "1" or "2".

As in Experiment 1, 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 duration of each sound file (i.e. the length of each word), there was a 500 ms pause, and then another fixation cross along with the second sound file, followed by another 500 ms pause and then a final crosshair and the last sound file. As soon as the participant made a response, there was a 2500 ms pause and the next trial started. The whole procedure lasted approximately 20 min.

Participants were seated in individual small, quiet rooms containing PCs and Sennheiser headphones. Participants were first given 8 practice trials to familiarize themselves with the task; there was no feedback in the practice.

3.2. Results

3.2.1. Accuracy

An analysis of variance was conducted to examine participants' performance on the ABX task. The within-subjects independent variables were sequence (FN, FS, SN, SS) and modification (insertion, prothesis, C_1 change, C_1 deletion). There were no significant differences in performance between the [Cn] and [Cm] stimuli for the FN category, so these were collapsed in the results. Subjects were included as a random factor. The dependent variable was d', which is described for ABX designs in Macmillan and Creelman (2005: 229–232). A graph for d' and a table of accuracy are shown in Fig. 3 and Table 3.

Results show that there was a significant effect of sequence [$F(3,\ 105)=9.06$, $p<0.001$], and modification [$F(3,\ 105)=5.02$, $p=0.003$], and a significant interaction [$F(9,\ 322)=9.35$, $p<0.001$]. A Tukey HSD post-hoc test showed that the significant effect of sequence was due to greater sensitivity to the fricative–initial sequences than to the stop–initial ones: FN (mean $d'=3.49$)=FS (3.41) > SS (3.14) > SN (2.78). The significant effect of modification was due to the following pattern: prothesis (3.46), deletion (3.35) > C_1 change (3.04)=insertion (3.04).

The interaction between sequence and modification was investigated by separate one-way ANOVAs for each sequence type with modification as the independent variable (see Fig. 3). For FN and FS sequences, there were no significant differences [FN: $F(3,\ 106)=2.19$, $p=0.09$; FS: $F(3,\ 117)=2.07$, $p=0.11$]. For SN, there was a significant effect of modification [$F(3,\ 103)=13.18$, $p<0.001$] due to the following pattern ($p<0.05$): prothesis > insertion = deletion > C_1 change. The significant effect of modification for SS [$F(3,\ 111)=16.67$, $p<0.001$] was due to the pattern: prothesis = deletion = C_1 change > insertion.

In order to investigate whether listeners showed increased sensitivity (higher d' scores) in Experiment 2 than in Experiment 1, a separate between-subjects ANOVA with experiment as an independent variable was run for the modification types that resulted in the lowest d' scores in Experiment 1: prothesis for FN and FS, C_1 change and C_1 deletion for SN, and prothesis and insertion for SS. For both FN and FS, listeners had significantly higher d' scores in the ABX experiment [FN: $F(1,\ 69)=7.36$, $p<0.01$; FS: $F(1,\ 70)=11.87$, $p<0.001$]. For SN, there were no significant differences for either C_1 change [$F(1,\ 70)<1$] or C_1 deletion [$F(1,70)=1.61$, $p<.21$]. For SS, listeners had significantly higher ABX d' scores for prothesis [$F(1,\ 70)=9.11$, $p<.01$], but there were no significant differences for insertion [$F(1,\ 70)<1$].

3.2.2. Reaction time

An ANOVA was also carried out with sequence and modification as independent variables, and reaction time in milliseconds (log transformed) as the dependent variable. A graph for reaction time is shown in Fig. 4.

Table 3
Accuracy (proportion of hits) for each sequence type and modification.

	FN	FS	SN	SS
C_1 change	0.79	0.85	0.74	0.83
C_1 deletion	0.88	0.84	0.78	0.87
Prothesis	0.85	0.87	0.90	0.88
Insertion	0.88	0.85	0.82	0.63

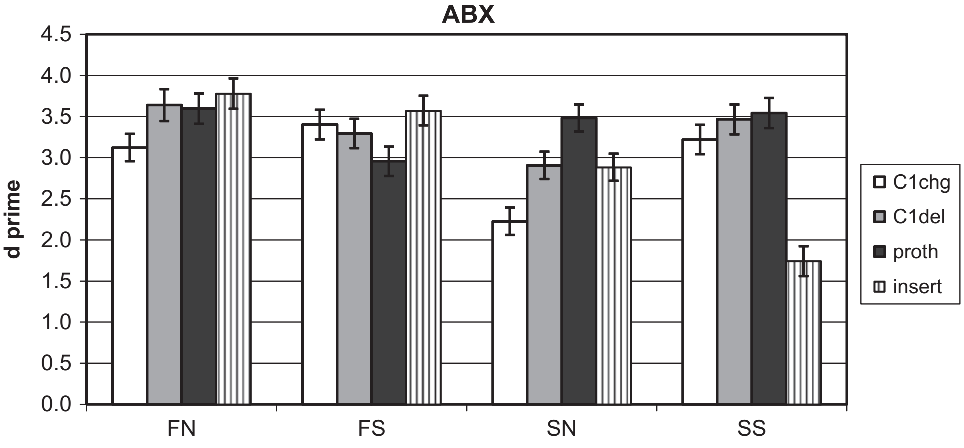

Fig. 3. d' scores for each type of modification for the ABX categorial discrimination task.

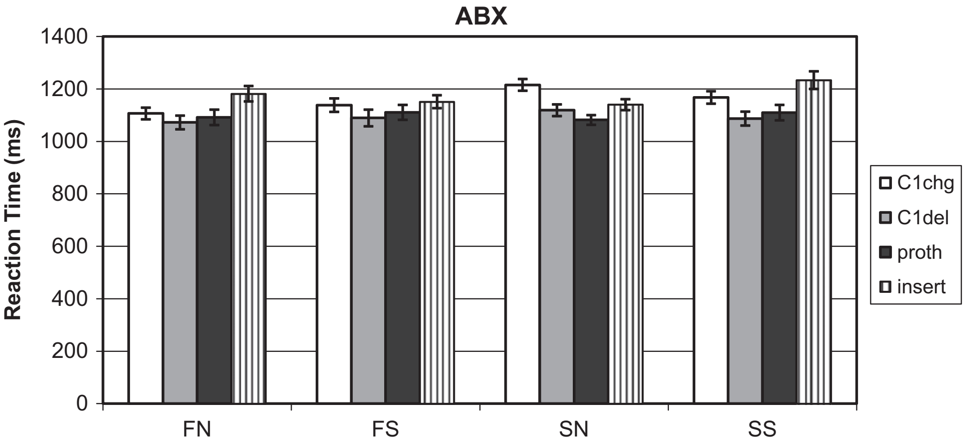

Fig. 4. Reaction time for each type of modification for the ABX categorial discrimination task.

Table 4
Means for d' measures for modification and place of articulation combinations (coronal–labial, labial–labial, coronal–velar, velar–coronal). Standard error is in parentheses.

