

Novel-category biases in second language perception and production

Kenneth de Jong¹²³, Yen-Chen Hao¹, Hanyong Park⁴, and Noah Silbert¹²

¹Dept. of Linguistics, Indiana University, 322 Memorial Hall, 1021 E. 3rd St.,
Bloomington, IN 47405

²Dept. of Cognitive Science, Indiana University

³Dept. of Second Language Studies, Indiana University

⁴Speech Research Lab, Dept. of Psychological and Brain Sciences, Indiana University

Abstract: Silbert & de Jong (2007) investigated correlations between response bias and segment frequency in a number of previously published segment identification data sets. In general, bias tends to be positively correlated with frequency. When applied to second language (L2) learning, this pattern predicts biases toward identifying novel L2 categories as L1 categories, since the number of instances of L1 categories in learners' experience far outweighs those of novel categories. This paper reports experiments in which 20 Korean EFL (English as a Foreign Language) learners identified English obstruents produced by native speakers of English. In a second experiment, the Korean EFL learners produced the same obstruents, which were then identified by native English listeners. Cases of novel category response bias are observed sporadically in L2 identifications, but not in L2 productions. Such novel category effects support a model in which novel categories serve, in part, as indicators of the L2 itself and stand out as response options in L2 identification tasks. These effects may be counteracted by motor difficulties in L2 production tasks that tend to preserve the substance of the L1 phonological categories.

1. Introduction.

1.1. Frequency Biasing

Frequency effects have been observed in linguistic behavior in a variety of tasks, and at a number of linguistic levels. For example, lexical decision and naming are faster for high frequency than low frequency words (Forster & Chambers, 1973), and lexical decision about a given target word is inhibited more by high frequency neighbors than low frequency neighbors (Luce & Pisoni, 1998). Frequency effects have also been observed with regard to the phonological structures of non-words; non-words with more frequent phonotactic patterns are judged to be more word-like than those with less frequent phonotactic patterns (Frisch, Large & Pisoni, 2000).

Silbert & de Jong (2007) examined frequency effects at the segment level. This study assumed that identification behavior has two components: perceptual similarity, which determines the confusability of targets, and decision bias, which modulates perceptual similarity to produce responses. The working hypothesis was that listeners would be systematically biased toward responding with higher frequency categories, regardless of the input. In Luce's Similarity Choice Model, this tendency is modeled with response bias parameters (b) for each response alternative. Silbert & de Jong (2007) estimated bias parameters for various previously published identification data sets and probed correlations between the bias parameters and measures of segmental token frequency, segmental type frequency (taken from the Hoosier Mental Lexicon; Nusbaum, Pisoni & Davis, 1984), and word frequency (taken from the Brown Corpus; Kucera & Francis, 1967).

Silbert & de Jong (2007) examined nine corpora of segment identification data. Five of these corpora are of native English listeners, listening in noise, and were taken from Miller & Nicely (1955), Luce (1986), and Cutler, Weber, Smits & Cooper (2004). Four additional corpora were of Koreans listening to English, taken from corpora collected as part of our research (Park & de Jong, 2006, 2008; Park, 2007). In eight of the nine corpora, bias was positively correlated with segmental type frequency and word frequency. Figure 1 presents correlation coefficients for segmental type frequency and bias parameters for these nine corpora. The values on the x-axis do not indicate prosodically restricted frequency, while those on the y-axis do.

Figure 1 plots Spearman rank order correlation coefficients (r_s) calculated between bias parameters (b) and two measures of segmental type frequency (overall segment frequency on the x-axis, frequency in the appropriate prosodic location (onset, coda, or intervocalic) in the y-axis). Each token indicates one of the identification data sets. The point of interest for the current paper is that all but one of the data sets exhibits a positive correlation for both frequency measures, the more frequent a segment is, the more positive the bias toward identifying a stimulus as that segment. Most of the data sets exhibited regression coefficients with both measures of frequency that are quite far from zero. Two of the subsets in the Miller & Nicely (1955) data exhibit very substantial correlations, one with $r_s > 0.8$. Bayesian analyses of these correlations support the overall reliability of these observations that bias pervasively appears as a function of frequency of occurrence.

1.2. L2 Old Category Biases

The current paper examines second language (L2) perception data in light of the connection between frequency and bias in native speakers. If this connection between frequency of experience and response bias also holds for L2 learners, we should observe that L2 learners exhibit what are typically called substitution errors. A substitution error is typically defined as occurring when a learner is faced with confusion between two L2 categories, and substitutes (responds with) the category that is more similar to a native category.

