

Language and Speech

<http://las.sagepub.com/>

Learning Foreign Vowels

John Kingston

Language and Speech 2003 46: 295

DOI: 10.1177/00238309030460020201

The online version of this article can be found at:

<http://las.sagepub.com/content/46/2-3/295>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Language and Speech* can be found at:

Email Alerts: <http://las.sagepub.com/cgi/alerts>

Subscriptions: <http://las.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations: <http://las.sagepub.com/content/46/2-3/295.refs.html>

>> [Version of Record](#) - Jun 1, 2003

[What is This?](#)

Learning Foreign Vowels*

John Kingston

University of Massachusetts

Key words

exemplars

features

German

learning

vowels

Abstract

Two hypotheses have recently been put forward to account for listeners' ability to distinguish and learn contrasts between speech sounds in foreign languages. First, Best's Perceptual Assimilation Model and Flege's Speech Learning Model both predict that the ease with which a listener can tell one non-native phoneme from another varies directly with the extent to which these sounds assimilate to different native phonemes (Best, 1994; also Best, McRoberts, & Goodell, 2001; Flege, 1991). Second, Logan, Lively, & Pisoni (1991) have argued that training listeners to identify non-native phonemes teaches them sets of exemplars rather than more abstract distinctive feature values. I report here the results of three sets of experiments designed to test these hypotheses, in which American English listeners were trained to categorize German nonlow vowels. The first set of experiments show that some instances of the same contrast between German vowels are more easily discriminated than others, a result incompatible with the predictions of either Best's or Flege's models, but compatible with the alternative category recognition interpretation. The second set of experiments reveals effects of contextual and speaker variation on listeners' ability to learn [tense] but not [high] contrasts between foreign vowels, and are thus at least partly compatible with an exemplar model of foreign category learning (Pisoni, Lively, & Logan, 1994; also Nosofsky, 1986). The third set of experiments compares the predictions of Nosofsky's (1986) selective attention exemplar

* *Acknowledgements:* This work was carried out together with Christine Bartels and Elliott Moreton, and with the advice and assistance of José Benkí, Deanna Moore, Elliott Moreton, Jeremy Rice, and Rachel Thorburn, all at that time of the Linguistics Department, University of Massachusetts, Amherst, and of Neil A. Macmillan, of the Psychology Department, Brooklyn College, CUNY. Parts of it have been presented at the meetings of the Acoustical Society of America, the International Conference on Spoken Language Processing, the fifth Laboratory Phonology Conference, and in invited presentations at the University of Buffalo, Universität Konstanz, Universität Kiel, the Max Planck Institute für Psycholinguistik, Nijmegen, and the First International Conference on Speech Sciences, Korean Association of Speech Sciences, Seoul, Korea. I thank the audiences at all these presentations for their very helpful comments and criticisms. This paper arises out of my presentation at the Workshop on Early Phonological Acquisition held at Carry-le-Rouet in October 2001. I thank the organizer of that workshop, Sharon Peperkamp, for inviting me both to present this work there and in this paper. The work was supported by grant R29 DC01708 from the National Institute of Deafness and Communicative Disorders, National Institutes of Health, which is gratefully acknowledged.

Address for correspondence: Linguistics Department, University of Massachusetts, South College 226, 150 Hicks Way, Amherst, MA 01003-9274. Phone: 1-413-545-6833; fax: 1-413-545-2792; e-mail: <jkingston@linguist.umass.edu>.

model of category learning with those of a feature learning model in tests of listeners' learning the natural classes to which the German vowels belong. The results are mixed: listeners learned the features that define the natural classes of [\pm high] and [\pm back] vowels, but could have learned either the feature that defines the natural classes of [\pm tense] vowels or sets of [\pm tense] exemplars. Natural classes defined by abstract distinctive feature values are thus learnable, even if their membership is phonetically polymorphous.

1 Introduction

Sounds' pronunciations in adults' native language and the differences between those sounds often interfere with recognizing a foreign speech sound or distinguishing one foreign speech sound from another. The occurrence, nature, and extent of such interference depend on how the foreign sounds are accommodated by native pronunciations and differences. Three general accounts of this accommodation have been influential: Best's Perceptual Assimilation Model (Best, 1994; Best et al., 2001), Flege's Speech Learning Model (Flege, 1991), and Kuhl's Natural Language Magnet model (Kuhl, 1991, 1993, 2000; Kuhl, Williams, Lacerda, Stevens, & Lindblom, 1992).

1.1

Models of assimilation

As its name indicates, the Perceptual Assimilation Model (PAM) treats accommodation as assimilation of foreign sounds to native ones. Best taxonomizes assimilations and uses that taxonomy to predict how well listeners will discriminate different foreign sounds from one another. The primary distinction in the taxonomy is between assimilation of two foreign sounds to two versus just one native sound, "two category" (TC) versus "single category" (SC) assimilations, respectively. PAM predicts that listeners will discriminate the members of TC assimilations far better than SC assimilations. SC assimilations are further taxonomized into those in which both foreign sounds assimilate equally to the single native category versus those in which one assimilates far more than the other. In the latter case, the two foreign sounds differ in "category goodness" (CG) with respect to the native category, and they are predicted to be discriminable to the extent that they do. The members of such CG assimilations are still less discriminable than the members of TC assimilations because they both assimilate to just one native category. PAM thus predicts that listeners' success in distinguishing different foreign sounds will be ranked: TC > CG > SC, with CG performance varying between TC and SC performance depending on whether the CG differences between the foreign sounds are larger or smaller.¹

¹ Best's taxonomy includes two other types: unassimilable and nonassimilable contrasts. Unassimilable contrasts are those whose members do not assimilate to any native categories, while the members of nonassimilable contrasts not only do not assimilate to any native categories but are also not even treated as speech sounds. The foreign sounds of interest in this paper are naturally produced vowels, presented in word contexts, and none of the listeners in these experiments thought they were anything other than speech. So it is safe to conclude that they were not members of nonassimilable contrasts. The specific results of the experiments show that they also were not members of unassimilable contrasts, so these two types will not be discussed further here.

Best et al. (2001) report that American English listeners differ in their ability to discriminate three pairs of different sounds in Zulu in just this way: they do best with the TC difference between the lateral fricatives /ɬ:ɣ/ because English distinguishes four pairs of fricatives for voicing, somewhat less well with the CG difference between aspirated /k^h/ and ejective /k'/ because both assimilate to English /k/ but /k^h/ does so much more than /k'/, and finally they do quite poorly with the SC difference between imploded /ɓ/ and unimploded /b/ because both assimilate equally well to English /b/, which itself varies between more and less imploded pronunciations.²

Up to this point, I have been purposefully vague in how I have referred to either the foreign or the native sounds or the differences between them, in particular, by using the words 'sound' and 'difference'. However, if PAM is to predict how any pair³ of different sounds from any foreign language will assimilate to the sounds of any native language, it must instead refer specifically to the phonemes and the phonological contrasts between them in the foreign and native languages. This reference to phonological categories was implicit in the account of Best et al.'s (2001) results in the preceding paragraph. For example, the [voice] contrast between Zulu /ɬ/ and /ɣ/ is a TC assimilation because English contrasts four pairs of fricatives, as well as three pairs of stops and one pair of affricates for [voice]. Generally, the foreign language's phonological contrasts can be compared with the native language's, and contrasts found in both should be TC assimilations while contrasts found in the foreign but not the native language should be SC assimilations instead.

The comparison cannot be only between the two language's phonologies, however, as that would predict no CG differences between SC assimilations. Best et al.'s finding that English listeners discriminate between Zulu /k^h/ and /k'/ far better than between Zulu /ɓ/ and /b/ depends on such phonetic differences. On the one hand, [k^h] is the typical allophone of English /k/ at the beginnings of words and stressed syllables, but [k'] does not occur in those contexts, although it may be produced occasionally by some speakers when they release /k/ at the ends of syllables. [ɓ] and [b], on the other hand, are free variants of /b/ at the beginnings of syllables, where producing voicing during the stop closure requires some degree of implosion, and speakers differ from token to token in how much they implode. English's lack of a [constricted glottis] contrast predicts that both these contrasts will be SC assimilations for English listeners, and that they will discriminate /k^h/ from /k'/ and /ɓ/ from /b/

² Best et al.'s acoustic description of the Zulu phonemes transcribed here as [ɓ] and [b] suggests that they might instead be prevoiced and voiceless unaspirated stops, that is, [b] and [p]. This alternative is acknowledged in their contribution to this volume (Best & McRoberts 2003). If this is correct, then these stops would indeed be hard to distinguish by English speakers as [b] and [p] are allophones of a single phoneme in this language, /b/.

³ Foreign and native contrasts are assumed to be binary by Best and in all the discussion in this paper. Nothing in her work or this depends on this assumption, and both her proposals and mine extend without change to contrasts with more values than two. The crucial distinction in her taxonomy is whether contrasting foreign sounds assimilate to different native sounds or the same native sound, and not whether assimilation is two-to-two versus two-to-one or many-to-many versus many-to-one.

less well than the /t/ and /k/ TC assimilation. However, the much larger CG difference between /k^h/ and /k'/ than between /b/ and /b/ predicts, correctly, that they will discriminate the two velar stops much better than the two bilabial ones.

These effects of CG differences should not, however, be limited to these two pairs of sounds. Zulu also contrasts labial and coronal voiceless stops for [constricted glottis], that is, /p^h/ versus /p'/ and /t^h/ versus /t'/, and they are predicted to assimilate like /k^h/ versus /k'/, that is, English listeners should be able to discriminate the members of all these pairs equally well. Unfortunately, the language does not contrast voiced stops at any other place of articulation than [labial] for this feature, but did it do so, English listeners would be also be predicted to discriminate them as poorly as they discriminate /b/ from /b/. These predictions follow from the fact that allophonic variation is systematic: one allophone may be typical in particular contexts, as aspirated pronunciations of [-voice] stops are at the beginnings of words and stressed syllables, or allophones may vary freely, as more or less imploded pronunciations of [+voice] stops do at the beginnings of syllables. Because both conditioned and free variation are systematic, the pronunciations of the foreign contrasts presented to listeners can be compared with the pronunciations of native contrasts, and predictions can be made from those comparisons about how well members of SC assimilations will be discriminated. Specifically, the CG differences between one instance of a foreign contrast should be the same as those for another instance. Indeed, one instance of a foreign contrast should generally assimilate in the same way as all other instances. This prediction of PAM is tested in the first set of experiments reported in this paper.

This account of PAM's predictions differs in two ways from that offered by Best (1994; Best et al., 2001) or by others (e.g., Polka, 1991, 1992) who have tested PAM's predictions. First, Best does not distinguish between phonological contrasts and allophonic variation because she represents differences between utterances in terms of the concrete gestures of articulatory phonology (Browman & Goldstein, 1986, 1989, 1990a,b,c 1992). In articulatory phonology, abstract contrasts for distinctive features such as [voice] aren't translated into concrete articulatory gestures such as glottal spreading but are instead represented directly in terms of those gestures. Moreover, systematic allophonic variation is represented by the same kinds of gestural differences as phonological contrasts. The /t:k/ contrast on this account would be represented as a choice between two gestures, the /k^h:k'/ contrast as the presence of a gesture versus one that does not occur in the English pronunciation of a voiceless velar stop, and the /b:b/ contrast as a difference in the size of a gesture. But for the purposes of predicting how a foreign contrast will assimilate, this difference between the two accounts is irrelevant. Whether the accounts are partly abstract and partly concrete or wholly concrete, their predictions follow from the systematicity of the correspondences between foreign and native sounds.