	CorCor	LabCor	CorLab	LabLab	CorVel	VelCor
C_1 chg	2.006 (0.396)	2.375 (0.298)	2.982 (0.233)	2.743 (0.234)	3.224 (0.277)	3.092 (0.324)
C_1 del	3.384 (0.382)	3.656 (0.367)	3.505 (0.236)	3.066 (0.273)	3.551 (0.400)	3.860 (0.338)
Insertion	4.096 (0.194)	2.121 (0.532)	3.410 (0.236)	3.495 (0.247)	2.478 (0.379)	1.963 (0.419)
Prothesis	4.199 (0.260)	3.554 (0.430)	3.727 (0.218)	3.544 (0.226)	3.988 (0.289)	3.813 (0.336)

Results show that there was no effect of sequence [F(3, 93)=1.36, p=0.26], but modification [F(3, 96)=14.17, $p < 0.001$] and the interaction between sequence and modification were significant [F(9, 307)=3.11, $p < 0.002$]. To investigate the interaction, Tukey's HSD post-hoc tests were carried out separately for each sequence type. The results showed that for FN, listeners were significantly faster on C_1 deletion ($p < 0.05$) and prothesis (p=0.003) than insertion. For FS, C_1 deletion trials were significantly faster than insertion (p=0.03), but no other differences were significant. For SN sequences, C_1 change was significantly slower than all other modifications ($p < 0.001$, except insertion where p=0.01), and SS showed the most complex pattern ('>' indicates a faster response), where C_1 deletion= prothesis > C_1 change > insertion (all $p < 0.01$, except C_1 change > insertion, p=0.05).

3.2.3. Place of articulation

The means for d' measures for modification and place of articulation combinations are in Table 4. An ANOVA with place combination (Coronal–Coronal, Labial–Coronal, Coronal–Labial, Labial–Labial, Coronal–Velar, Velar–Coronal) and modification (C_1 change, C_1 deletion, insertion, prothesis) as independent variables, and d' as the dependent variable was carried out. Subjects were treated as a random variable. Results showed significant effects of place combination [F(5, 168)=5.58, $p < 0.001$], modification [F(3, 98)=24.55, $p < 0.001$] and an interaction between modification and place combination [F(15, 523)=6.33, $p < 0.001$]. Tukey's HSD post-hoc tests within place combination for each modification type were also carried out. ('>' indicates significantly higher d' score, $p < 0.05$; Coronal–Coronal: prothesis=insertion=C_1 deletion > C_1 change; Labial–Coronal: C_1 deletion=prosthesis > C_1 change=insertion; Coronal–Labial: prothesis > C_1 change (no other significant differences); Labial–Labial: prothesis=insertion > C_1 change (C_1 deletion not significantly different from any other modification); Coronal–Velar: prothesis=C_1 deletion=C_1 change > insertion; Velar–Coronal: prothesis=C_1 deletion=C_1 change > insertion.)

3.2.4. Discussion

The results of Experiment 2 are partially consistent with the prediction that the ABX categorial discrimination task directs listeners' attention to phonetic information in the stimuli in order to create labels for A and B that help to categorize X. For the fricative–initial sequences, listeners did not show significant differences among the modification types for d' scores and the accuracy proportions were above 80% for almost all of the sequences (C_1 change for FN was at 79%). For SN, the most confusable modifications were C_1 change and C_1 deletion. For SS, the cluster was confused most often with the insertion modification. This pattern of confusions was the same in Experiment 1. The persistence of error patterns for stop–initial clusters across tasks may indicate that the phonological or phonetic cues to non-native cluster identity are either particularly weak in these cases, or are consistent with cues that listeners use to identify the most confusable modification in native language speech processing. This will be addressed further in Section 4.

The reaction time findings showed that for the stop–nasal sequences, listeners were slowest on C_1 change trials (and slower on C_1 deletion trials, though not significantly), and for stop–stop sequences, listeners were slowest on insertion trials; both of these findings are consistent with the accuracy results. For fricative–nasal clusters, listeners were slower on the insertion modification than on some of the other modifications (C_1 deletion and prothesis), and for fricative–stop clusters, listeners were slightly faster on C_1 deletion trials than some of the others. The latter point is also consistent with the accuracy/d' results, but the reason for increased reaction time for insertion in FN clusters is not immediately evident. The differences between reaction time findings for the AX discrimination and ABX categorial discrimination paradigms will be discussed further in Section 4.

For ABX categorial discrimination, again the place results mostly reflect the patterns present for the manner combinations. Coronal–coronal and labial–coronal combinations, which were only used in the ABX experiment, were comprised only of stop–nasal clusters and consequently, the d' score for C_1 change was significantly lower than other modifications. The only exception to this is that for the labial–coronal combination only, C_1 change and insertion were not significantly different from one another and were both significantly lower than the other modifications. This difference between coronal–coronal and labial–coronal sequences is likely due to the acoustic reflexes of homorganicity: bursts are short for both types of cluster, but they are 20 ms shorter for homorganic coronal–coronal sequences (see the table containing "segmental durations in CC sequences" for the ABX task in Appendix B). Thus, it is more obvious to the listener when a cluster is paired with the insertion modification for homorganic sequences (shorter burst) than for heterorganic ones (longer burst). Similar to AX, the coronal–labial and labial–labial sequences were comprised of fricative–initial and stop–nasal sequences. In the ABX study, since listeners were now very accurate on fricative–initial sequences, the main finding is that confusions on C_1 change trials for these place combinations were significantly worse than the other modifications. Finally, in ABX, velar–coronal and coronal–velar sequences showed an identical pattern. For both place combinations, performance on insertion modifications was significantly worse than all other modifications. Again, this reflects the fact that these place combinations were comprised of stop–stop sequences.

4. General discussion

A main purpose of using illusions in cross-language perception studies has been to investigate the relative markedness of two or more phonotactically unattested sequences. Markedness, here, refers to the degree to which the constraints of the grammar favor one structure to another. For example, Berent et al. (2007, 2008, 2009) used a vowel illusion to investigate the relative markedness of initial consonant clusters with different sonority profiles: falling sonority (e.g., [lbif]), rising sonority (e.g., [bnif]), and plateauing sonority (e.g., [bdif]). That research claims that listeners are less likely to have accurate perception of marked clusters (i.e., those with falling or plateauing sonority) than unmarked clusters (i.e., those with rising sonority) even when none of the clusters are permitted in the language of the experimental participants. Alternatively, Davidson (2011b) used illusory vowels to demonstrate that a sonority sequencing account of perception accuracy does not hold up when there are much smaller sonority differences between the consonants in the cluster.