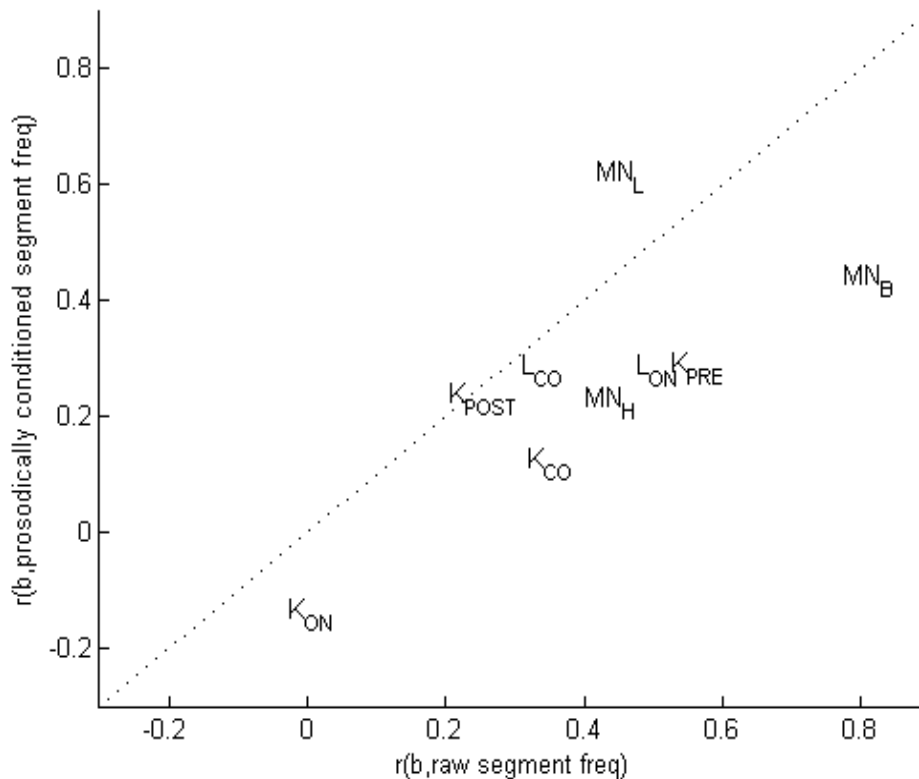


Figure 1. Spearman rank order correlation (r_s) for correlations between response bias (b) for each segment and estimates of that segment's frequency in a particular prosodic location (y -axis) plotted against r calculated with frequency estimates across prosodic locations (x -axis). Frequency estimates are from the Hoosier Mental Lexicon. Each data point represents a corpus of segmental identification data: Miller & Nicely (1955 – MN) in broadband (MN_B), low-pass filtered (MN_L), and high-pass filtered (MN_H) conditions; Luce (1986 – L), in onsets (L_{ON}) and codas (L_{CO}); or Park & de Jong (2008) in onsets (K_{ON}), Park & de Jong (2007) in codas (K_{CO}), Park (2007) in pre-stress intervocalic (K_{PRE}) and post-stress intervocalic (K_{POST}) conditions.

The substitution error pattern fits that predicted by frequency biasing. L2 segments that correspond closely to L1 segments, if they count as instances of the L1 segment, should be much more common in the experience of the learner. The amount of exposure to L1 sounds in less experienced learners is vastly larger than is their exposure to L2 sounds. In cases of Two-to-One Mapping, i.e., two L2 categories to one L1 category, there will be confusions (high perceptual similarity) between two categories in the L2. In such cases, we expect the confusion to be resolved in the direction of the more frequently experienced category, which, in the case of a second language, would be the category that is most neatly mapped onto a familiar L1 category. Thus, we expect biases toward L1-similar sounds, and away from novel categories in L2 production and perception tasks. For example, Korean learners of English often will exhibit confusions between English

/f/ and /p/; in such cases, there is anecdotal evidence for a tendency for Korean learners to produce /p/, rather than /f/, creating lexical errors such as ‘fork’ → ‘pork’. Here, the /p/ is much more frequent in the learner’s experience, and hence, this might be a case of frequency biasing.

1.3. Novel Category Biases?

There are occasional observations, however, of response outcomes that run opposite these predictions, that is, of biases *toward* a novel category, away from the category that is similar to that in the L1. In production, one case can be seen in the data in Flege’s (1987) study of high vowels produced by various English/French bilinguals. Flege measured F2 values in productions of words with /u/, in order to examine how various learner groups cope with the fact that /u/ in French has typically lower F2 values than /u/ in English. One of the learner groups examined consisted of relatively inexperienced American native English speakers learning French in an American university setting. While the general pattern in the various groups was to produce French /u/ with F2 values between those of monolingual French and English talkers, this inexperienced French L2 group actually produced F2 values that were higher than those in English /u/. The most likely explanation for this is that some (or all) of these English speakers tended to produce French /u/ as /y/. If this explanation is right, these higher F2 values would suggest a tendency toward novel category substitution, exactly in the opposite direction of the prediction described above.