The second difference is more practical. Best and others do not predict how foreign contrasts assimilate in the way that I have sketched here but instead by whether listeners give the same or different labels to the foreign sounds and if they give them different labels, how consistently they do so. Listeners either name the native sound closest to each foreign sound or identify the foreign sound with a native sound in a

key word. If listeners consistently label the members of the foreign contrast differently, then it is treated as a TC assimilation, otherwise as an SC assimilation. If they sometimes label the members differently, then it is an SC assimilation with CG differences. There are three objections to these procedures. First, they are post hoc: no predictions can be made beforehand about how listeners will behave by comparing the two languages' contrasts and the systematic variation in their pronunciations. Second, these labeling tasks are unnecessary if the two languages' contrasts and their pronunciations are compared. Those comparisons, moreover, make much more general predictions than labeling does. Third, there's no reason to be confident that the native label a listener consciously assigns to a foreign sound will predict how they will identify that sound or discriminate it from a minimally contrasting sound in the foreign language. Therefore, the predictions of PAM tested by the first set of experiments reported in this paper are obtained from systematic comparison of the phonologies of German and English vowels.

The systematic similarities and differences between the actual pronunciations of foreign and native sounds is the foundation of Flege's Speech Learning Model (SLM). SLM predicts that a phonetic difference that distinguishes contrasting foreign sounds but does not also distinguish contrasting native sounds will be poorly detected and reproduced. SLM makes the same predictions as PAM about the three Zulu contrasts, and does so, moreover, in equally concrete terms. Flege also shows that listeners are more able to detect and reproduce a foreign phonetic difference when they encounter it early in life and when their encounter with that difference is lengthy. He also documents changes in the detection and reproduction of native differences when the encounter with the foreign language is early and lengthy. Although the experiments reported here exposed listeners to the foreign sounds for much longer than do the typical tests of perceptual assimilation and even sought to teach the foreign contrasts to the listeners, they still first heard these sounds as mature adults and only heard them for 90 mins a day for at most 12 days. These listeners encounter with the foreign sounds is too late and too brief to qualitatively change which phonetic they can detect—all changes observed are at best modest quantitative improvements. Moreover, listeners' pronunciations of their native or foreign contrasts were not measured in any of these experiments. Nonetheless, because the members of one instance of a foreign contrast differ phonetically from one another in the same ways as other instances, SLM like PAM predicts that the phonetic differences between all instances should be equally easy or hard to detect and reproduce.

The predictions of the third perceptual assimilation model, Kuhl's Natural Language Magnet model (NLM), are based on representations that are at least as concrete as those in PAM or SLM, if not more so. Kuhl proposes that native language categories are prototypes. Each prototype occupies a specific location in a space whose dimensions are the phonetic properties that define that class of categories, as, for example, the vowel space is defined by vowels' formant frequencies. Tokens near a prototype are drawn perceptually to it; this is why Kuhl also refers to the prototypes as "magnets." Foreign as well as native sounds are drawn to these prototypes as a function of their distance from them in the phonetic space. More distant foreign sounds either assimilate to another prototype if they're closer to it, or do not

assimilate if there's no nearby prototype. That is, assimilations could be either TC or SC with CG differences. Two foreign sounds that are the same distance in the phonetic space from a native prototype are predicted to assimilate to it equally so long as one is not closer to another prototype. In this case the assimilation is SC without CG differences.

The essential difference between NLM and either PAM or SLM is that the prototypes' locations are fixed in a phonetic space. Neither PAM nor SLM relies on a spatial representation; instead, sounds differ from one another in one or more constituent gestures or phonetic properties. The prototypes representing different instances of what would be the same phonological contrast⁴ might therefore not be equally far apart in the phonetic space. NLM predicts better discriminability for foreign sounds that assimilate to more well separated prototypes than to less well separated ones. For example, the prototypes of high rounded vowels contrasting for backness, /u/ versus /y/, might be farther apart than those of mid rounded vowels, /o/ versus /ø/, because vowels are more dispersed higher in the vowel space. Foreign vowels that assimilate to the high vowels would therefore be predicted to be more discriminable than those that assimilate to mid vowels. Neither PAM nor SLM provides in any principled way for gestures or phonetic properties to differ more for one instance of a contrast than another, so neither predicts discriminability differences between different instances of the same foreign contrast.

1.2

Models of learning by infants and adults

What have just been described as models of the perceptual assimilation of foreign categories to native ones are also all explicit models of how these categories are learned by infants during the first year of life, and how that learning changes their response to contrasts that will not occur in the ambient language.

Before six to nine months of age, infants can discriminate any sounds that contrast in any language. The age they lose this ability depends on the contrast: detection of differences between foreign versus ambient contrasts fails earlier for vowels than consonants. By the end of the first year, however, they apparently can no longer discriminate most sounds that do not contrast in the ambient language—the exception is members of nonassimilable contrasts such as Zulu clicks contrasting for place of articulation (Aslin, Jusczyk, & Pisoni, 1998; Best, McRoberts, LaFleur, & Silver-Isenstadt, 1995; Polka & Werker, 1994; Werker & Lalonde, 1989; Werker & Tees, 1984). This developmental change is accounted for in all three assimilation models as a side effect of the infant having learned the categories of the ambient language during this six to nine month period. By attracting both the ambient and foreign sounds the infant hears, these categories learned by 12 months deafen it to

⁴ It is actually quite unclear how different contrasts can be identified as instances of the same contrast in a model where categories are prototypes distributed in a phonetic space. The most one can do is identify pairs of prototypes that differ along parallel axes in the space, for example, the prototypes of the high rounded vowels /u/ and /y/ might differ along the same axis as those of the mid rounded vowels /o/ and /ø/.

differences it could detect six months earlier when it had not yet learned any categories with such attractive power.

Of course, the infant is not literally deafened to these contrasts. I adopt this vivid term from Dupoux and his coworkers to refer to the developmental weakening in the behavioral response to differences within foreign than ambient contrasts. At 12 months of age, an infant is behaviorally deafened in this sense to foreign contrasts because it no longer responds differently when the stimulus changes from one member of a foreign contrast to the other. To be sure, at this age behavioral deafening is only temporary, as an infant can still for a time learn new categories if the ambient language changes. However, if the infant does not get sustained exposure to the foreign language early enough, apparently before 5–6 years of age, it is permanently deafened behaviorally to foreign contrasts not shared with the native language.

Monolinguals are deafened. For example, adult speakers of languages with fixed stress (French, Finnish, Hungarian) are significantly less able to detect a shift in stress position in a word than a change in one of its segments (Dupoux, Pallier, Sebastian, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Peperkamp & Dupoux, 2002). Even highly fluent bilinguals are deafened, too, if their exposure to the second language is too late. For example, some Spanish-dominant Spanish-Catalan bilinguals who did not learn Catalan before 5–6 years of age cannot discriminate Catalan contrasts not shared with Spanish, high mid versus low mid vowels, /e, o/ versus /ɛ, ɔ/, or voiced versus voiceless fricatives, /z/ versus /s/ (Pallier, Bosch, & Sebastián-Gallés, 1997). These sounds are in fact so thoroughly homophonous for the Spanish-dominant bilinguals that one member of a minimal pair for these Catalan contrasts produces repetition priming for the other; no repetition priming occurs within minimal pairs for Catalan-dominant bilinguals (Pallier et al., 2001). Guion (2003) reports similar effects of age of acquisition on Spanish vowel production by Ecuadorian Quichua-Spanish bilinguals. Bilinguals who learned Quichua and Spanish simultaneously pronounce the five Spanish vowels /i, e, a, o, u/ distinctly as well as distinguishing their pronunciations from those of the three Quichua vowels /ɪ, a, ʊ/. Bilinguals who learned Spanish early but not before they started school at 5–6 years pronounce the high Quichua vowels indistinguishably from either the high or mid Spanish vowels, although they do discriminate high from mid vowels when speaking Spanish. Finally, bilinguals who learned Spanish in puberty or later produce only three distinct vowels when speaking either language, with Quichua pronunciations.

Adults not only have difficulty learning foreign categories at all, but when they do learn foreign categories, what they learn still differs in generalizability from native categories, at least when the learning takes place over a relatively brief period of time in an experimental setting. Pisoni and his colleagues (Lively, Logan, & Pisoni, 1993; Lively, Logan, Pisoni, Yamada, R., Tohkura, & Yamada, T., 1994; Logan, Lively, & Pisoni, 1991; Pisoni, Lively, & Logan, 1994) have extensively studied the learning of the contrast between the American English liquids /l/ versus /ɹ/ by native speakers of Japanese. Their procedures are described in more detail below; for now, all that need be said is that the participants in their experiments were trained for multiple days with a large number of minimal pairs for this contrast in which the contrasting sounds occurred in a variety of positions, for example, ‘lock’ versus ‘rock’, ‘blade’

versus 'braid', 'mellow' versus 'marrow', 'steel' versus 'steer', and 'bold' versus 'board', spoken by a number of different speakers. Training with these multiple and varied instances of naturally produced minimal pairs improved the participants' performance on the training stimuli and even generalized to novel tokens, unlike training with a single synthetic minimal pair (cf. Strange & Dittman, 1984). However, participants consistently discriminated these two sounds better in some positions than others and when spoken by some speakers than others. They were also more successful at generalizing to novel tokens spoken by a familiar speaker than by a novel speaker. Finally, the position and speaker effects both persisted, being still detectable 3 and 6 months after the original training, even though both accuracy and generalization diminished with time. These findings support the interpretation that Japanese listeners learned the English categories as sets of exemplars of /l/ and /ɹ/ and did not abstract much, if at all, away from their experience of the training stimuli, even after considerable time had elapsed.

Pisoni and his colleagues argue that the benefits of their training procedures and these exemplar effects both follow from Nosofsky's attention-weighted exemplar model of category learning (Nosofsky, 1986). In that model, listeners learn categories as sets of exemplars. When a novel token is encountered, its similarity is calculated to all previously encountered exemplars of all competing categories, and it is assigned to the category with which it has the highest aggregate similarity. Moreover, irrelevant variation between exemplars, for example, that between word positions or speakers, helps the learner by showing which dimensions of difference between the stimuli are relevant to categorizing them and which are irrelevant. Perceptual distances stretch along the relevant dimensions and shrink along the irrelevant ones, and as a result, the relevant dimensions are eventually weighted more heavily in calculating aggregate similarity for any new token.

Distinguishing relevant from irrelevant dimensions can apparently be a lengthy process, for the Japanese listeners in these studies have clearly not finished sorting the relevant from the irrelevant dimensions, even though they received multiple days of training with a large variety of stimuli. Nor did the passage of considerable time make the irrelevant dimensions any less salient compared to the relevant dimensions. Evidently, these listeners acquired quite stable sets of exemplars of /l/ and /ɹ/ during training, to which they could still refer three and six months later, rather than abstracting the categories away from the instances that originally exemplified them. The second set of experiments reported in this paper test the generality of Pisoni and colleagues' findings in the learning of German vowels by American English listeners, by manipulating the extent and kind of irrelevant variation in the training stimuli.⁵

⁵ Although exemplar effects like these have been reliably established for native categories, too, in adults' performance in recall or matching tasks (Goldinger, 1996, 2000, Goldinger et al., 1991), they apparently do not affect category recognition, except in adverse listening conditions (Mullenix et al., 1989) or under time pressure (Mullenix & Pisoni, 1990). Exemplar effects are also much shorter lived for native categories than foreign ones, presumably because so many exemplars are experienced and new ones constantly refresh the sets stored in memory.