The perceptual confusion of illegal phonotactics with legal counterparts, as in those studies described above, is often discussed in terms of segmental level patterns. This is in accord with the traditional view that the domain of phonotactics is governed by phonological processes operating over discrete categories. However, description at this level tells only part of the story. While it is evident that listeners do not always veridically perceive non-native sequences (e.g. Best & Tyler, 2007; Dupoux et al., 1999; Flege, 1995; Flege & Wang, 1990; Kuhl et al., 2008), it is possible to distinguish two levels at which native phonotactic patterns can influence the perception of novel phonotactic sequences. At the level of segmental patterns, perceptual confusion may arise because top down phonological knowledge prohibits accurate perception of unattested sequences. At the level of phonetic categorization, veridical perception of novel phonotactic sequences may be difficult because the phonetic environment obscures familiar cues to phonological categories (see also Dupoux et al., 2011 for a related but slightly different perspective on these issues). That is, the phonetic cues to a phonological category change according to phonetic environment. English listeners who lack experience with the consonant clusters in our study do not necessarily have the phonetic knowledge to parse familiar consonants from novel phonetic contexts. The unfamiliar phonetic context is a source of perceptual confusion that is distinct from direct top-down phonological repair.

In experimental designs that pair an illegal consonant cluster #CC with one possible alternative, e.g., #CCVX vs. #CəCVX, it is difficult to evaluate whether discrimination errors are due to phonetic or phonological interference. Both types of interference predict degraded performance. Low accuracy may therefore be attributable to language-specific phonetic knowledge, phonological knowledge, or to some combination of the two. In the current study, we investigated how unattested consonant clusters are perceived by examining the patterns of confusion between such clusters and a range of phonotactically permissible repairs. Our results reveal clear cases in which patterns of perceptual confusion are driven by language-specific phonetic knowledge, as opposed to direct interference from the phonology. We argue that confusions that persist across tasks – insertion in SS sequences and C_1 change/deletion in SN sequences – result from language-specific phonetic experience (indirectly influenced by the phonotactics of English), whereas confusions that only surface in the AX task – such as prothesis for fricative–initial sequences – have their basis in yet another source of illusion, namely the acoustic similarity of the stimulus items.

We begin the discussion with patterns that reveal the role of language-specific phonetic knowledge (Section 4.1) and then progress to patterns of confusion better accounted for by the acoustic similarity of the stimulus (Section 4.2). We conclude (Section 4.3) by summarizing the influence of four sources of illusion in consonant cluster perception (phonetic knowledge of the listener, acoustic properties of the stimuli, the experimental task, and the stimulus design) that need to be considered alongside top-down phonological knowledge when interpreting experimental results.

4.1. The role of phonetic experience in perceptual confusion

We begin our discussion of the role of language-specific phonetic knowledge with stop–stop (SS) sequences. For these sequences, confusion with the insertion modification persisted across both experimental tasks. This confusion has a likely basis in the phonetic experience of English listeners. Given the patterns of pre-tonic schwa reduction to which English speaking listeners are exposed, the stop burst after C_1 may indicate the presence of a reduced vowel. English listeners typically do not have extensive experience with a stop burst followed by another stop in the same word, since stops before stops word-medially or word-finally (e.g. napkin, act) are generally produced such that the burst of the initial stop is not audible (Crystal & House, 1988; Davidson, 2011a; Ghosh & Narayanan, 2009; Henderson & Repp, 1982). If English speakers do have experience with stop–stop sequences in which the first stop is released, they may be ones that are derived from intended #CəC sequences. Though deletion of pre-tonic schwa (e.g. [bɡ]in, [pt]ato) is not as common as is often assumed in American English (Patterson, LoCasto, & Connine, 2003), the vowel can be substantially reduced, leaving the stop burst as the primary indication that the speaker is producing a #CəC sequence (Davidson, 2006b). Consequently, the range of acoustic parameters associated with #SS–#SəS variants map consistently, for English speakers, to #CəC words. It is therefore not surprising that listeners tend to respond that the insertion modification is the same as the cluster for the SS sequence. This response can be attributed to experience with phonetic variation in English.

In the absence of evidence to the contrary, it is also possible that confusion distinguishing between #SS and #SəS sequences could be due, at least in part, to perceptual epenthesis, i.e., the top down insertion of a vowel. For comparisons between SS sequences and the insertion modification, it is difficult to determine whether the effect of native language phonotactics is direct, i.e., takes the form of perceptual epenthesis, or indirect, i.e., determines a listener's phonetic experience. However, if top-down phonological knowledge is responsible for the confusions for SS sequences, this would seem to be the only manner combination that is targeted by a phonological repair process. We now turn to stop–nasal sequences, where we show that it is unlikely that a top-down phonological repair can account for the confusion patterns.

For stop–nasal (SN) sequences, listeners were most likely to confuse the cluster with either the C_1 deletion or C_1 change modification, a result that, like insertion for stop–stop sequences, persisted across both AX and ABX tasks. These confusions are likely due in part to the fact that stop bursts have lower intensity before nasals than before other sounds. This can be seen by comparing the intensity of the stop burst in the stimuli (Appendix B) for SN sequences with SS sequences, which were rarely confused with C_1 change or C_1 deletion modifications. This comparison shows that the burst is indeed lower in intensity in SN sequences (SN mean=64 dB, s.d.=6; SS mean=67 dB, s.d.=5; $t(65)=2.25$, $p<0.03$).[2] The intensity of the stop burst is known to have a significant effect on the perception of place of articulation by English listeners (Ohde & Stevens, 1983). As shown in Appendix A, except for a modification of [d] to [z] for the [dm] cluster, all other C_1 change modifications involved a change to another stop (based on repairs observed in Davidson, 2010). The lowered intensity of the stop burst for SN sequences makes the identity of C_1 less perceptible to English-speaking participants, and therefore more confusable with either silence (C_1 deletion) or a different stop (C_1 change).

We extrapolate from the results on SN clusters that part of the phonetic knowledge of Russian listeners (but not English listeners) involves the ability to distinguish places of articulation on the basis of weak stop bursts (in combination with other cues) (Davidson, 2011b). As SN clusters are unattested in English, English listeners do not have experience parsing stop-place identity from the #_N environment.

English listeners also performed poorly on voicing distinctions preceding nasals, e.g., [tN] vs. [dN]. The most robust cue to voicing in English is voice onset time, the time between stop release and the onset of voicing of the following segment (Carney et al., 1977; Lisker & Abramson, 1964), and not voicing during the stop closure. However, since all Russian stops are produced with short voice onset time, and voicing during the closure distinguishes voiced from voiceless stops (Petrova, Plapp, Ringen, & Szentgyörgi, 2006), English listeners' language-specific experience regarding voiced stops may have made it difficult to discover that voicing during closure is what differentiates the Russian stimuli. In contrast, speakers of Slavic languages that contain these sequences are presumably sensitive to voicing during stop closure and can recover stop place from information in the burst alone. (This has not been verified experimentally with these stimuli, but the Russian listeners in Davidson (2011b) distinguished significantly better than English listeners on both long and short stimuli containing similar clusters.)