Such effects have also been noted with perceptual data, again with relatively inexperienced learners. Nagao, Lim & de Jong (2003) examined perceptual identifications of syllable affiliation in ambiguously syllabified stop consonants. This study examined the nature of previously-found rate effects on perceived syllabification of singleton consonants in intervocalic position. A rate variation paradigm (de Jong, 2001) was used to induce native English talkers to produce stop consonants in postvocalic position (e.g. as in a word such as *eeɸ*) repetitively at fast speech rates. Such productions have been found to be ambiguous between pre-vocalic and post-vocalic position for native English listeners (de Jong, Nagao, & Lim, 2004). In order to examine syllabification patterns for non-native listeners, the same stimuli were played to three groups of Japanese L2 speakers, ESL (English as a Second Language) learners in their late 20’s and 30’s in the United States with a large amount of experience with native spoken English, EFL listeners in their 50s with very little experience with native spoken English, and EFL learners in an undergraduate population, who also had limited experience with spoken English.

Figure 2 gives results for syllabic affiliation responses for the three groups with native listener controls. While the native speaker controls identified the stimuli as being words of a CV form (‘Onset Response’) at a rate of just over 60%, non-native listeners exhibited a stronger tendency to label the stimuli as CV forms. Since Japanese does not allow coda stops, this would constitute an L1 substitution pattern; i.e. what we characterize above as an old category biasing effect. Curiously, the younger, relatively inexperienced Japanese EFL learners (third bar) tended to label ambiguously syllabified

consonants as VC forms ('Coda Response') more often than did experienced listeners. An analysis of variance showed this difference to be significant, and also post-hoc tests indicated that, while the other two groups of Japanese listeners were significantly different from native listeners, these younger Japanese EFL listeners were not different from the native speaker controls.

Nagao, *et al.* (2003) suggested that this difference between the inexperienced college-aged group and the other non-native groups constitutes a novel category bias, and that this might be typical of relatively inexperienced learners. However, as is also clear from these results, the effect was not obtained with the older listeners who rarely, if ever, used their English training from school, and were not particularly oriented towards English speaking culture. Thus, the effect could be determined by a number of factors, including degree of L2 experience, type of training in the language, the predisposition of the learner toward the culture, and even the effects of aging.

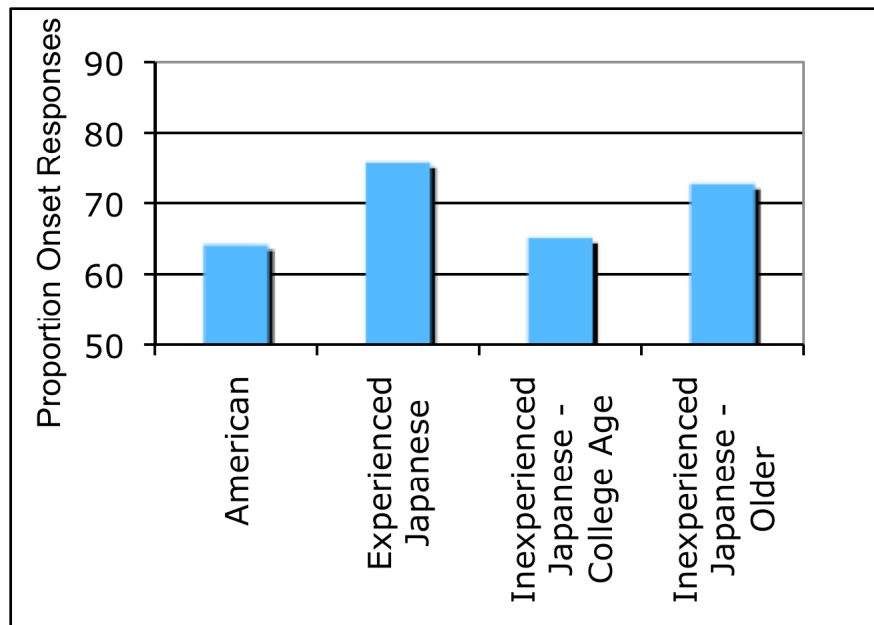


Figure 2. Proportion of onset responses to fast-rate repetitions of syllables with singleton coda consonants for four listener groups. Data taken from Nagao *et al* (2003).

Two general questions are posed in the current paper. First, are novel category biases evident in a corpus of segment identification data from non-native EFL listeners? The corpus of data examined here is that which is indicated in Figure 1 as K_{on} , that is, Korean identification of L2 onset consonants. It is obvious from Figure 1 is that K_{on} is peculiar in that no overall segmental frequency – bias correlations were evident. It might be that novel category biases override any segmental frequency effects in the data.

Second, are novel category biases equally evident in segmental perception and production tasks? While many treatments of second language acquisition do not discuss differences between production and perception, it is possible that biases function quite differently in the two modes. Most cases of substitution are generally found in the production mode, while substitution in the perceptual mode is not as commonly noted in the literature. Again, it is possible that such differences may partially be due to the effects of novel category biases in a particular mode.