Exemplar models of category learning like Nosofsky's represent categories in spaces defined by the perceptual correlates of the physical dimensions along which they differ. Moreover, perceptual similarity is modeled as the distance in this space between exemplars, and distances stretch or shrink depending on whether a dimension is relevant to categorizing the stimuli. In all these respects, they resemble prototype models of category learning like Kuhl's. Indeed, the perceptual attraction of tokens by a nearby prototype in her model could readily be obtained in Nosofsky's model by stretching perceptual distances along relevant dimensions and shrinking them along irrelevant dimensions (cf. Guenther, 2000; Iverson & Kuhl, 2000; Lotto, 2000; Lotto, Kluender, & Holt, 1998).

Kuhl et al. (1992) showed that the locations of prototypes are determined by experience as early as six months of age. Infants of this age who were being raised in English or Swedish speaking households were habituated either to prototypical English /i/ or prototypical Swedish /y/, and then presented with vowels surrounding these prototypes in the dishabituation phase. English, of course, has no /y/, and Swedish /i/ is located differently in the vowel space from English /i/. Results depended on the infants' linguistic experience: infants from English speaking households discriminated vowels surrounding /i/ much less well than those surrounding /y/, while infants from Swedish speaking households discriminated vowels surrounding /y/ much less well than those surrounding /i/.

More recently, results reported by Polka and Bohn (1996; see also Polka & Werker, 1994) undermine the argument that experience determines a language's prototypes. They tested 6–8 and 10–12 month old children being raised in English and German speaking households on the English contrast /ε/ versus /æ/ and the German vowel contrast /y/ versus /u/. These vowels were pronounced by a single native speaker of each language in the context [d_t]. Infants were habituated to one vowel from a pair and then tested with the other. If the vowel they're habituated to is a prototype, they should discriminate tokens of the test vowel from it less well than if it is not. Because English has an /u/ but no /y/, while German has an /ε/ but no /æ/, NLM predicts that both /u/ and /y/ will act like prototypes for German infants but only /u/ will do so for English infants and that both /ε/ and /æ/ will act like prototypes for English infants but only /ε/ will do so for German infants. Moreover, the prototypes' effects should be stronger for the older than the younger infants. Polka and Bohn's findings were quite at odds with both these predictions: younger and older infants from both linguistic backgrounds discriminated the test vowels /y/ and /ε/ significantly less well from the habituation vowels /u/ and /æ/ than vice versa, indicating that /u/ and /æ/ are prototypes for all four groups of learners. These prototypes apparently aren't determined by the infants' different linguistic experience but instead apparently by these vowels' more peripheral positions in the vowel space.

Two substantial differences in procedure might underlie the discrepancies between Polka and Bohn's results and Kuhl et al.'s: (1) Polka and Bohn used multiple tokens of naturally produced vowels in a CVC syllable, Kuhl et al. used arrays of synthetic, isolated vowels equally spaced around the English and Swedish prototypes, and (2) Polka and Bohn tested infants' ability to discriminate two vowel categories

from one another, Kuhl et al. tested their ability to discriminate different variants of a single category. However, until the discrepancies are shown to follow from one or the other of these procedural differences, no conclusion can be safely drawn about whether experience determines prototypes' locations. Experience and peripherality are unfortunately also confounded in the results reported here in that the German vowels that American English listeners are likely to find most familiar are also more peripheral in the vowel space. Nonetheless, it is possible to test whether the less familiar and less peripheral German vowels assimilate to the more familiar, more peripheral ones by measuring response biases in the first set of experiments.

Of more interest, however, is whether the participants in these experiments merely learned sets of exemplars, like the Japanese listeners apparently did in Pisoni and colleagues' experiments, or if they also generalized away from the properties of the particular exemplars they were exposed to in training. Ashby and his colleagues have developed an alternative model of category learning in which categorization depends not on aggregate similarity to competing clumps of exemplars but on distance from a decision boundary separating the competing categories in the perceptual space (Ashby & Gott, 1988). Although this work has established that the geometry of this boundary depends on the distributions of exemplars of the competing categories, once the boundary is established it determines observers' categorization of novel stimuli independently of the exemplar distributions. This independence is the first step toward establishing representations of contrasting categories that are more general and abstract than the specific exemplars encountered during learning, that is, the first step toward distinctive features. The last set of experiments reported here test the competing predictions of a model in which the categories listeners learn are the natural classes defined by distinctive feature values rather than clumps of exemplars.

1.3

Prologue to the experiments

Category learning is the *primum mobile* of these experiments. They investigate how categories are learned, and what it is about these categories that gives them the power to attract other similar sounds. Ideally, one wants to simulate the infant's experience of spoken sounds experimentally, to the extent that it is practical to do so.

In these experiments, all listeners are adults, so the learners are not in the state of an infant who has yet to learn any categories, indeed they are in the state of permanent behavioral deafness to differences between foreign categories that a person suffers from after the age of five or six. However, the target stimuli, German nonlow vowels, were spoken by multiple speakers in multiple consonantal contexts, so they vary irrelevantly in the same way if not to the same extent as these vowels do in a German-learning infant's experience. Like these infants, the listeners in these experiments had to learn which acoustic properties were relevant for identifying and categorizing the vowels and which ones were not.

The stimuli in the first set of experiments were also chosen so that listeners heard two instances of each contrast, for example, they might hear two pairs of vowels that contrast for [round], /i/ versus /y/ and /e/ versus /ø/, and two pairs that contrast for [high], /i/ versus /e/ and /y/ versus /ø/. As with these four vowels, the

contrasts between the stimuli were orthogonal as well as multiple. Using more than one instance of a contrast was essential to testing PAM's or SLM's prediction that all instances of the same foreign contrast will assimilate equally well or poorly to native contrasts. Using multiple instances of orthogonal contrasts also simulates category learning in nature, where many of the categories the infant has to learn have more than one member, which contrast with one another in terms of orthogonal features, that in turn define other categories with multiple members.

The second set tests the generality of Logan et al.'s finding (1991; see also Lively et al., 1993, 1994; Pisoni et al., 1994) that Japanese listeners learn the American English /ɪ:ɪ/ contrast better with some speakers and contexts than others. These results suggest that listeners learn sets of exemplars and that they cannot entirely separate relevant dimensions for identifying the vowels from irrelevant ones.

The exemplar effects found in the second set of experiments motivate the third set, which directly compares Nosofsky's (1986) selective attention exemplar model of category learning with a model in which listeners learn the abstract features (or prototypes) that define natural classes in a language (Iverson & Kuhl, 1995; Jakobson, Fant, & Hallé, 1953; Kuhl, 1991; Kuhl et al., 1992; Trubetzkoy, 1939/1969; **but cf.** Bybee, 2001; Johnson, 1997; Maye, Werker, & Gerken, 2002; Pierrehumbert, 2002).

In the next three sections of this paper, I present the methods used and the results obtained in each set of experiments, followed by discussion of each set of results. What the package of results reveals about the perceptual nature of phonological categories is discussed at the end, and a recommendation is made about a new kind of research that needs to be done with infant listeners.

2 Testing perceptual assimilation

I argued in the introduction that PAM and SLM predict that each instance of the same foreign contrast should assimilate equally well or poorly. This prediction is tested in the experiments reported in this section by comparing American English listeners' ability to identify different pairs of German vowels that contrast in the same way. I refer to those categories as distinctive feature values, for example, [+high], but I could, like Best, also refer to them as gestural parameter values, that is, a close but not critical constriction by the tongue dorsum. Either way, the categories contain multiple members that contrast minimally with the members of another category, whether that other category is referred to as [-high] or as a not close constriction by the tongue dorsum. Furthermore, referring to the foreign categories in terms of distinctive feature or gestural parameter values permits me to predict very specifically which American English vowels the German vowels will assimilate to and what type of assimilation will occur, TC, CG, or SC (see immediately below).

2.1

Methods

Introduction. I adapted a method introduced by Logan et al. (1991) to assess the learning of the American English /ɪ:ɪ/ contrast by Japanese listeners. The experiments all had three phases. First, "pretraining" assessed listeners' ability to identify

the individual stimuli, and to categorize them for their distinctive feature values. Pretraining was followed by multiple days of “training” on complete identification of the individual stimuli. Finally, in “post-training and generalization,” listeners were presented with new tokens of vowels they had been trained on, for complete identification and categorization by distinctive feature value.

Stimuli. The stimuli in the experiments were German words and pseudowords of the form [ʼCVCŋ], that is, a two-syllable utterance with primary stress on the first syllable and an unstressed syllabic nasal in the second syllable (that nasal is pronounced with the same place of articulation as the preceding C). This is a common word shape in German, occurring for example as the infinitive of verbs with monosyllabic stems. The stimuli were produced and presented in isolation. The vowel that listeners identified or categorized is represented by V in the first, stressed syllable. The C preceding this vowel could be any of /b,d,g,p,t,k/, and the C following it could be any of /p,t,k/. These stimuli were produced by five native German speakers, referred to henceforth as A–E. Three speakers were women (A, B, E), and two were men (C, D). All three women spoke a northern dialect of German (Hamburg or Kiel), as did one of the men (D); the other man (C) spoke a strongly northern-influenced southern dialect (Bavaria). One of the women was in her mid forties (B), the other women and the men were in their late twenties or early thirties. Recordings were made in an anechoic chamber onto cassette tape, digitized at 16kHz with appropriate anti-aliasing filtering, down-sampled to 10 kHz, and equalized for peak amplitude, which always occurred during the first, stressed vowel. When the stimuli were presented to listeners they were low-pass filtered with an 11-pole filter at 4.167kHz (energy was down at least 80 dB at the Nyquist frequency).

In all the experiments reported in this section, a subset of four vowels was drawn from the 12 nonlow German vowels shown in Table 1.