In our account of SN confusions, native phonotactics plays an indirect role. Because word-initial SN sequences are not attested in English, English listeners do not have the requisite phonetic knowledge to reliably parse the cues for stop-place in initial position, leading to confusions with C_1 change and C_1 deletion modifications. These confusions may not be the same for listeners of other language backgrounds, even if the segmental sequences used in this study are similarly unattested in those languages. Such a finding would be consistent with previous research, which has shown that speakers of disparate language backgrounds produce and perceive non-native sounds differently from one another (Flege, Bohn, & Jang, 1997; Ingram & Park, 1997; McAllister, Flege, & Piske, 2002; Willerman & Kuhl, 1996). Although language-specific "filters" affect perception, even in a domain like phonotactics that is traditionally considered to be governed by categorical phonological processes, segmental phonological processes are not sufficient to account for the observed confusions in SN clusters. To demonstrate this, we take a closer look at the confusion data.

As mentioned earlier in this section, most of the C_1 change modifications for the SN trials involved a change from one stop to another, whether that change is to a different place of articulation, or to a different value for voicing (e.g. [dm]→[bm] or [dm]→[tm]). Although these C_1 change modifications do not result in attested English clusters, it might be expected that listeners are biased toward confusing the SN cluster with another one that is less marked either for place or voice. Such an argument has been presented for other findings in studies of second language speech (e.g., Eckman, 1977). For example, Broselow, Chen, and Wang (1998) found that speakers of Mandarin sometimes devoiced voiced stops at the ends of English words, even though no stops at all are permitted word finally in Mandarin. They attributed this to a phenomenon called "emergence of the unmarked" (McCarthy & Prince, 1994), which occurs when speakers produce (or perceive) a phonological structure that is still unattested in their language, but one that is less marked than the target structure.

If the English listeners in our study are biased to "repair" their percept of SN clusters with one that is less marked along some dimension, then they should show confusions when a target voiced stop, e.g. [d̲m], is paired with its voiceless counterpart, e.g. [t̲m], because voiceless stops are less marked in this environment (sonority sequencing markedness, Parker, 2008), or when a labial or velar consonant, e.g. [b̲m] or [g̲m], is paired with a coronal consonant, e.g. [d̲m], because coronal consonants are less marked than labials and velars (place of articulation markedness, de Lacy, 2006). However, focusing on the ABX task in Experiment 2 as an example, the data in Table 5 demonstrate that listeners frequently confused the *X* stimulus with one that was more marked along either the voicing or place of articulation dimensions.

The notation "N" collapses over both C_2=[n] and [m], since place of articulation of C_2 had no significant effect. Bolded entries

Table 5
The number of C_1 change confusions for SN clusters in Experiment 2 (ABX) (n=186).

bN→		dN→		pN→		tN→	
pN	11	tN	11	**bN**	**15**	**dN**	**21**
dN	25	*bN*	*18*	tN	23	*pN*	*19*
vN	5	zN	4	*kN*	*12*	*kN*	*6*

See the text for further explanation of this table.

[2] Besides maximum burst intensity, we also considered other potential indices of stop burst salience including the rms amplitude of the burst, the rms burst amplitude relative to the following consonant, the rms burst amplitude relative to the following vowel, the duration of the burst, and duration weighted rms amplitude of the burst. Each of these measures show the same results that C_1 release burst is stronger in SS sequences than in SN sequences.

are the number of trials for which the listeners incorrectly responded that the *X* stimulus was the same as the more marked confusion for voicing. Italics indicate the number of trials for which the X stimulus was incorrectly confused with a cluster that was more marked for place of articulation.

Another possible explanation of the data in terms of segmental phonological processes is that listeners confuse some SN clusters because of regressive place assimilation – for example, they incorrectly respond 'A' in an ABX trial like [bmafa]-[dmafa]-[dmafa] (e.g., Dilley & Pitt, 2007; Gaskell, 2003; Gaskell & Snoeren, 2008). If so, then performance should be worse on trials in which the *X* stimulus is compared with a SN cluster in which the *S* and *N* differ in place features, the context for an assimilation rule, than on trials in which the *X* stimulus is compared to an *S* and *N* with the same place features, the context for a dissimilation rule. However, this prediction was not borne out in the data. The number of errors were equivalent in both dissimilation contexts ($N=44$) and assimilation contexts ($N=43$). These findings are inconsistent with an account positing that segmentally based phonological processes are responsible for preventing listeners from veridically hearing a cluster. Instead, the language-specific phonetic account detailed above better explains the pattern of results for stop–initial clusters.

In sum, the perceptual confusions that persisted across both the AX and ABX tasks, C_1 change and C_1 deletion for SN sequences and insertion for SS sequences, have a likely basis in the language-specific phonetic knowledge of the listener. For SN sequences we were able to distinguish between the plausible phonetic basis for confusion and a direct top-down influence of phonological knowledge. We identified a likely source of perceptual confusion grounded in language specific phonetic experience (indirectly influenced by native language phonotactics) and demonstrated that top-down phonological repair is not a plausible alternative. Perceptual modifications did not improve on markedness and, moreover, they did not pattern consistently, e.g., /b/ was heard as /p/ 11 times and /p/ was heard as /b/ 15 times. The inconsistency with which listeners misperceived consonants supports our claim that perceptual confusion in these cases is due to phonetic knowledge – listeners are unable to identify cues to consonant identity in the novel phonotactic environments – and not to a top-down phonological bias, which would presumably influence listener perception in a consistent direction.

We now turn to perceptual confusions that arose only in the AX task and discuss these along with the demands of the AX and ABX tasks.

4.2. *Fricative–initial sequences and the role of acoustic similarity*

In addition to the patterns of perceptual confusion that persisted across experimental tasks, we found some patterns that occurred only in the AX task. Each of these patterns involved the prothesis modification. We expected that listeners would be more accurate in the ABX task if the categorization demands of the ABX task required listeners to make use of the acoustic differences between A and B in order to appropriately label the A and B stimuli. This prediction was upheld only for prothesis modifications in fricative–initial sequences and SS clusters. The insertion modification for SS and both the C_1 change and C_1 deletion modifications for the SN sequences (discussed above in Section 4.1) were not significantly different in the AX and ABX experiments. Prothesis as a modification of fricative–initial clusters is a common cross-linguistic process that may be attributable to general properties of speech perception. If this type of influence is not due to perceptual organization based on native-language experience, but rather more general perceptual properties, it may be that listeners can overcome it more easily when the task requires that the participants label the stimuli.