2. Methods

2.1. Materials

The targets for the frequency analyses are 10 anterior obstruents in English, as identified and produced by 20 Korean EFL learners. Based on orthographic mapping data in Park & de Jong (2008), the segments can be broken down into similar (old) and new classes. The segments similar to segments in Korean are /p, b, t, d/, while the segments not similar to Korean segments are /f, v, θ, ð/. The orthographic mapping data in Park & de Jong (2008) was collected by presenting stimuli containing consonant segments of various types to the Korean listeners, and asking them to identify the segment according to Korean orthographic renditions of native Korean consonant categories. With each response, the listeners were asked to judge the goodness of the fit between the English production and the chosen Korean category. The four stop consonants were systematically rated as better fits to Korean categories than the fricative consonants. In addition, the mapping for the fricatives tended to be inconsistent, eliciting a larger number of different Korean categories as responses. Our current corpus also includes /s/ and /z/, which were not in the Park & de Jong (2008) corpus. We do expect the /s/ and /z/ to be treated as ‘old’ (like the stops) and ‘new’ (like the non-sibilant fricatives) respectively. While Korean has /s/, there are no categories transcribed as /z/ in typical Korean productions, and learners’ productions of English words with /z/ tend to exhibit the substitution of a voiced affricate.

2.2. Stimuli

Stimuli for the identification experiment consisted of productions of these 10 consonants (and others, not analyzed here), preceding the vowel /a/. These were produced in isolation by 4 midwestern American speakers (2 male, 2 female). The native English talkers were cued with cards, spelling the target syllables with normal English orthography. In addition, the dental fricatives were cued with IPA renderings of the two fricatives, and the talkers were checked for their understanding of the symbols before recording.

Recordings were done in a quiet room, using a DAT recorder sampling at 44.1 kHz, and ported to various computer platforms for editing and randomization. For presentation, the stimuli were burned into compact disks that were then presented to listeners in free-field conditions in a non-noisy environment.

2.3. Participants

Participants were 20 Korean learners of English, who were undergraduate students at Kyonggi University, near Seoul, South Korea. As undergraduates, their ages ranged from 22 – 28 years at the time of experimentation (2003). 14 of the participants were female, and 6 were male. None of the participants had spent more than 3 months in an English speaking country. Thus, though each of the students had studied English for an extensive period of time in EFL classes, the degree of exposure to English spoken by native speakers was low.

2.4. Tasks

Each of the participants was involved in three tasks. The second and third tasks were grouped together, and the order of tasks A and tasks B and C was counterbalanced across participants. In our discussion of the results of these tasks, we will assume that most of the confusions in the data occur because of the EFL learners, so evaluations of the **Id** data will reflect confusions introduced in their perceptions, and evaluations of the **Rd** data will reflect confusions introduced in their productions. The **Mim** data is more complicated, involving both perception and production on the part of the EFL learners.

A. Identification (Id): listeners were presented with the stimuli with 5 second intervals in a quiet classroom, and were asked to identify the consonant by marking the appropriate symbol on a response form. On the response form for each trial were a set of 15 response options, and an additional option of writing in an alternative not on the list (“Other ____”). The response alternatives, along with the instructions in English are included in the Appendix. After a block of trials, the experimenter entertained questions from the participants to insure they understood the task. The alternatives were determined by means of a pilot experiment to contain all of the likely responses for the consonant set. The write-in option was rarely (3.2% of the time) used for the obstruents analyzed here.

B. Reading (Rd): talkers were asked to read the same list of forms as the American speakers who produced the stimuli, and were recorded at 44.1 kHz using a portable mini-disc recorder. Each consonant was repeated twice. Recordings were converted to .wav files using *Audacity*.

C. Mimicry (Mim): in the same session, either before or after the reading tasks, the talkers listened to the American stimuli and were asked to repeat the syllable (with no visual cues). Since there are four stimuli per consonant, there are four mimicked productions of each consonant from each talker.

All of the productions from tasks 2 and 3 were, and were randomized in blocks of four talkers and presented to 10 native English speaking judges each. Much as in the Identification task, the native English speaking judges were seated in a room in groups of 1 to 5, and were presented the stimuli over a loud speaker and asked to identify the produced consonants using the same response options as in Task 1.

2.5. Analyses

Similarity Choice Model bias parameters (Luce, 1963), estimates of an overall tendency for a particular category to be used as a response label, were calculated for each segment according to the following formula:

$$b(j) = \frac{1}{\sum_{i=1}^n \sqrt{\frac{p(i|j)p(i|i)}{p(j|i)p(j|j)}}}$$

Here, $p(j|i)$ is the probability of responding with the j^{th} response label when presented with the i^{th} stimulus. The larger the probability of giving the j^{th} response, the larger the denominator in the ratio of the products of the probabilities, and hence, the larger the estimated bias. Bias estimates from the three tasks, averaged across all talkers and listeners for each segment, were then compared to each other.