TABLE 1

The nonlow vowels of German, cross-classified by [high], [back], [round], and [tense]

<i>Non-low German vowels</i>	front [–back]				back [+back]	
	unrounded [–round]		rounded [+round]			
	<i>long</i> [+tense]	<i>short</i> [–tense]	<i>long</i> [+tense]	<i>short</i> [–tense]	<i>long</i> [+tense]	<i>short</i> [–tense]
high [+high]	i	ɪ	y	ʏ	u	ʊ
mid [–high]	e	ɛ	ø	œ	o	ɔ

Seven of these vowels, the unrounded front vowels and the rounded back vowels other than [ɔ], are phonologically identical and phonetically similar to American English vowels. There are only two striking phonetic differences in the pronunciation of these vowels between the two languages (cf. Strange & Bohn, 1998). The first is that the [+tense] German vowels aren't diphthongized in these speakers' north

German varieties, while their American English counterparts are, particularly mid /e,o/. Otherwise, the [+ tense] vowels last longer and have higher and more peripheral articulations in both languages (Fischer-Jørgensen, 1990; Hillenbrand, Getty, Clark, & Wheeler, 1995). Second, the mid lax back rounded vowel /ɔ/ in German is shorter and less peripheral than its tense counterpart /o/, like all the other lax vowels in this language, while the English vowel with this quality is long, indeed the longest vowel in a recent study of American English vowels (Hillenbrand et al., 1995). This vowel is also not distinct from [a] either in production or perception for a large number of American English speakers, most of whom have [a] everywhere. This vowel quality may also be foreign for these listeners. Means and standard deviations across consonantal contexts for the spectral properties and durations of the German vowels are given for each speaker in the *Appendix*. Table A3 shows that the speaker who originally spoke a southern rather than northern variety of German, C, does not pronounce the tense and lax high front vowels, /i/ versus /ɪ/ or /y/ versus /ɻ/, with different F1 values, although he does pronounce the tense and lax high back vowels and all the tense and lax mid vowels with different values for this formant. The tense and lax high front vowels still differ in duration for this speaker.

The other vowels which are novel or foreign to American English listeners are of course the front rounded vowels /y, ø, ɤ, œ/, but as shown in Table 1, the contrasts between them for [high] and [tense] are not. Because the contrasts among the front rounded vowels are familiar to American English listeners, they may find them relatively easy to discriminate from one another. What they may instead find difficult is distinguishing the front rounded vowels from either front unrounded or back rounded vowels. Front rounded vowels occur in no more 6% of the UPSID sample of 451 languages (Maddieson & Precoda, 1992), and they may be infrequent because violating the usual redundancy between the features [back] and [round] makes them hard to discriminate from vowels that do not violate this redundancy constraint. This violation also makes the front rounded vowels less peripheral in the vowel space. American English listeners may therefore find them hard to discriminate from one another, despite the familiarity of the contrasts between them.

The experiments differed in which four vowels were drawn from this set, but in all the experiments the four vowels could be cross-classified into a 2×2 subset by two distinctive features. The features held constant in the 2×2 subsets give the names to these experiments; Table 2 (overleaf) lists these names, the vowels used, and the features that cross-classify the vowels in each experiment.

One of the features was always [high], the other feature was [tense] in the Front Round experiment, [round] in the Tense and Lax Front experiments, and [back] in the Tense and Lax Round experiments. At least one pair of vowels was always front rounded, as I always wished to test listeners responses to genuinely foreign categories. At the same time, both features in all five experiments also distinguish many pairs of English vowels from one another, so all of these contrasts between the front rounded vowels themselves are potentially TC assimilations. However, the lack of front rounded vowels in English and the redundancy between [back] and [round] in this language probably make the [round] contrast in the Tense and Lax Front experiments and the [back] contrast in the Tense and Lax Round experiments CG

assimilations instead. PAM's and SLM's predictions are the same regardless of whether a contrast is a TC or CG assimilation: both instances of the same contrast in each stimulus set, for example, the two [back] contrasts in the Tense Round experiment, /u:y/ and /o:ø/, are predicted to assimilate equally well or poorly.

TABLE 2

2 × 2 vowel subsets used in the five experiments testing PAM and SLM, cross-classified by the two features defining the subset. Names of experiments refer to the features held constant within the subset

Front Round	+ tense	– tense	
	y	ɤ	+ high
	ø	œ	– high
Tense Front	– round	+ round	
	i	y	+ high
	e	ø	– high
Lax Front	– round	+ round	
	ɪ	ɤ	+ high
	ɛ	œ	– high
Tense Round	– back	+ back	
	y	u	+ high
	ø	o	– high
Lax Round	– back	+ back	
	ɤ	ʊ	+ high
	œ	ɔ	– high

Participants. Listeners were all native speakers of American English, recruited by advertisement from the undergraduate student body of the University of Massachusetts, Amherst. They were paid \$6–\$7 per hour for participating. No listeners reported any speech or hearing pathology. Listeners were carefully screened beforehand to ensure that they had never heard German except very occasionally in films or other media or during brief visits to German-speaking countries. Anyone was excluded who had studied German, lived for any period longer than two weeks in a German-speaking country, or had studied or been exposed for any length of time to any other language with front rounded vowels: Dutch, French, any Scandinavian language, any Chinese language, Korean, or Finnish. Listeners were screened by asking them what languages other than English they spoke, had studied, or had been exposed to for any length of time. Potential listeners were not explicitly asked about German as I wanted to keep the identity of the language they would be listening to secret from them until after the experiment was over. Many listeners correctly guessed that the language was German during the course of the experiment, but these guesses were not confirmed by the experimenters until the experiment was over.

Procedures. Listeners heard the stimuli binaurally over TDH-49 headphones at self-selected comfortable intensities in a sound-isolated quiet room, where they sat partly

separated by partitions from other listeners. Up to four listeners were run at a time. Once all the listeners in a group had responded to a stimulus, a light came on above the button they should have pressed in response to that stimulus. They were instructed that they would hear two-syllable words from a foreign language. They were to listen to the first vowel in these words and to identify it by pressing one of the buttons on the box, and they were to learn which button to press in response to each vowel by paying attention to the feedback provided by the lights that came on after they had responded. Each block of trials began by presenting example words containing each of the four vowels in a fixed order. These orientation trials were intended to teach the listeners how the vowels corresponded to response buttons. Otherwise, they were told that at first they would simply have to guess which button to press for a particular vowel, but that if they attended carefully to which feedback light came on after their response, they would eventually learn which vowel went with which button. They were also told that the words would be spoken by different people and that the consonants on either side of the vowel would differ, but that they were to pay attention only to the differences between vowels. The only other instruction they were given was that the words contained just four different vowels, which they were to identify individually. In this complete identification task, listeners responded by pressing one of four buttons on the box in front of them. Listeners had up to 2000 ms to respond. The feedback light came on for 500 ms after the slowest listener responded or 2000 ms, whichever was shortest, and it was in turn followed by 1000 ms before the next stimulus was presented. No response was recorded for listeners on a trial if they did not respond within 2000 ms, and all responses given less than 100 ms after the stimulus began were discarded.⁶

During pretraining and training in all but one subexperiment, listeners heard stimuli produced by speakers A, B, and C in a representative two-thirds of the consonantal contexts. The exception is that four of the listeners in the Front Round experiment were trained with speakers B, D, and E. In post-training and generalization, tokens produced by the other two speakers and the remaining third of the consonantal contexts were presented to test generalization of training. In some blocks of trials in post-training and generalization, new stimuli were mixed with the old ones presented during training, in others the new stimuli were presented alone.

⁶ In addition to this complete identification task, during pre- and post-training phases, listeners categorized the vowels into two sets of two vowels each according to their values for the distinctive features defining the 2×2 stimulus arrays. In categorization, listeners responded by pressing one of the two outermost buttons on the box. After responding, they also indicated how confident they were about their judgment, on a 1–4 scale. A brief triplet of tones was played after all listeners had responded to tell them to indicate their confidence by pressing one of the four buttons. The feedback light did not come on until after the confidence rating was entered or the allotted time had elapsed. The category to which each pair of vowels belonged was also taught entirely by providing the listeners with this feedback after their response, although they were told in this task that the two vowels were assigned to each response. 2000 ms were allotted for each response, and the feedback light was on for 500 ms and was followed by a 1000 ms pause before the next stimulus was presented. Categorization results aren't reported here, but the way in which categorization was run is described for each experiment.

Pretraining began by familiarizing the listeners with the identification task. Familiarization used pseudowords produced in isolation by a single male native speaker of American English (the author), whose vowels are more-or-less typical of the northern Midwestern cities — he lived in Detroit, Toronto, north of Pittsburgh, and Chicago up to the age of 22. The shape of these pseudowords, [‘CVCŋ], was very similar to that of the German stimuli. The stressed vowel was one of the tense nonlow vowels of American English, /i,e,u,o/. Listeners were presented with a single block of identification trials in which they had to respond differently to each of the four vowel categories and two blocks of categorization trials in which they had to sort the four vowels into the sets defined either by the feature [high], /i,u/ versus /e,o/, or the features [back] and [round], /i,e/ versus /u,o/. Both identification and categorization blocks began with 12 orientation trials in which the stimuli cycled through the four vowel categories or alternated between the two values of the relevant distinctive feature, followed by 48 randomized test trials.

Once listeners were familiarized with the tasks, German stimuli were used in the rest of pretraining, which took place on the same day. In pretraining and subsequent training, a block of identification trials consisted of 16 unscored orientation trials in which stimuli cycled in a fixed order between the four vowels of interest, followed by 72 randomized test trials in which each of the four vowel categories occurred equally often. The categorization blocks that the listeners performed during pre- and post-training also began with 16 unscored orientation trials in which stimuli alternated between the two values of the distinctive feature, followed by 72 randomized test trials with each vowel category presented equally often. In this task, listeners grouped vowel categories into pairs according to their values for one of the two distinctive features that defined the 2×2 array of vowels in that experiment. For both identification and categorization, two blocks of trials of these sizes were needed to present all the stimuli (3 speakers \times 12 C_C contexts \times 4 vowels = 144 trials). In new + old post-training complete identification and categorization, a block of trials began with 16 unscored orientation trials, followed by 90 randomized test trials. Four blocks of this length were needed to present all combinations of five speakers, 18 contexts, and four vowels. In new-alone post-training complete identification and categorization, a block began with 12 unscored orientation trials, followed by 48 randomized test trials, in which each of the four vowels was presented 12 times in representative new tokens differing in speaker and/or C_C context from the training tokens. The blocks of new-alone stimuli were presented twice, once at the beginning of post-training and generalization and once at the end.

The Front Round experiment was run first. Training in this experiment consisted only of complete identification, in which listeners identified all four vowels on every trial. Because listeners improved only modestly over the course of training in this experiment, in some blocks just two vowels were presented during training for the other four experiments. In these experiments, each training day consisted of two blocks of trials in which listeners had to identify all four vowels, and one block each for the six possible pairs of vowels. In two- as well as four-vowel blocks, the stimuli varied in speaker and/or context. For four listeners in the Front Round experiment and four listeners in the Lax Front experiment, the two-vowel task was yes-no, that is, the listeners heard a single stimulus per trial and had to identify its vowel as one

of two alternatives. This task was easy enough when just two stimuli had to be identified that a number of listeners were able to perform it perfectly. To avoid such ceiling effects, the more difficult same-different task was used in its place for the other four listeners in the Lax Front experiment and for all listeners in the Tense Front and the Tense and Lax Round experiments. Stimuli in both same and different trials always differed in speaker and/or context, so listeners had to judge whether the vowel was the same or different in distinct tokens. They were instructed to judge these stimuli solely on the identity of the vowel. The blocks of trials for both yes-no and same-different two-stimulus tasks also consisted of 12 unscored, ordered orientation trials and 72 randomized test trials.

To continue in the experiment, listeners had to get above 60% correct for identifying each vowel by the end of pretraining. Nearly all listeners reached this criterion, but in a few instances listeners did not manage to and were excused from further participation, and other listeners were recruited in their place. Subsequently, performance was assessed by calculating the perceptual distance measure d' using the constant ratio rule for all six pairs of vowels in the complete identification tasks (Macmillan & Creelman, 1991).