Cross-linguistic surveys of vowel insertion have shown that it is common across languages to repair ill-formed word-initial fricative–consonant sequences (especially sibilant–initial sequences) with prothesis (Broselow, 1992; Fleischhacker, 2005; Gouskova, 2004; Karimi, 1987; Zuraw, 2007). Fleischhacker (2005) and Zuraw (2007) discuss an explanation for the preferred prothetic repair of fricative–consonant sequences in terms of the perceptual similarity between FC and əFC sequences. Fleischhacker conducted several experiments demonstrating that frication followed by silence or nasality, as in FC sequences, is more salient than frication followed by formant structure, as in FəC sequences. However, when the contiguity of the frication and either silence or nasality is retained, as in a comparison between the cluster, FC, and the prothetic modification, əFC, listeners rate the two types of utterances as very similar. By extension, this type of explanation – the contiguity of coarse-grained categories of waveform properties – can account for the relatively accurate perception of the FC contrasts with C_1 change and C_1 deletion. Since the duration of aperiodic noise is long, fricatives contain robust cues to their identity (Wright, 1996, 2004). Thus, changing or deleting a fricative C_1 would make it perceptibly different from the cluster it is being compared to (see also a similar claim in Kabak (2003) for perception of fricatives by Korean listeners). Given the apparent generality of the claim that prothetic modifications are perceptually similar to fricative–initial clusters and the fact that this does not persist in the ABX task, we propose that the pattern is attributable to the acoustic similarity of these sequences as opposed to language-specific knowledge of English speakers.

Finally, to reiterate, the results of the AX discrimination task suggest that for the most confusable modifications for each sequence type, listeners are behaving conservatively and only responding 'different' if they are certain of a difference. This may account for the poor sensitivity and accuracy for prothesis confusions in the AX task. This may occur even if listeners are primarily attending to information at an acoustic level in this task; that they are is indicated by the reaction time data for the AX task. Rather than mimicking the accuracy results, the AX reaction time results demonstrated that listeners are slower on C_1 deletion trials regardless of sequence. As noted in Section 2.3, this result could arise because the proportion of material shared between the C_1 deletion and CC stimuli is higher than for any other pair. The greater acoustic similarity between these items would most likely cause slower reaction times if speakers are performing the task by comparing continuous representations of acoustic information without labeling the items. In contrast, the results for the ABX task show that when there are significant differences in accuracy (i.e., for stop–initial sequences), there are similar differences in reaction time. It seems that for the ABX task, reaction time and accuracy are both more influenced by listeners' phonetic knowledge as it is deployed to label acoustic information in the A and B stimuli and to make decisions about *X* on the basis of the labels.

4.3. *Sources of illusion in consonant cluster perception*

We have demonstrated that, in addition to phonological knowledge, there are other sources of perceptual confusion that must be taken into account when interpreting experimental results. The phonetic knowledge of the listener develops in response to experience that is indirectly shaped by native language phonotactics. Listeners learn to parse phonological categories from the phonetic signals to which they are exposed. Since English lacks the consonant clusters presented in this study, English listeners lack the phonetic knowledge necessary to recover consonant sequences in these contexts. We argued that this type of language-specific phonetic knowledge is a distinct source of perceptual confusion from top-down phonological knowledge and we showed how only the former can account for the patterns of perceptual confusion that persisted across tasks.

In addition to phonetic and phonological knowledge, our results also reveal a number of task and stimulus-related influences on listener behavior. Some patterns of perceptual confusion persisted across both AX and ABX tasks, while others surfaced only in the AX task. For confusions that occurred only in the AX task, both reaction times and accuracy patterns suggest processing based on continuous representations. In this task only, comparisons that maintained contiguous gross waveform properties, e.g., the frication-nasality of fricative–nasal and frication-silence of fricative–stop clusters were frequently confused with prothetic modifications. This points to acoustic similarity as a cause of perceptual confusion that is distinct from language-specific phonetic knowledge and relevant only (or primarily) in the AX task.

Lastly, we point out that the design of the stimulus may also have an impact on accuracy patterns. This can be illustrated by comparing the results of Experiment 1 with those of the long words in the AX task in Davidson (2011b). Some of the #CC combinations were the same in both of these experiments, including FN, FS, and SN. In Davidson (2011b), where listeners were presented with a #CC–#CəC contrast on every trial, American English listeners had an accuracy of 34% on FN sequences, 27% on FS, and 42% on SN. All of these values were below chance, indicating that listeners could not tell the difference between #CC and #CəC. Because the stimuli in Experiment 1 of this study and the AX task in Davidson (2011b) used different stimuli, it is possible that differential performance across studies on the same sequences in the same kind of task (AX) are due to differences in the phonetic properties of the stimulus items. That is, the stimuli items in the two studies were produced by speakers of different languages (a Russian/English bilingual in this study and a native speaker of Catalan in Davidson, 2011b). Another possibility, however, is that the presence of variation in the stimulus set improved performance on the cluster-insertion trials in this study relative to Davidson (2011b). By presenting a #CəC item in the majority of the trials (the different trials and #CəC–#CəC same trials), the likelihood that the listeners are hearing a #CəC token at any particular time remains high and may have contributed to listeners' below-chance ability in Davidson (2011b) to distinguish between #CC and #CəC regardless of sequence type (as compared to Experiment 1, which showed differential performance across sequence types). For now, the comparison of this study and Davidson (2011b) is tentative, and we leave further exploration of this potential stimulus design effect to future research.

Overall, we have identified four factors – language-specific phonetic knowledge, the acoustic similarity of the stimulus items, the task (AX vs. ABX), and the stimulus design (one or multiple modification types) – that may contribute to perceptual illusions. Although perceptual illusions in non-native consonant clusters are often discussed in terms of top down phonological processes, we found little evidence for perceptual repairs of a categorical nature. Only in the case of SS-SəS comparisons did we see a possible role for perceptual repair and even for this case, the evidence is inconclusive. Instead we found consistent evidence for native phonotactics influencing perception indirectly, by conditioning the phonetic knowledge of the listeners. The most likely confusions for each consonant cluster have a likely basis in listener experience mapping continuous phonetic dimensions to discrete phonological categories.

5. Conclusion

The results of this study provide insight into how English listeners process the acoustic cues present in initial consonant clusters that are not possible in their native language. While listeners do not faithfully perceive non-native sequences, there is no one single modification that is the most perceptually confusable across the board. Rather, the most likely confusions between illicit and licit structures depend largely on the manner combination of segments that compose the sequence. For English, fricative–initial sequences may lead to prothesis illusions, stop–nasal sequences to deletion or change of the first consonant, and stop–stop sequences to vowel insertion. Extended analysis of these results revealed several sources of illusion in the perception of non-native sequences including the acoustic similarity of the licit–illicit stimuli pairs, the task itself (AX vs. ABX), and the number of different modifications included in the design (just insertion vs. insertion, prothesis, C_1 change, and C_1 deletion). Native language phonotactics, as emphasized in past studies, also