3. Results

Figure 3 presents bias parameters for each of the 10 consonants taken from the data in each of the three tasks. Both panels plot production parameters on the y -axis against perception parameters on the x -axis, the top panel corresponding to reading data and the bottom panel mimicry data. While bias parameters vary from segment to segment, there appears to be a systematic difference between the ‘new’ and ‘old’ segments across the L2 production and L2 perception tasks. For ‘old’ segments, the bias parameters are higher for the L2 production tasks than for the L2 identification task; for the ‘new’ segments, the bias parameters are higher for the L2 identification task than for either of the L2 production tasks. This is evident in the ‘old’ segments consistently appearing above the diagonal, while the ‘new’ segments consistently appear below the diagonal.

Note that this difference between the two sets of sounds is only found in comparing the production and perception data. If we examine the production data alone, there are pretty clear tendencies for higher bias parameters for ‘old’ segments, though /s/ and the voiced stops are quite similar to the novel non-sibilant fricatives. If we examine the perception data alone, there is a considerable degree of overlap between ‘old’ and ‘new’ segments. For example, the bias parameters for the new segments /v/ and /θ/ are quite similar to that of the old segment /b/. However, regardless of this range of bias parameter values, the difference in bias parameters *between* the tasks seems to be systematically related to the ‘new’ vs. ‘old’ distinction.

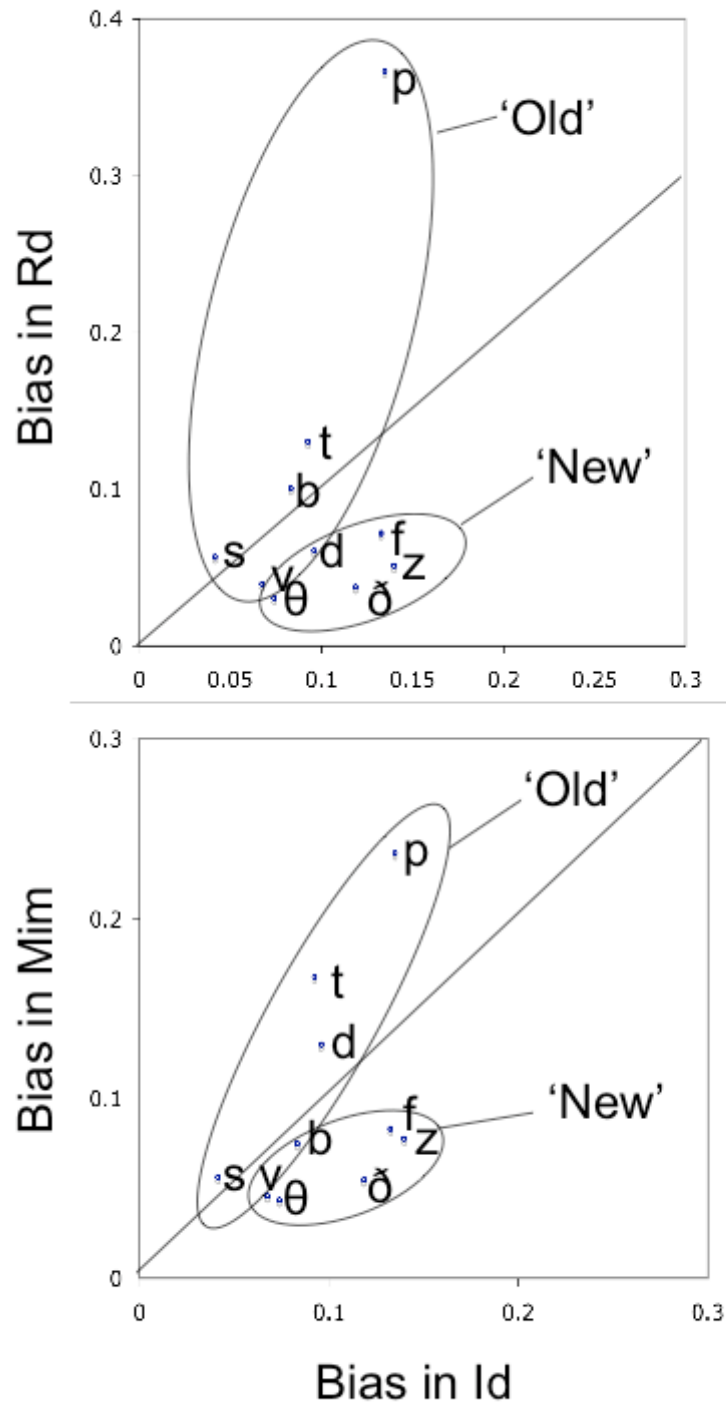


Figure 3. Bias for each segment found in Reading (top panel) and Mimicry (bottom panel) plotted against bias in Identification.