Blocks of trials lasted between 4–8 mins depending on how many stimuli they contained. Blocks were separated by brief breaks during which the responses were stored and the next block set up. Listeners also took a longer break midway through each day. Generally, only the two phases of pretraining could be run on the first day, and training did not begin until the second day. Training took place in two-hour sessions over 4–8 days, usually 2–3 days a week for 2–4 weeks, depending on the experiment, and was followed by two days of post-training and generalization. The three stages of the experiment were not run on consecutive days, nor were successive training days or the two days of post-training usually consecutive. In extreme cases, as many as 3–4 days might pass between sessions, but we did, nonetheless, generally run 2–3 days each week.

Three–12 listeners were run in each experiment. The lower end is a small number, but it is compensated for by the very large number of responses obtained in all experiments.

2.2

Results

All the results are presented in the same format. First, d' values averaged across the listeners for each pair of vowels contrasting for the same distinctive feature are displayed across training blocks. Only the d' values from the four-vowel complete identification training blocks are presented. Second, the difference between d' values for vowel pairs contrasting for the same distinctive feature is assessed separately for each distinctive feature with a repeated measure ANOVA in which vowel Pair and training Block are within-subjects independent variables. The scores obtained during pretraining are treated as the first training block in these analyses (and displayed as training Block 1 in the figures). Finally, listeners' ability to generalize to complete identification of new pairs of vowels contrasting for these features is assessed. Categorization data aren't presented in this paper because they aren't relevant to

testing PAM's predictions. In the figures, the vowels are represented as follows: 'ii' = /i/, 'ee' = /e/, 'i' = /ɪ/, 'e' = /ɛ/, 'yy' = /y/, 'ooe' = /ø/, 'y' = /ʏ/, 'oe' = /œ/, 'uu' = /u/, 'oo' = /o/, 'u' = /ʊ/, and 'o' = /ɔ/, that is, with two letters for [+tense] vowels and just one for their [-tense] counterparts.

Front Round. Twelve listeners were run in this experiment; four heard speakers B, D, and E, and the other eight heard speakers A, B, and C in training. Of the eight listeners who heard speakers A, B, and C in training, four did only four-vowel complete identification tasks during training, but the other four did two- as well as four-vowel tasks during this stage. These three groups of listeners are referred to as the BDE, ABC, and ABC2 listeners, respectively, and any differences between them are assessed with a between-subjects variable in the ANOVAs. The two panels in Figure 1 show average d' values across training blocks for the two pairs of vowels contrasting for [high] (Fig. 1a) and [tense] (Fig. 1b), separately for these three groups of listeners. From the first training block, ABC listeners (circles) were better at distinguishing the lax vowels /ʏ/ and /œ/ than the tense vowels /y/ and /ø/ for [high] but BDE and ABC2 listeners (squares and diamonds) were equally good at distinguishing both pairs of vowels for this feature. Nonetheless, for the high:mid contrasts (Fig. 1a), the main effect of the difference between Groups of listeners was not significant, $F(2, 9) = 1.809, p > .10$, nor was the main effect of the tense versus lax vowel Pairs, $F(1, 9) = 1.607, p > .10$, and these two variables also did not interact significantly, $F(2, 9) = 1.571, p > .10$. The main effect of training Block was highly significant, $F(15, 135) = 8.010, p < .001$; the significant upward linear trend across Blocks, $F(1, 9) = 26.104, p = .001$, shows that listeners got better at identifying the vowels as a result of training. For the tense:lax contrasts, Figure 1b shows that from the beginning of training all three groups of listeners were better at distinguishing the mid vowels /ø/ and /œ/ than the high vowels /y/ and /ʏ/ for the feature [tense]. The effect of listener Group was not significant, $F(2, 9) = 1.497, p > .10$, but the main effect of mid versus high vowel Pairs was highly significant, $F(1, 9) = 25.095, p = .001$. Vowel Pair did not interact significantly with Group in this comparison, $F(2, 9) = 1.716, p > .10$. The main effect of training Block and the upward linear trend were both highly significant, $F(15, 135) = 9.099, p < .001$; $F(1, 9) = 28.654, p < .001$.

Figure 2 (p.314) shows average d' values from post-training generalization for the "old" stimuli used for training (black bars) and for new stimuli not heard before (grey and white bars). Responses were collected to the new stimuli when they were presented together with the old stimuli (grey) as well as alone (white), and separate d' values were calculated for each set of responses. There was no significant effect of listener Group on listeners' ability to discriminate these vowels for either [high] [$F < 1$] or [tense], $F(2, 9) = 1.591, p > .10$. For the [high] contrasts (Fig. 2a), the difference between tense /y:ø/ versus lax /y:œ/ vowel Pairs was not significant [$F < 1$] and that between old, new /old, and new alone Stimuli was at best marginally significant, $F(2, 18) = 2.681, p = .096$. However, Group and Stimuli did interact significantly, $F(4, 18) = 10.190, p < .001$, because when the new stimuli were presented together with the old, BDE listeners were able to discriminate the new vowels for [high] better than the old, ABC listeners discriminated the new vowels far less well than the old, and ABC2 listeners discriminated new stimuli about as well as

Figure 1

Front Round. d' values averaged across listeners for (a) high:mid (+ high vs. – high) tense vowels (filled squares, circles, diamonds) versus lax vowels (open squares, circles, diamonds), and (b) long:short (+ tense vs. – tense) high vowels (filled squares, circles, diamonds) versus mid vowels (open squares, circles, diamonds). Squares for listeners trained with speakers B, D, and E (BDE), circles for listeners trained with speakers A, B, and C with four-vowel complete identification tasks alone (ABC), and diamonds for listeners trained with speakers A, B, and C with two- as well as four-vowel complete identification tasks (ABC2)

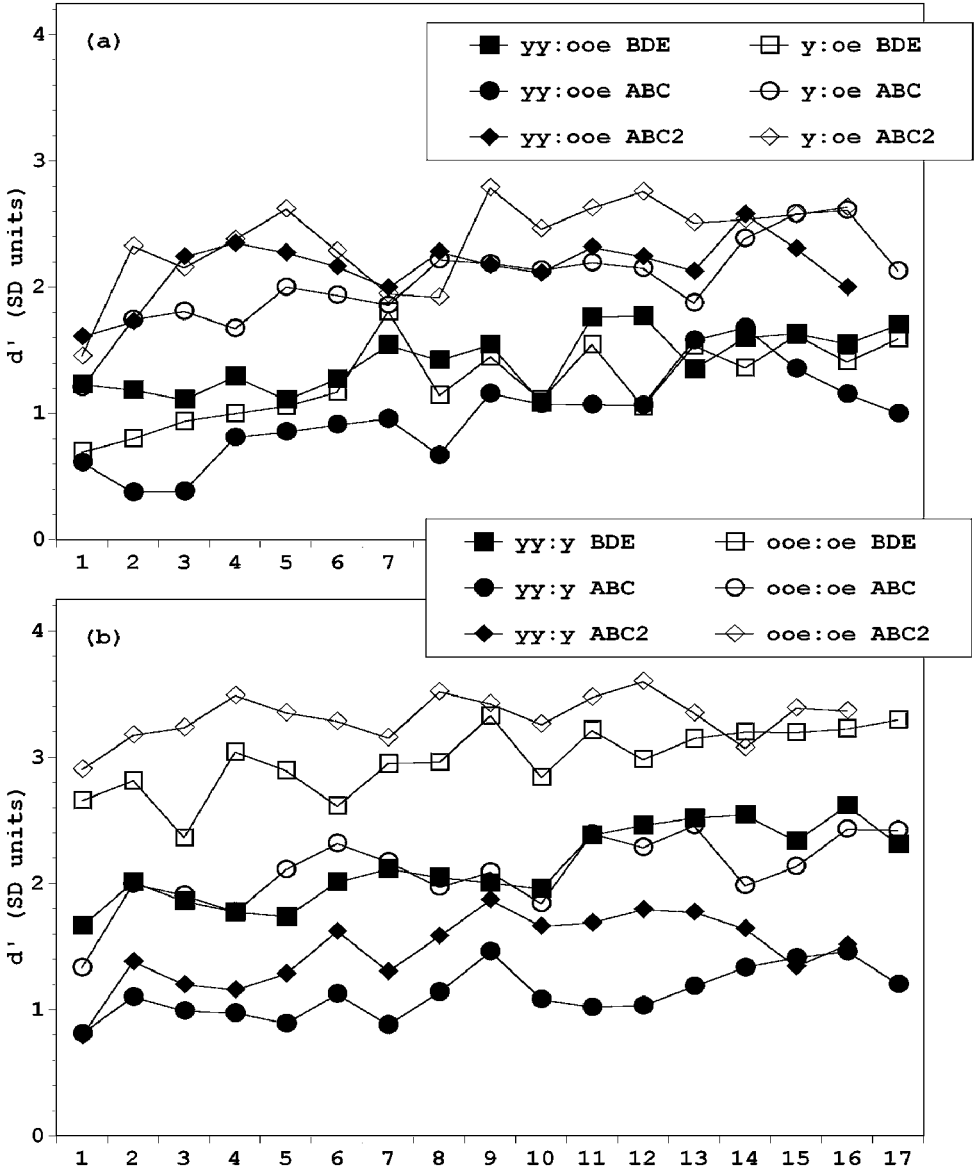
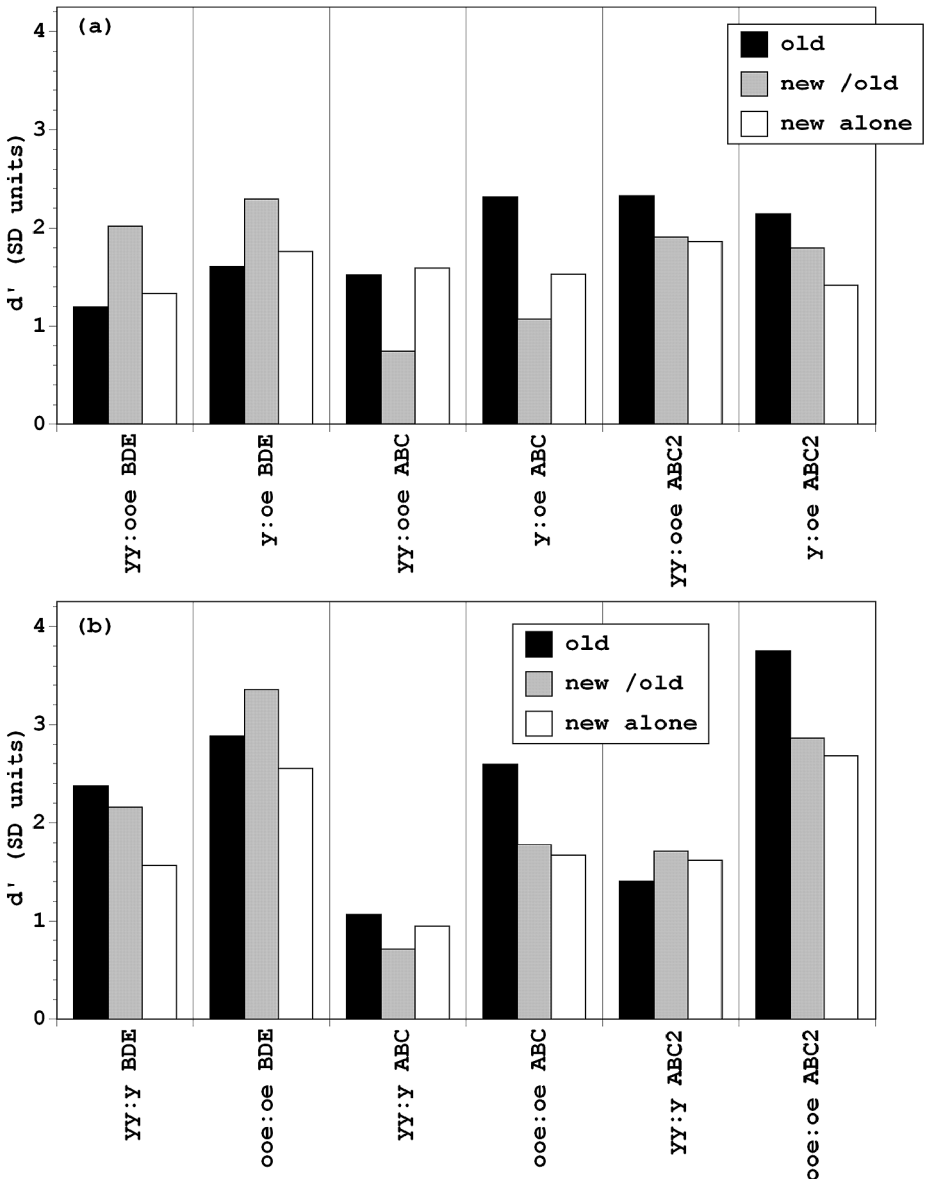