Table A1

	Fricative–Nasal		**Fricative–Stop**		**Stop–Nasal**		**Stop–Stop**	
Insertion	vmatu	vəmatu	vbano	vəbano	dmafa	dəmafa	ɡdase	ɡədase
C_1 del	vmatu	matu	vbano	bano	dmafa	mafa	ɡdase	dase
Prothesis	vmatu	əvmatu	vbano	əvbano	dmafa	ədmafa	ɡdase	əɡdase
C_1 chg	vmatu	fmatu	vbano	spano	dmafa	zmafa	ɡdase	ktase
	vmatu	smatu	vbano	fpano	dmafa	tmafa	ɡdase	bdase
	vmatu	zmatu	vbano	zbano	dmafa	bmafa		
Insertion	zmatu	zəmatu	zbano	zəbano	bmafa	bəmafa	ktase	kətase
C_1 del	zmatu	matu	zbano	bano	bmafa	mafa	ktase	tase
Prothesis	zmatu	əzmatu	zbano	əzbano	bmafa	əbmafa	ktase	əktase
C_1 chg	zmatu	fmatu	zbano	spano	bmafa	tmafa	ktase	ptase
	zmatu	smatu	zbano	fpano	bmafa	pmafa		
Insertion	fmatu	fəmatu	fpano	fəpano	pmafa	pəmafa	dɡase	dəɡase
C_1 del	fmatu	matu	fpano	pano	pmafa	mafa	dɡase	ɡase
Prothesis	fmatu	əfmatu	fpano	əfpano	pmafa	əpmafa	dɡase	ədɡase
C_1 chg	fmatu	smatu	fpano	spano	pmafa	tmafa	dɡase	tkase
	fmatu	zmatu			pmafa	dmafa	dɡase	bɡase
Insertion	smatu	səmatu	spano	səpano	tmafa	təmafa	tkase	təkase
C_1 del	smatu	matu	spano	pano	tmafa	mafa	tkase	kase
Prothesis	smatu	əsmatu	spano	əspano	tmafa	ətmafa	tkase	ətkase
C_1 chg							tkase	pkase

plays an important role in processing. We discussed two ways in which phonotactic knowledge can affect perception of illicit sequences: directly, through top-down phonological repair, and indirectly, through the language-specific interpretation of acoustic cues. Our results provide strong evidence for the indirect influence of phonotactics.

Acknowledgments

We would like to thank Tuuli Morrill for her assistance in collecting data. We also thank Catherine Best, audiences at the NYU Phonetics and Experimental Phonology Lab, the Acoustical Society of America, and the Linguistic Society of America, especially Paul Smolensky, for their comments on this study. This research was supported by the National Science Foundation CAREER Grant BCS-0449560.

Appendix A: experimental stimuli

Different trials for AX discrimination in Experiment 1 are shown below. Presentation of the items in the pairs was counterbalanced for order. In Experiment 2, the same stimuli were presented in ABX trials. Additional stop–nasal sequences beginning with Cn were also presented in Experiment 2 (e.g. dnalu, tnalu, bnalu, pnalu). These were also paired with insertion, C_1 deletion, prothesis, and C_1 change modifications (see Table A1).

Appendix B: Stimuli measurements

See Table B1.

Table B1
AX stimuli.

Stimulus item	Segmental durations in CVC sequences					C_1 burst intensity
	Total duration	C_1	Burst	Vowel	C_2	
fematu	369	216	NA	80	74	NA
sematu	302	173	NA	55	74	NA
vematu	276	104	NA	92	81	NA
zematu	256	117	NA	60	78	NA
fepano	246	98	NA	51	98	NA
sepano	342	191	NA	45	106	NA
vebano	333	154	NA	89	90	NA
zebano	279	123	NA	76	80	NA
bemafa	279	98	8	93	81	84
demafa	370	181	12	90	87	86
pemafa	192	NA	15	93	84	82
temafa	165	NA	23	61	81	80
degase	215	78	12	59	66	82
tekase	143	NA	17	52	75	82
gedase	247	87	18	58	84	70
ketase	212	NA	29	56	126	80

Stimulus item	Segmental durations in VCC sequences					C_1 burst intensity
	Total duration	C_1	Burst	Vowel	C_2	
efmatu	387	86	NA	140	161	NA
esmatu	293	65	NA	141	87	NA
evmatu	295	97	NA	104	94	NA
ezmatu	259	69	NA	100	89	NA
efpano	240	61	NA	87	92	NA
espano	274	59	NA	115	100	NA
evbano	218	80	NA	59	80	NA
ezbano	234	52	NA	117	65	NA
edmafa	254	55	74	16	110	59
etmafa	237	45	82	27	83	65
ebmafa	269	54	111	NA	104	NA
epmafa	303	59	143	NA	101	NA
egdase	177	44	65	14	54	72
ektase	248	45	63	35	105	62
etkase	234	50	60	39	84	60
edgase	212	58	57	39	58	77

Stimulus item	Segmental durations in CC sequences					C_1 burst intensity
	Total duration	C_1	Burst	Vowel	C_2	
fmatu	291	191	NA	NA	100	NA
smatu	272	195	NA	NA	77	NA
vmatu	209	115	NA	NA	93	NA
zmafa	234	112	NA	NA	122	NA
zmatu	232	127	NA	NA	106	NA
fpano	238	141	NA	NA	96	NA
spano	270	177	NA	NA	93	NA
spano	255	159	NA	NA	96	NA
vbano	196	112	NA	NA	84	NA
zbano	292	181	NA	NA	111	NA
tmafa	154	NA	27	NA	127	79
bmafa	234	107	NA	NA	127	NA
dmafa	260	151	NA	NA	109	NA
pmafa	161	NA	NA	NA	161	NA
bgase	202	105	8	NA	89	71
dgase	175	81	13	NA	81	67
gdase	241	109	30	NA	95	75
ptase	120	NA	26	NA	94	59
ktase	137	NA	30	NA	107	69
tkase	190	NA	37	NA	152	67

ABX stimuli. Durations are averages of four individual tokens, with standard deviation in parentheses

Stimulus item	Segmental durations in CVC sequences					C_1 burst intensity
	Total duration	C_1	Burst	Vowel	C_2	
fematu	288(19)	137(17)	NA	75(20)	75(2)	NA
sematu	316(12)	168(15)	NA	76(7)	72(4)	NA
vematu	257(23)	89(13)	NA	93(10)	76(9)	NA
zematu	269(10)	96(13)	NA	96(10)	77(2)	NA
fepano	308(25)	134(16)	NA	81(4)	94(9)	NA
sepano	322(4)	155(27)	NA	71(2)	97(21)	NA
vebano	237(9)	81(10)	NA	81(3)	75(2)	NA
zebalo	269(25)	114(10)	NA	83(8)	73(8)	NA
zebano	280(21)	116(14)	NA	81(9)	83(11)	NA
bemafa	244(39)	73(33)	9(3)	83(11)	80(3)	76(2)
benalu	244(10)	81(13)	9(2)	92(9)	61(3)	76(3)
demafa	203(16)	41(9)	17(10)	65(5)	80(5)	72(3)
denafe	224(36)	59(32)	14(1)	90(1)	62(4)	73(2)
denalu	234(24)	73(22)	9(2)	88(9)	65(2)	74(3)
gemafa	261(18)	91(15)	14(3)	77(2)	79(4)	72(0)
pemafa	161(5)	NA	15(2)	77(5)	70(1)	72(1)
penalu	157(22)	NA	15(3)	79(16)	63(4)	72(4)
temafa	183(14)	NA	31(11)	73(8)	79(7)	67(2)
tenalu	183(17)	NA	37(8)	79(5)	67(4)	69(1)
degafe	192(15)	55(15)	11(3)	74(5)	52(2)	73(2)
degase	210(16)	67(15)	15(3)	77(8)	50(8)	74(3)
gedase	260(60)	73(26)	44(34)	74(11)	69(7)	75(6)
ketase	171(14)	NA	32(5)	49(14)	91(6)	69(3)
tekase	167(9)	NA	25(6)	64(8)	79(5)	68(2)