To get a feel for how these parameters play out in specific combinations of sounds from previously noted substitution errors in Korean, Tables 1 and 2 give portions of the larger confusion matrix. Table 1 gives matrices for labial stops and fricatives, collapsed across voicing. The reason for collapsing across voicing here is that, in the data analyzed here, voicing interacts with the manner errors highlighted in the table, so that errors in identification from stop to fricative tend to create voiceless outputs, /p/ → /f/, while errors from fricative to stop tend to yield voiced responses /f/ → /b/. Thus if we compare confusions of /p/ and /f/ only, this will not account for the fact that many errors are occurring with /f/, but they do not yield /p/ as a response. Table 1, thus, focuses on the overall pattern with manner errors, such as the ‘fork’ – ‘pork’ confusion noted above.

Table 1. Proportional confusion matrices for labial manner contrasts for three tasks. Rows tabulate input, columns tabulate percentage of identification of those inputs.

Identification	/f/ & /v/	/p/ & /b/
/f/ & /v/	70.2	12.6
/p/ & /b/	15.3	83.3
Reading		
/f/ & /v/	68.8	26.4
/p/ & /b/	4.8	92.1
Mimicry		
/f/ & /v/	44.1	16.0
/p/ & /b/	11.3	69.9

Table 1 shows a bi-directional error pattern in the identification task, with roughly the same proportion of fricatives being called stops as stops being called fricatives. In the production tasks, however, there are more instances of the fricative-to-stop errors than the other way around. In the reading task, the stop-to-fricative error rate is exceptionally low. (A majority of the additional errors for the mimicry production of /f/ & /v/ are confusions with the dental non-sibilants, so they would count as new-to-new errors, increasing the 44% correct responses. Similarly a majority of the additional /p/ & /b/ confusions were with /t/ & /d/.)

Table 2 presents confusions for the voiced coronal obstruents, and shows an even more extreme case of asymmetrical errors. The rate of stop-to-fricative errors is almost double that of fricative-to-stop errors in identification. The production tasks are strikingly in the opposite direction, with the /ð/-to-/d/ errors actually being more common than the correct transmission of the /ð/ fricatives. We also note, here, some very low accuracy rates in the production data (Rd and Mim), since errors with both /ð/ and /d/ are with a number of different segments. Summing up all of these errors, we find the same sorts of asymmetries. For example, in the Mim data /ð/ is called a stop 39.3% of the time, while /d/ is called a fricative only 16.8% of the time.

Table 2. Proportional confusion matrices for manner contrasts with voiced coronal obstruents for three tasks. Rows tabulate input, columns tabulate proportion of identification of those inputs.

Identification	/ð/	/d/
/ð/	63.8	20.0
/d/	36.7	62.0
Reading		
/ð/	25.7	34.7
/d/	7.1	49.5
Mimicry		
/ð/	21.5	29.0
/d/	11.0	66.6

We can also take a broader view of the database, and group all of the stops as ‘old’ segments, and all of the non-sibilant fricatives as ‘new’ segments, yielding the proportional matrix in Table 3. (Sibilants are not included here.) The same general pattern is found. There is a slightly larger proportion of stop-to-fricative errors than fricative-to-stop errors in identification. The pattern in both of the production tasks is strongly in the opposite direction. In mimicry, there the fricative-to-stop errors are double the stop-to-fricative errors; in reading the fricative-to-stop errors are more than quadruple the stop-to-fricative errors. (Note that the numbers in each row do not add up to 100, since there were a number of non-responses in the database.)

Table 3. Proportional confusion matrices for stops and non-sibilant fricatives for three tasks. Rows tabulate input, columns tabulate proportion of identification of those inputs.

Identification	Fricatives (‘new’)	Stops (‘old’)
‘new’	83.8	15.6
‘old’	18.2	81.5
Reading		
‘new’	59.5	31.7
‘old’	6.2	91.7
Mimicry		
‘new’	60.7	26.2
‘old’	12.1	82.1

4. Discussion and Conclusion

The current analysis of a corpus of perception and production data by Korean EFL learners reveals systematic ‘old category’ biases in production. As would be expected on the basis of L2 acquisition research from the 1950s onward (e.g., Lado, 1957), this data reveals systematic effects which would appear as substitution of an L2 category that maps onto an L1 category for an L2 category that poorly maps onto any L1 categories. Comparisons between the bias parameters plotted in Figure 3 with the mapping data in Park & de Jong (2008) suggest a systematic relationship between mapping and production data. Park & de Jong (2008) note that mapping goodness among these obstruents is not simply a binary difference between ‘old’ and ‘new’, but clearly has gradient structure. In these obstruents, the best fits are found with voiceless stops. The voiced stops, while being better than the non-sibilant fricatives, are clearly not as good a fit to the Korean stop categories as are the voiceless stops. Similarly, there is a bias in production toward stops, which is higher than that for non-sibilant fricatives. In addition, this difference between goodness measures for voiced and voiceless stops also shows up in the bias parameters (Figure 3), where the bias in production toward voiceless stops is higher than that for voiced stops.