Figure 2

Front Dose. d' values from post-training generalization averaged across listeners for (a) tense /y:ø/ versus lax /y:œ/ vowels contrasting for [high] and (b) high /y:y/ versus mid /ø:œ/ vowels contrasting for [tense]. Black bars represent responses to training stimuli (“old” stimuli), gray bars to new stimuli presented together with old stimuli (“new /old”), and white bars to new stimuli presented alone (“new alone”). BDE represents data from listeners trained with speakers B, D, and E, ABC data from listeners trained with speakers A, B, and C with four-vowel identification tasks alone, and ABC2 data from listeners trained with two- as well as four-vowel identification tasks



old. The planned comparison for new /old versus old was highly significant, $F(2, 9) = 13.855, p = .002$. This difference is inexplicable, as is the fact that the new stimuli were not discriminated better than the old when presented alone by any Group of listeners. The planned comparison for new alone versus old was at best marginally significant, $F(2, 9) = 3.654, p = .069$. Even so, all three Groups of listeners were significantly better at distinguishing the mid vowels /ø/ and /œ/ for [tense] than the high vowels /y/ and /ɥ/, $F(1, 9) = 25.058, p = .001$. A significant three-way interaction was obtained between listener Group, vowel Pair, and Stimuli, $F(4, 18) = 3.558, p = .026$ because listeners' ability to discriminate new stimuli relative to their ability to discriminate old ones varied both with Group and vowel Pair. The planned comparison for this interaction between new /old and old stimuli was significant, $F(2, 9) = 7.493, p = .012$, while that between new alone and old stimuli was marginally significant, $F(2, 9) = 3.794, p = .064$.

Tense Front. Three listeners were run in this experiment. Figure 3a shows that from the beginning of training they were better at distinguishing the front unrounded vowels /i/ and /e/ for [high] than the front rounded vowels /y/ and /ø/. Nonetheless, the main effect of the unrounded /i:e/ versus rounded /y:ø/ vowel Pairs on [high] judgments did not reach significance, $F(1, 2) = 2.320, p > .10$. Listeners' performance

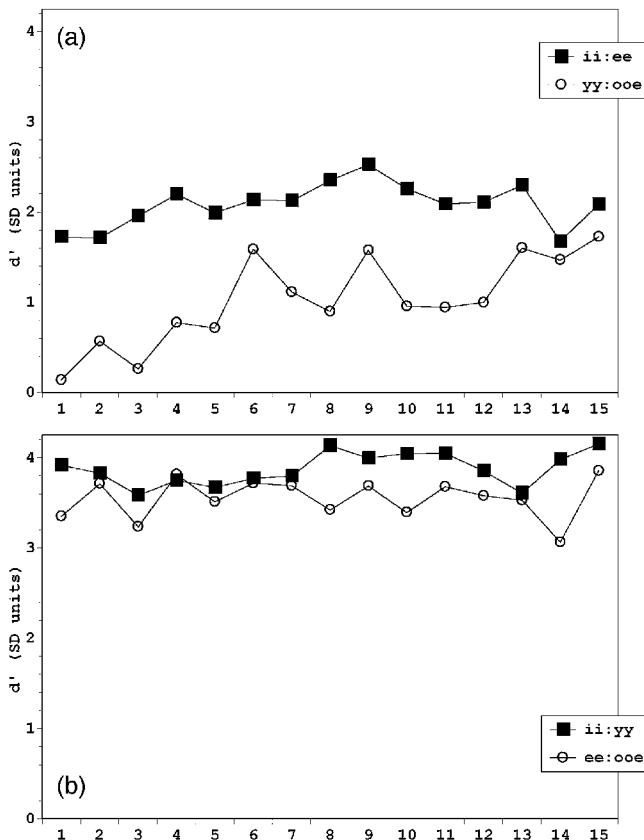


Figure 3

Tense Front. d' values averaged across listeners for (a) high:mid (+ high vs. - high) unrounded vowels (filled squares) versus rounded vowels (open circles), and (b) unrounded:rounded (-round vs. + round) high vowels (filled squares) versus mid vowels (open circles)

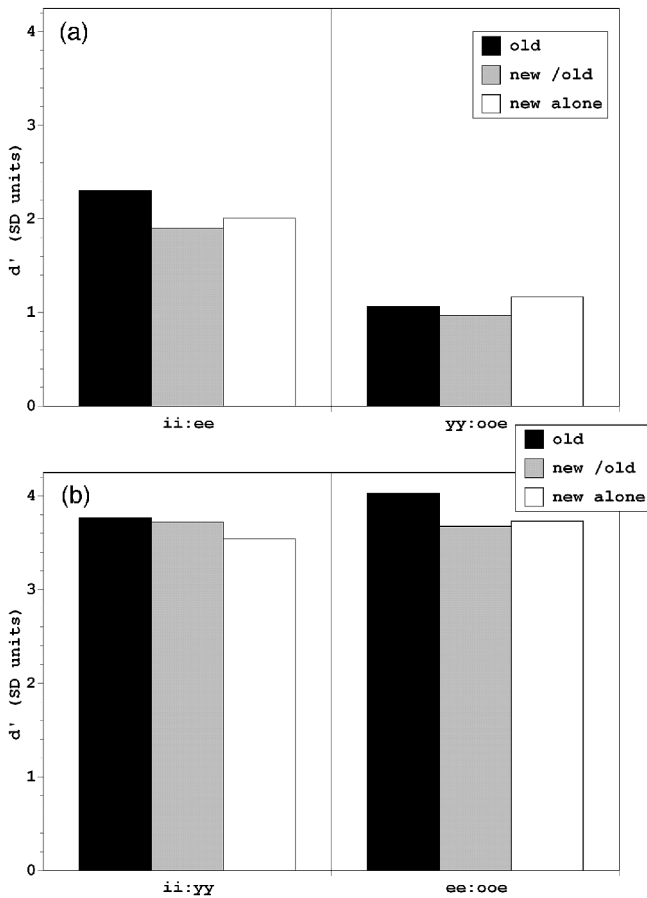


Figure 4

Tense Front. d' values from post-training generalization averaged across listeners for (a) unrounded /i:e/ versus rounded /y:ø/ vowels contrasting for [high] and (b) high /i:y/ versus mid /e:ø/ vowels contrasting for [round]. Black bars represent responses to training stimuli (“old” stimuli), gray bars to new stimuli presented together with old stimuli (“new /old”), and white bars to new stimuli presented alone (“new alone”)

did differ significantly across training Blocks, $F(14, 28) = 5.624, p < .001$, but there was no significant upward linear trend, $F(1, 2) = 7.827, p > .10$. Figure 3b shows that these listeners were equally good at distinguishing the high vowels /i:y/ as the mid vowels /e:ø/ for [round]; indeed, they were nearly at ceiling for both instances of this contrast. The main effect of high versus mid vowel Pairs was not significant for this contrast, $F(1, 2) = 2.846, p > .10$, nor was performance significantly different across training Blocks, $F(14, 28) = 1.702, p > .10$.

Figure 4a shows average d' values from post-training generalization for front unrounded /i:e/ and rounded /y:ø/ vowels contrasting for [high]. The difference in rounding did not have a significant effect on listeners' ability to discriminate vowels for this feature, $F(1, 2) = 3.465, p > .10$. They were also no better at the old than the new Stimuli [$F < 1$], no matter how the new Stimuli were presented. A similar absence of effect of vowel Pair or Stimuli was obtained in the comparison of high /i:y/ versus mid /e:ø/ vowels contrasting for [round], $F(1, 2) = 3.974, p > .10; F < 1$.

Lax Front. Eight listeners were run in this experiment, four with the yes/no two-stimulus tasks during training and four with the same-different two-stimulus tasks.

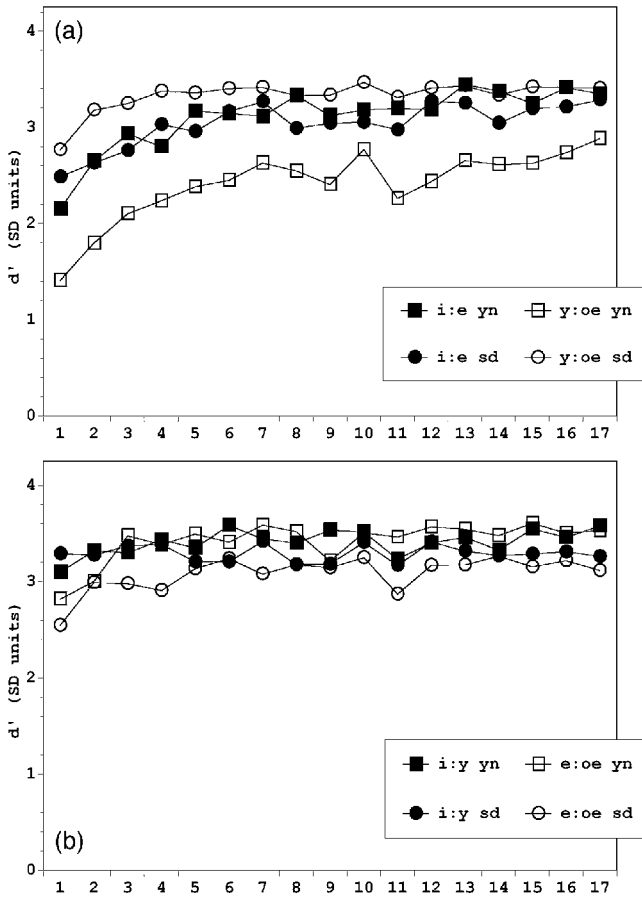


Figure 5

Lax Front. d' values averaged across listeners for (a) high:mid (+ high vs. –high) unrounded vowels (filled squares and circles) versus rounded vowels (open squares and circles), and (b) unrounded:rounded(–round versus + round) high vowels (filled squares and circles) versus mid vowels (open squares and circles). Squares for listeners trained with yes/no two-stimulus tasks (yn) and circles for listeners trained with same-different two-stimulus tasks (sd)

Two-Stimulus (YN vs. SD) is a between-subjects variable in the ANOVAs. Figure 5a shows that YN listeners were better at distinguishing unrounded / $i:\epsilon$ / than rounded / $y:\alpha$ / for [high], but the SD listeners were somewhat better at distinguishing the rounded than the unrounded vowels for this feature. Nonetheless, for this contrast, neither the Two-Stimulus nor vowel Pair effects were significant on their own, $F < 1$ for both, and their interaction was only marginally significant, $F(1, 6) = 4.249, p = .085$. Listeners improved significantly in their ability to discriminate both pairs of vowels across training Blocks, $F(16, 96) = 10.230, p < .001$; upward linear trend: $F(1, 6) = 19.025, p = .005$. Figure 5b shows that both YN and SD listeners were slightly better at the beginning of training in distinguishing the high vowels / i / and / y / than the mid vowels / e / and / α / for [round]; this difference was reflected in a significant interaction between high versus mid vowels Pairs and training Block, $F(16, 96) = 2.315, p = .006$. The upward linear trend for this interaction was also significant, $F(1, 6) = 25.551, p = .002$, because performance improved more for mid than high vowel Pairs. YN listeners were almost significantly better overall than SD listeners, YN $d' = 3.415$, SD $d' = 3.187$; $F(1, 6) = 5.820, p = .052$.