Stimulus item	Segmental durations in VCC sequences					C_1 burst intensity
	Total duration	Vowel	C_1	Burst	C_2	
efmatu	289(4)	77(5)	131(8)	NA	81(9)	NA
esmatu	283(23)	71(23)	124(9)	NA	88(10)	NA
evmatu	250(11)	84(10)	77(14)	NA	89(11)	NA
ezmatu	253(28)	76(18)	85(11)	NA	92(5)	NA
efpano	273(16)	76(8)	108(7)	NA	89(5)	NA

Table B1 (*continued*)

Stimulus item	Segmental durations in VCC sequences					C_1 burst intensity
	Total duration	Vowel	C_1	Burst	C_2	
evbano	236(17)	84(11)	70(14)	NA	82(9)	NA
ezbano	228(22)	69(9)	82(16)	NA	77(2)	NA
ebmafa	244(14)	63(5)	87(9)	6(7)	88(21)	67(1)
ebnalu	226(26)	51(18)	82(9)	17(9)	76(10)	67(5)
edmafa	256(14)	62(12)	84(10)	15(2)	95(16)	65(4)
ednalu	258(19)	76(9)	87(12)	13(4)	83(17)	66(2)
epmafa	291(25)	64(8)	118(20)	27(6)	83(11)	60(3)
epnalu	259(15)	54(7)	98(4)	43(5)	64(15)	61(2)
etmafa	255(10)	55(19)	93(15)	30(7)	76(20)	62(3)
etnalu	257(13)	66(14)	98(7)	24(6)	68(9)	62(1)
edɡase	206(26)	59(14)	60(6)	17(8)	69(11)	66(4)
eɡdase	230(20)	71(19)	58(3)	27(4)	75(2)	69(1)
ektase	243(23)	58(10)	72(5)	33(10)	81(9)	68(2)
etkase	214(14)	52(4)	62(6)	22(4)	77(8)	66(2)

Stimulus item	Segmental durations in CC sequences				C_1 burst intensity
	Total duration	C_1	Burst	C_2	
fmatu	271(48)	183(41)	NA	88(14)	NA
smatu	270(8)	180(10)	NA	90(9)	NA
vmatu	215(33)	112(22)	NA	103(13)	NA
vnalu	211(18)	108(15)	NA	103(7)	NA
zmafa	242(8)	142(3)	NA	99(8)	NA
zmatu	220(13)	125(16)	NA	96(6)	NA
fpano	244(25)	154(23)	NA	90(3)	NA
vbano	175(28)	99(24)	NA	76(6)	NA
zbano	205(19)	118(21)	NA	87(8)	NA
bmafa	225(8)	108(26)	21(9)	105(26)	65(4)
bnalu	200(29)	76(23)	31(17)	96(5)	71(4)
dmafa	218(12)	101(28)	23(10)	105(22)	70(1)
dnalu	235(19)	119(16)	4(8)	112(12)	NA
ɡnalu	224(18	92(19)	33(7)	99(6)	72(2)
knalu	154(5)	NA	91(6)	66(5)	60(1)
pmafa	132(10)	NA	41(8)	90(16)	59(4)
pnalu	134(11)	NA	50(6)	85(13)	62(4)
tmafa	154(19)	NA	76(14)	80(19)	60(3)
tnalu	131(10)	NA	30(1)	102(11)	63(3)
bdase	190(4)	84(12)	28(10)	78(8)	69(7)
bɡase	197(26)	89(21)	27(10)	84(4)	72(3)
dɡase	177(34)	57(24)	31(6)	89(11)	71(2)
ɡdase	191(14)	83(13)	33(2)	75(6)	72(4)
ktase	131(11)	NA	40(2)	91(11)	67(1)
pkase	117(15)	NA	24(5)	92(17)	61(5)
ptase	117(5)	NA	22(10)	94(4)	64(3)
tkase	114(8)	NA	29(9)	86(4)	67(4)

References

Berent, I., Lennertz, T., Jun, J., Moreno, M., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, *105*, 5321–5325.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology*, *26*, 75–108.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*, 591–630.

Best, C., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, *109*(2), 775–794.

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In: M. Munro, & O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13–34). Amsterdam: John Benjamins.

Broselow, E. (1992). Transfer and universals in second language epenthesis. In: S. Gass, & L. Selinker (Eds.), *Language transfer in language learning (revised edition)* (pp. 71–86). Amsterdam: John Benjamins.

Broselow, E., Chen, S., & Wang, C. (1998). The emergence of the unmarked in second language phonology. *Studies in Second Language Acquisition*, *20*, 261–280.

Brown, R., & Hildum, D. (1956). Expectancy and the perception of syllables. *Language*, *32*(3), 411–419.

Carney, A. E., Widin, G., & Viemeister, N. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, *62*(4), 961–970.

Crystal, T., & House, A. (1988). The duration of American-English stop consonants: An overview. *Journal of Phonetics*, *16*, 285–294.

Davidson, L. (2006a). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, *34*(1), 104–137.

Davidson, L. (2006b). Schwa elision in fast speech: Segmental deletion or gestural overlap?. *Phonetica*, *63*(2-3), 79–112.

Davidson, L. (2007). The relationship between the perception of non-native phonotactics and loanword adaptation. *Phonology*, *24*(2), 261–286.

Davidson, L. (2010). Phonetic bases of similarities in cross-language production: Evidence from English and Catalan. *Journal of Phonetics*, *38*(2), 272–288.

Davidson, L. (2011a). Characteristics of stop releases in American English spontaneous speech. *Speech Communication*, *53*(8), 1042–1058.

Davidson, L. (2011b). Phonetic, phonemic, and phonological factors in cross-language discrimination of phonotactic contrasts. *Journal of Experimental Psychology: Human Perception & Performance*, *37*(1), 270–282.

de Lacy, P. (2006). *Markedness: reduction and preservation in phonology*. Cambridge: Cambridge University Press.