One cause for this old category bias we also considered is that it reflects perceptual bias in the native listeners who judged the productions. Thus, when productions are ambiguous, frequency biases in the listeners begin to become apparent. Examining the frequency counts used in Silbert & de Jong (2007) might support this explanation. For example, if we examine the number of words with each segment in onset position, stop frequency is generally higher than fricative frequency. However, there are some parts of the overall patterns that do not match. For example, /s/ has the largest number of words in which it appears in onset position. In addition, the patterns for mimicry are somewhat different than those in reading; these differences cannot be due to listener bias, since the listeners were the same for the two corpora.

Another explanation for this old category bias in L2 mimicry and production tasks would lie in a motor component in the acquisition process. While perceptual tasks, such as identification, do not require any more difficult motor skill to respond with a ‘new’ than an ‘old’ category response, obtaining a ‘new’ category response from the English-speaking judges does require a novel motor skill. Thus, less accurate production regimes for the novel segments could produce more errors, reducing the overall number of ‘new’ category responses, and lowering the bias parameter for the ‘new’ segments. It is further possible that attraction of the productions of a novel L2 sound toward an L2 sound that is similar to one of the L1 sounds could be due to a merging of the motor skills for the ‘new’ and ‘old’ segment. This is the sort of explanation that has been suggested in various L2 production studies, such as Sancier & Fowler (1997). The motor skills for the two segments somehow share a component, and this component is linked to the experience of the learner with L1 segments. Since this component arises in connection with producing the L1 segment, the productions of the ‘new’ L2 segment tend to be similar to those of the ‘old’ L2 segment. It seems that motor learning has a tendency to conserve the substance of L1 phonological categories.

Turning to the identification patterns, one might conclude simply by explaining the tendency toward novel category biases here as being just the absence of old category preservation in the motor system. That is, since there are no systematic biases in identification, the biases that are relatively higher than those in production are simply due to the absence of motor biases. In fact, if we compare bias parameters for stops and their matched fricatives in Figure 3, it is apparent that two of the fricatives have lower bias parameters, and three have higher ones.

What this does not explain, however, is the oddity of the overall corpus in that it does not exhibit correlations between bias parameters and segment frequency. Specifically, it is not clear why it is that /f/ and /ð/ exhibit such large biases. It is unlikely that this is a simple frequency effect, since the frequency of occurrence for these segments, depending on the method of assessing frequency, tends to be very low. It is true that /ð/ is very high in token frequency, based on the corpora used in Silbert & de Jong (2007), if we only consider onset position. This is undoubtedly due to its predominant occurrence in function words. However, the frequency of /f/, assessed the same way, is less than /b,s,t/, all of which do not exhibit the same sorts of bias.

We suspect that the reason for biases toward /f/ and /ð/ have to do with the prototypicality of these segments as being decidedly ‘English’. That is, during the course of instruction, the presence of these non-Korean segments is highlighted, and they become strongly associated with the language. Since the language being listened to is English, the salience of /f/ and /ð/ as response options is constantly being strengthened throughout the listening task.

What this highlights is that what has been identified as frequency biasing may have many higher-level factors. That is, the frequency of a token is critically contingent on its frequency in a particular context. In the current case, the context is the new language, English, and hence it is not clear how experience with the L1 is to get factored into experience with the L2. That is, it is not clear how frequency biasing in the L1 plays out in their assessment of segments in the L2. A more targeted analysis of relative frequencies in both languages would illuminate this point.

The current analyses also highlight the fact that linguistic tasks are always happening in the context of tasks that we often would not classify as indicating part of the linguistic system. Thus, in the current instance we focus on identification of production events as instances of linguistic categories. However, while this is occurring, the listener is also cognizant of the fact that these production events are also identifiable (to some extent) as instances of the larger, social category of ‘English’. It is this intersection of the two classification systems, the segmental system within the language, and the social classification of the language as Korean or English, that is playing out in the listeners’ performance on identification and production tasks. In production, there may be an implicit goal of trying to sound English while getting the segmental content right. In perception, there may be an implicit estimate of the likelihood of a segmental category in the context of a language that seems to be populated with lots of specific novel

categories. While the larger novel language category is still fairly poorly separated from the preexisting language, instances of novel categories will be particularly salient markers of the novel language, and hence the novel language will tend to encourage perceptual responses of these marker segments. Either of these scenarios would give rise to a novel category bias effect. To the extent that the context of the novel language is well-formed and distinguished from the original language, this novel category effect would tend to disappear.

The results of the current analyses suggest that, at least with EFL learners with relatively little exposure to native spoken English, there is little in the way of novel production effects, but some suggestion of novel category perceptual effects. Future work should examine potential task effects that highlight the social classification of the language itself and how these impact identification behavior, in addition to the obvious extension of examining the effect of different kinds of L2 experience.