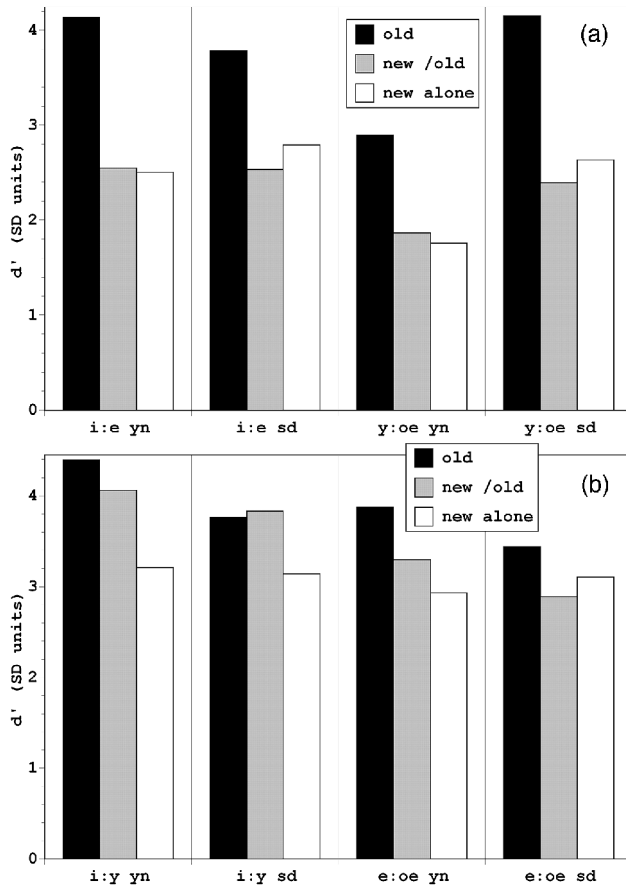


Figure 6

Lax front. *d'* values from post-training generalization averaged across listeners for (a) unrounded /i:ε/ versus rounded /y:œ/ vowels contrasting for [high] and (b) high /i:y/ versus mid /ε:œ/ vowels contrasting for [round]. Black bars represent responses to training stimuli (“old” stimuli), gray bars to new stimuli presented together with old stimuli (“new/old”), and white bars to new stimuli presented alone (“new alone”). YN for listeners trained with yes-no two-stimulus tasks and SD for those trained with same-different two-stimulus tasks

Figure 6a shows that YN but not SD listeners were better at distinguishing unrounded /i:ε/ than rounded /y:œ/ vowels for [high] during post-training generalization; the interaction between Two-Stimulus and vowel Pair was significant, $F(1, 6) = 12.123, p = .013$. Otherwise, both groups of listeners were much better at distinguishing the old than the new stimuli for this feature, $F(2, 12) = 61.892, p < .001$; planned comparisons show that this was so regardless of whether the new stimuli were presented together with the old, $F(1, 6) = 161.884, p < .001$, or alone, $F(1, 6) = 61.487, p < .001$. Figure 6b suggests that YN but not SD listeners were also better at distinguishing high /i:y/ than mid /ε:œ/ vowels for [round], but the Two-Stimulus main effect was at best marginally significant, $F(1, 6) = 4.218, p = .086$, the high:mid vowel Pair main effect was significant, $F(1, 6) = 7.670, p = .032$, and their interaction was not [$F < 1$]. Listeners were significantly better at distinguishing the old than new stimuli for this feature, $F(2, 12) = 14.589, p = .001$. Planned comparisons showed that this was marginally true when the new stimuli were presented together with the old, $F(1, 6) = 4.638, p = .075$ but significantly so when they were presented alone, $F(1, 6) = 35.078, p = .001$.

Tense Round. Four listeners were run in this experiment. Figure 7a shows that listeners

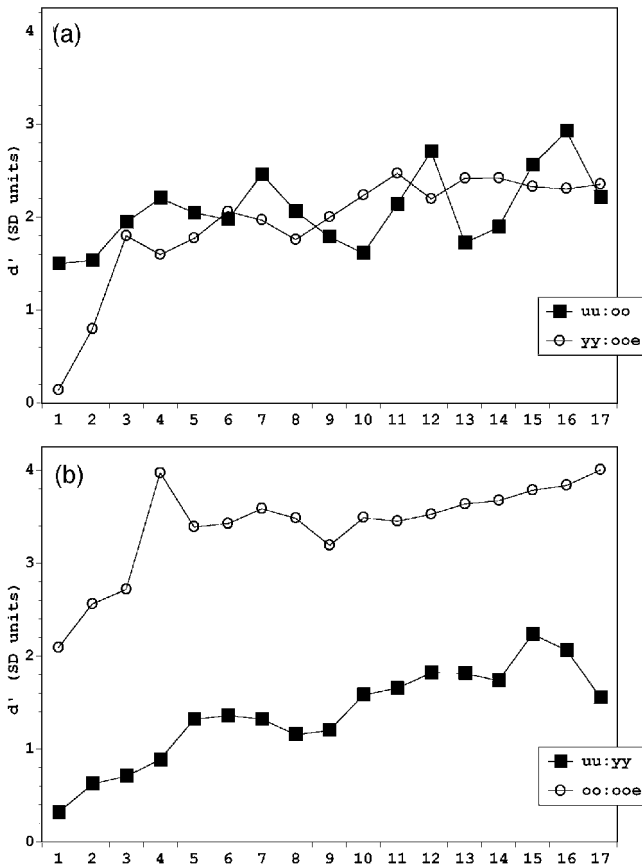


Figure 7

Tense Round. d' values averaged across listeners for (a) high:mid (+ high versus -high) back vowels (filled squares) versus front vowels (open circles), and (b) front:back (-back versus +back) high vowels (filled squares) versus mid vowels (open circles)

discriminated the front vowels /y/ and /ø/ as easily as the back vowels /u/ and /o/ for [high], except at the very beginning of training. The main effect of the difference between the front /y:ø/ versus back /u:o/ vowel Pairs for the [high] contrast was not significant [$F < 1$]. Listeners' ability to discriminate both pairs of vowels improved significantly across training blocks, $F(16, 48) = 8.991, p < .001$; upward linear trend: $F(1, 3) = 83.084, p = .003$. Vowel Pair also interacted significantly with training block, $F(16, 48) = 2.566, p = .006$, because listeners could not discriminate the front vowels /y/ and /ø/ for [high] as well as the back vowels /u/ and /o/ at the beginning of training. Figure 7b shows that listeners discriminated the mid vowels for [back] far better than the high vowels. The main effect of high /u:y/ versus mid /o:ø/ vowel Pairs on the [back] contrast was highly significant, $F(1, 3) = 37.964, p = .009$. Listeners also improved significantly for both pairs of vowels across training blocks, $F(16, 48) = 6.123, p < .001$; upward linear trend, $F(1, 3) = 13.362, p = .035$.

One of the listeners dropped out before post-training generalization, so the results displayed in Figure 8 and the statistical comparisons reported in this paragraph only represent data obtained from the remaining three listeners. Figure 8a shows that listeners were able to discriminate the front vowels /y/ and /ø/ as easily for

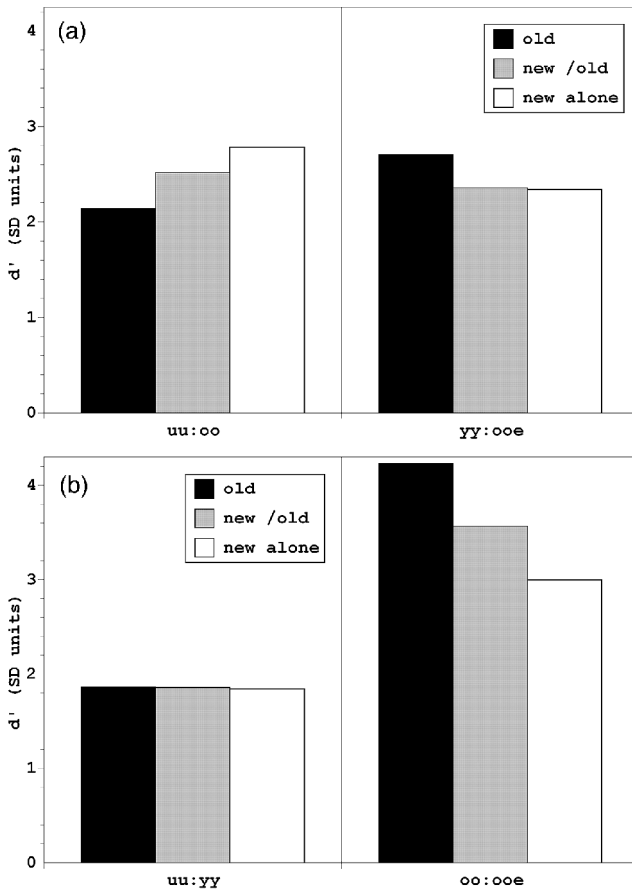


Figure 8

Tense Round. d' values from post-training generalization averaged across listeners for (a) back /u:ɔ/ versus front /y:ø/ vowels contrasting for [high] and (b) high /u:y/ versus mid /o:ø/ vowels contrasting for [back]. Black bars represent responses to training stimuli (“old” stimuli), gray bars to new stimuli presented together with old stimuli (“new /old”), and white bars to new stimuli presented alone (“new alone”)

[high] as the back vowels /u/ and /o/ [$F < 1$]. Responses to old versus new Stimuli also did not differ significantly in accuracy from one another [$F < 1$]. The interaction between vowel Pair and Stimuli was marginally significant, $F(2, 4) = 4.567, p = .093$. Planned comparisons show that this was a result of listeners being somewhat better at distinguishing new than old instances of the back vowels /u/ versus /o/ but less good at distinguishing new than old instances of the front vowels /y/ versus /ø/ when the new stimuli were presented together with the old stimuli, $F(1, 2) = 23.238, p = .040$. They probably were not better with new instances of the back than the front vowels when they were presented alone, $F(1, 2) = 8.541, p = .10$. Figure 8b shows that listeners were better distinguishing the mid vowels /o/ and /ø/ for [back] than the high vowels /u/ and /y/ during post-training. However, the main effect of vowel Pair does not reach significance, $F(1, 2) = 4.571, p > .10$, even though it does interact significantly with Stimuli, $F(2, 4) = 9.717, p = .029$. Planned comparisons for this interaction show that for the mid but not the high vowel Pair, new stimuli were only marginally harder to identify than the old when they’re presented together with the old stimuli, $F(1, 2) = 9.717, p = .094$, but they were significantly harder to identify when presented alone, $F(1, 2) = 23.426, p = .040$.