Dehaene-Lambertz, G., Dupoux, E., & Gout, A. (2000). Electrophysiological correlates of phonological processing: A cross-linguistic study. *Journal of Cognitive Neuroscience*, *12*, 635–647.

Demuth, K. (1996). The prosodic structure of early words. In: J. Morgan, & K. Demuth (Eds.), *Signal to syntax*. Hillsdale, NJ: Erlbaum.

Dilley, L., & Pitt, M. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *Journal of the Acoustical Society of America*, *122*, 2340–2353.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion? *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1568–1578.

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. *Language and Cognitive Processes*, *16*(5/6), 491–505.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from? *Journal of Memory and Language*, *64*, 199–210.

Durlach, N., & Braida, L. (1969). Intensity perception I: Preliminary theory of intensity resolution. *Journal of the Acoustical Society of America*, *46*, 372–383.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, *27*, 315–330.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In: W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research* (pp. 229–273). Timonium, MD: York Press.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers production and perception of English vowels. *Journal of Phonetics*, *25*, 437–470.

Flege, J. E., & Wang, C. (1990). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics*, *17*, 299–315.

Fleischhacker, H. (2005). *Similarity in phonology: evidence from reduplication and loan adaptation*. PhD dissertation, UCLA, Los Angeles, unpublished.

Gallagher, G. (2010). Perceptual distinctness and long-distance laryngeal restrictions. *Phonology*, *27*, 435–480.

Gaskell, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics*, *31*(3–4), 447–463.

Gaskell, M. G., & Snoeren, N. (2008). The impact of strong assimilation on the perception of connected speech. *Journal of Experimental Psychology: Human Perception & Performance*, *34*(6), 1632–1647.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics*, *66*(3), 363–376.

Ghosh, P. K., & Narayanan, S. (2009). Closure duration analysis of incomplete stop consonants due to stop–stop interaction. *Journal of the Acoustical Society of America*, *126*(1), EL1–EL7.

Gouskova, M. (2004). Relational hierarchies in Optimality Theory: The case of syllable contact. *Phonology*, *21*(2), 201–250.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Hallé, P., & Best, C. (2007). Dental-to-velar perceptual assimilation: A cross-linguistic study of the perception of dental stop+/l/ clusters. *Journal of the Acoustical Society of America*, *121*(5), 2899–2914.

Hallé, P., Dominguez, A., Cuetos, F., & Segui, J. (2008). Phonological mediation in visual masked priming: Evidence from phonotactic repair. *Journal of Experimental Psychology: Human Perception & Performance*, *34*(1), 177–192.

Hallé, P., Segui, J., Frauenfelder, U., & Meunier, C. (1998). Processing of illegal consonant clusters: A case of perceptual assimilation? *Journal of Experimental Psychology: Human Perception and Performance*, *24*(2), 592–608.

Henderson, J., & Repp, B. (1982). Is a stop consonant released when followed by another stop consonant?. *Phonetica*, *39*, 71–82.

Ingram, J., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics*, *25*, 343–370.

Kabak, B. (2003). *The perceptual processing of second language consonant clusters*. PhD dissertation, unpublished.

Kabak, B., & Idsardi, W. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints? *Language and Speech*, *50*, 23–52.

Karimi, S. (1987). Farsi speakers and the initial consonant cluster in English. In: G. Ioup, & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 305–318). Cambridge, MA: Newbury House.

Kuhl, P., Conboy, B., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., & Nelson, T. (2008). Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B*, *363*, 979–1000.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Massaro, D., & Cohen, M. (1983). Phonological context in speech perception. *Perception & Psychophysics*, *34*(3), 338–348.

Matthews, J., & Brown, C. (2004). When intake exceeds input: Language specific perceptual illusions induced by L1 prosodic constraints. *International Journal of Bilingualism*, *8*(1), 5–27.

McAllister, R., Flege, J., & Piske, T. (2002). The influence of the L1 on the acquisition of Swedish vowel quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, *30*, 229–258.

McCarthy, J., & Prince, A. (1994). The emergence of the unmarked: Optimality in prosodic morphology. In: M. Gonzales (Ed.), *Proceedings of the 24th North East Linguistics Society*. Somerville: Cascadilla Press.

McCarthy, J., & Prince, A. (1996). Prosodic morphology. In: J. Goldsmith (Ed.), *The handbook of phonological theory* (pp. 318–366). Cambridge, MA: Blackwell.

Moreton, E. (2002). Structural constraints in the perception of English stop–sonorant clusters. *Cognition*, *84*, 55–71.

Ohde, R. N., & Stevens, K. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. *Journal of the Acoustical Society of America*, *74*, 706–714.

Parker, S. (2008). Sound level protrusions as physical correlates of sonority. *Journal of Phonetics*, *36*(1), 55–90.

Patterson, D., LoCasto, P. C., & Connine, C. M. (2003). Corpora analyses of frequency of schwa deletion in conversational American English. *Phonetica*, *60*, 45–69.

Peperkamp, S. (1999). Prosodic words. *GLOT International*, *4*, 15–16.

Petrova, O., Plapp, R., Ringen, C., & Szentgyörgi, S. (2006). Voice and aspiration: Evidence from Russian, Hungarian, German, Swedish, and Turkish. *The Linguistic Review*, *23*, 1–35.

Pisoni, D. (1973). Auditory and phonetic codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, *13*, 253–260.

Pitt, M. (1998). Phonological processes and the perception of phonotactically illegal consonant clusters. *Perception & Psychophysics*, *60*(6), 941–951.

Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic and acoustic contributions. *Journal of the Acoustical Society of America*, *89*(6), 2961–2977.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Selkirk, E. (1984). On the major class features and syllable theory. In: M. Aronoff, & R. Oehrle (Eds.), *Language sound structure*. Cambridge, MA: MIT Press.

Shaw, J., & Davidson, L. (2011). Perceptual similarity in input–output mappings: A computational/experimental study of non-native speech production. Lingua, doi:10.1016/j.lingua.2011.03.003.

Strange, W., & Shafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In: J. G. Hansen Edwards, & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 153–192). Philadelphia: John Benjamins.

Werker, J., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, *37*(1), 35–44.

Willerman, R., &Kuhl, P. (1996). Cross-language speech perception: Swedish, English, and Spanish speakers' perception of front rounded vowels. In W. Idsardi & T. Bunnell (Eds.), *Proceedings of ICSLP 96*.

Wright, R. (1996). *Consonant Clusters and Cue Preservation in Tsou*. PhD dissertation, UCLA, unpublished.

Wright, R. (2004). A review of perceptual cues and cue robustness. In: B. Hayes, R. Kirchner, & D. Steriade (Eds.), *Phonetically based phonology*. Cambridge: Cambridge University Press.

Zuraw, K. (2007). The role of phonetic knowledge in phonological patterning: corpus and survey evidence from Tagalog infixation. *Language*, *83*(2), 277–316.