Acknowledgments

Work supported by the NSF (grant #BCS-04406540). We thank Mi-Hee Cho for her help in collecting the consonant data and Kyoko Nagao for her work on the design and processing of the identification data. We also thank Kenji Yoshida and Volya Kapatsinski for their helpful comments on an earlier version of this ms. Many of the analyses reported here were also presented at the Acoustical Society of America, abstract published as de Jong, K.J., Y.C. Hao, H. Park, & N.H. Silbert (2008). "The '[+English]' Effect in English Acquisition: Novel Category Biases in Second Language Production and Perception." *Journal of the Acoustical Society of America*, **123**: 3332, 2pSCc53.

References

- Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *Journal of the Acoustical Society of America*, 116 (6), 3668-3678.
- de Jong, K.J., B.J. Lim, and K. Nagao (2004). The perception of syllable affiliation of singleton stops in repetitive speech." *Language and Speech*, **47**: 241 - 266.
- Flege, J. (1987) The production of "new" and "similar" phones in a foreign language. *Journal of Phonetics*, 15, 47-65
- Kucera, Henry, & W. Nelson Francis (1967). *Computational Analysis of Present-Day American English*.
- Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, Mich.: The University of Michigan Press.
- Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology* (pp. 103-189). New York: John Wiley & Sons.
- Luce, P.A. (1986). Neighborhoods of words in the mental lexicon. *Research on Speech Perception Progress Report*, 14.

- Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusion among some English consonants. *Journal of the Acoustical Society of America*, 27 (2), 338-352.
- Nagao, K., B.J. Lim, and K.J. de Jong (2003). Perceptual acquisitions of non-native syllable structures by native listeners of Japanese. In M.J. Sole, D. Recasens, & J. Romero (eds.), *ICPhS*, pp. 1565-1568.
- Nusbaum, H. C., D.B. Pisoni, & C.K. Davis (1984). Sizing up the Hoosier Mental Lexicon: Measuring the familiarity of 20,000 words. *Research on Speech Perception Progress Report*, 10: 357-276.
- Park, H. (2007). Varied adaptation patterns of English stops and fricatives in Korean loanwords: the influence of the P-map. *IULC Working Papers On-line*, 07:01.
- Park, H., and K.J. de Jong (2006). Naïve Koreans' perception of voicing in VC position: Prosodic restructuring effects on consonant identification. *IULC Working Papers On-line*, 06:04.
- Park, H., and K.J. de Jong (2008). Perceptual category mapping between English and Korean prevocalic obstruents: Evidence from mapping effects in second language identification skills. *Journal of Phonetics*, 36: 704 - 723.
- Sancier, M.L., & C.A. Fowler (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, 25: 421 – 436.
- Silbert, N.H., and K. J. de Jong (2007). Response bias, type and token frequency, and prosodic context in segment identification. *Journal of the Acoustical Society of America*, 122: 3019, 3aSCa15.

Appendix. Instructions for the Perceptual Experiment

You will hear a series of items spoken by native speakers of English. For each item, you will hear a number followed by a one or two syllable nonsense word. For each nonsense word, please identify the consonant you hear by circling the appropriate symbol. If you do not find a symbol for the consonant that you hear, you may write the appropriate English consonant in the space marked other (). The symbols that you may choose from are:

symbol	description	symbol	description
p	as in the words <u>p</u> it, <u>app</u> le, and <u>stop</u>	b	as in the words <u>b</u> ad, <u>tab</u> le, and <u>rob</u>
t	as in the words <u>t</u> en, <u>beaut</u> y, and <u>cat</u>	d	as in the words <u>d</u> oor, <u>bo</u> dy, and <u>mad</u>
f	as in the words <u>f</u> an, <u>beautif</u> ul, and <u>half</u>	v	as in the words <u>v</u> an, <u>cover</u> , and <u>save</u>
θ	as in the words <u>th</u> ink, <u>math</u> , and <u>thank</u> you	ð	as in the words <u>th</u> ey, <u>broth</u> er, and <u>this</u>
s	as in the words <u>s</u> alt, <u>list</u> , and <u>pass</u>	z	as in the words <u>z</u> ebra, <u>amaz</u> ing, and <u>size</u>
r	as in the words <u>r</u> ock, <u>hear</u> ing, and <u>cover</u>	l	as in the words <u>l</u> ight, <u>feel</u> ing, and <u>ball</u>
w	as in the words <u>w</u> ood, <u>tow</u> el, and <u>cow</u>	y	as in the words <u>y</u> es, <u>lawy</u> er, and <u>toy</u>
h	as in the words <u>h</u> appy, <u>ah</u> ead, and <u>hand</u>		

After identifying the **consonant** that you heard, indicate how certain you are that you have chosen the appropriate symbol by circling a number to the right of the **consonant** symbols. The number 1 indicates that you are **not confident** that you have chosen the appropriate symbol for the **consonant** that you heard (you are just guessing). The number 7 indicates that you are **very confident** that you have chosen the appropriate symbol for the **consonant** that you heard.

Keyword tell dog thin that fall vase sit zip pin ball rain law hall wood yes - confidence +

1. t d θ ð f v s z p b r l h w y other () 1 2 3 4 5 6 7