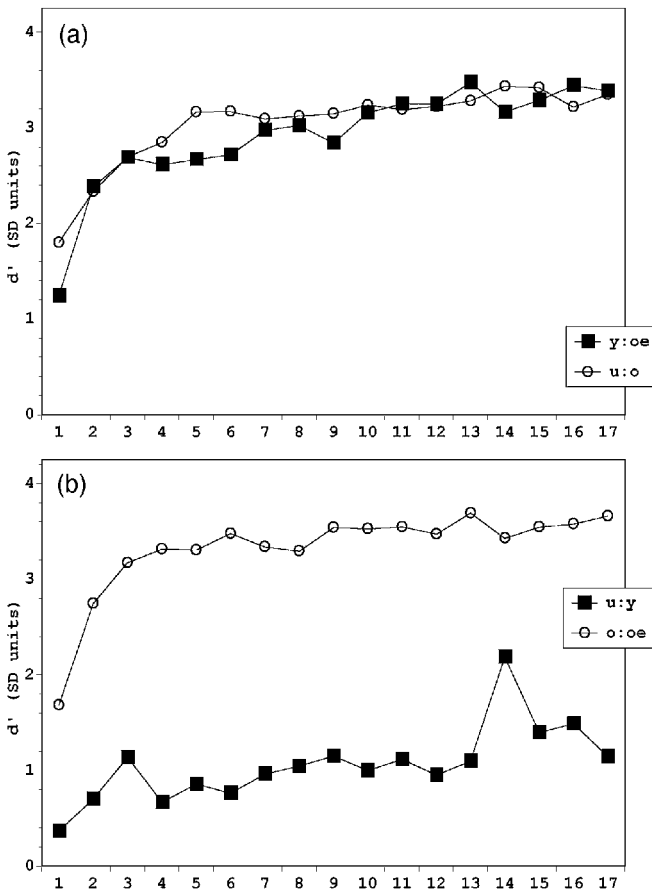


Figure 9

Lax Round. d' values averaged across listeners for (a) high:mid (+ high versus - high) front vowels (filled squares) versus back vowels (open circles), and (b) back:front (+ back versus - back) high vowels (filled squares) versus mid vowels (open circles)

Lax Round. Four listeners were run in this experiment. Figure 9a shows that listeners were equally good at distinguishing back as front vowels for [high], but Figure 9b shows that they were far better at distinguishing mid than high vowels for [back]. For the [high] contrast, the main effect of back / υ : ω / versus front / γ : ϵ / vowel Pairs was not significant, $F(1, 3) = 1.288, p > .10$. The main effect of training Block was significant, $F(16, 48) = 7.512, p < .001$, and improvement with training was indicated by a significant upward linear trend, $F(1, 3) = 23.550, p = .048$. For the [back] contrasts, on the other hand, the main effect of high / υ : γ / versus mid / ω : ϵ / vowel Pairs was significant, $F(1, 3) = 24.655, p = .016$. The main effect of training Block was significant, $F(16, 48) = 6.011, p < .001$, and the upward linear trend almost reached significance, $F(1, 3) = 7.343, p = .073$. Vowel Pair also interacted significantly with training Block in this comparison, $F(16, 48) = 2.283, p = .014$, because performance improved faster for the mid than the high vowel pair during the early training blocks.

In Figure 10a (overleaf), one can see that listeners were equally good at distinguishing the front vowels / γ / and / ϵ / for [high] as the back vowels / υ / and / ω / [$F < 1$]. They were also not significantly worse at distinguishing the new than the old stimuli for this feature, $F(2, 6) = 1.863, p > .10$. However, as can be seen in Figure 10b, listeners

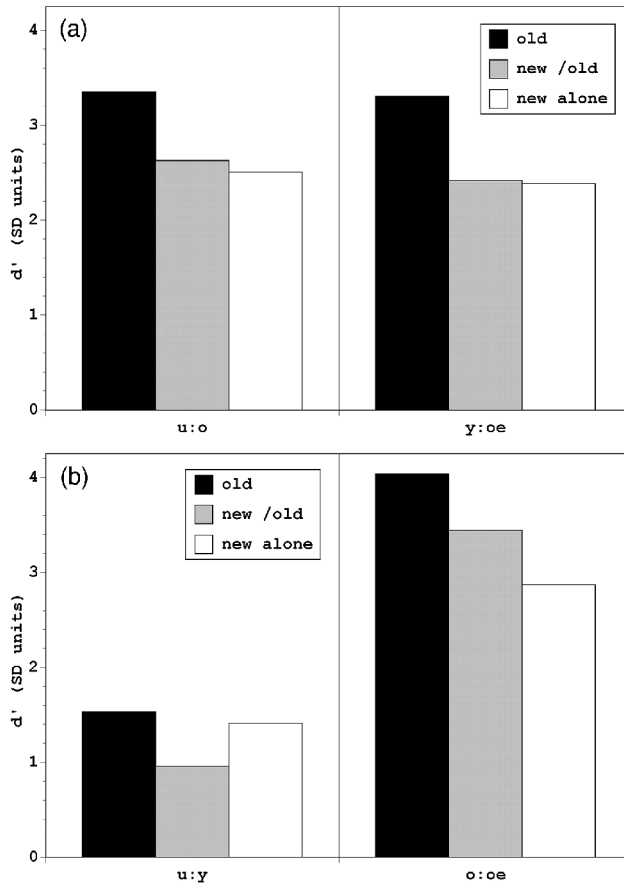
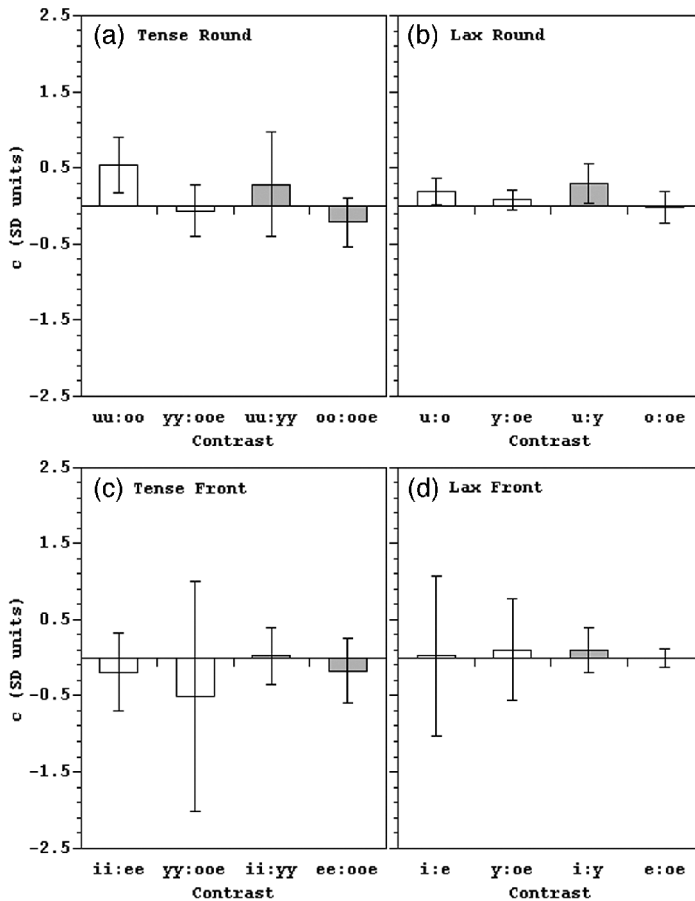


Figure 10

Lax Round. d' values from post-training generalization averaged across listeners for (a) back / $u:o$ / versus front / $y:œ$ / vowels contrasting for [high] and (b) high / $u:y$ / versus mid / $o:œ$ / vowels contrasting for [back]. Black bars represent responses to training stimuli (“old” stimuli), gray bars to new stimuli presented together with old stimuli (“new /old”), and white bars to new stimuli presented alone (“new alone”)

were far better at distinguishing the mid vowels / o / and / $œ$ / for [back] than the high vowels / u / and / y /, $F(1, 3) = 24.697, p = .016$. They were not significantly worse at distinguishing the new than the old vowels for this feature either, $F(2, 6) = 2.940, p > .10$.

Bias. In all but the Front Round experiment, one pair of vowels corresponds reasonably closely to American English vowels: Tense Round / u, o /, Lax Round / $u, ɔ$ /, Tense Front / i, e /, and Lax Front / $i, ε$ / . These vowels are, moreover, more peripheral in the vowel space than the front rounded vowels they’re paired with, / $y, ø$ / in the Tense Round and Front experiments, and / $y, œ$ / in the Lax Round and Front experiments. Results reported by Kuhl et al. (1992) and Polka and Bohn (1996) predict that American English listeners would be biased by their language experience and peripherality to respond with the back rounded vowels / $u, o; u, ɔ$ / or front unrounded vowels / $i, e; i, ε$ / rather than the front rounded vowels / $y, ø; y, œ$ / in these experiments. If peripherality biases responses, listeners would also have preferred the high over the mid vowel response in pairs contrasting for height. Mean bias measures (c) calculated for the backness, rounding, and height contrasts are displayed with

**Figure 11**

Mean bias (c) values with 95% confidence intervals for the (a) Tense Round (b) Lax Round (c) Tense Front, and (d) Lax Front experiments. White bars for [high] contrasts, gray bars for (a,b) [back] and (c,d) [round] contrasts

their 95% confidence intervals in Figure 11. A c value of 0 indicates no bias, a negative value a bias to respond with the first member of each pair, a positive value a bias to respond with the second member. In nearly all cases, the 95% confidence intervals for these means included 0, which indicates that for those pairs listeners did not consistently prefer one response over the other. The three cases where the confidence intervals did not include 0, /u:o/ in the Tense Round Experiment (Figure 11a) and /v:ɔ/ and /v:y/ in the Lax Round Experiment (Figure 11b) indicate a bias to respond with the familiar or more peripheral vowel; however, none of these three biases was particularly strong.

2.3

Discussion

For assessing PAM's or SLM's prediction that all instances of the same contrast should assimilate equally, the results are essentially identical for post-training generalization as training in these five experiments, so the two phases of each experiment will not be discussed separately in this section. Table 3 summarizes the results.

TABLE 3

Summary of results obtained in the first set of experiments. ‘=’ indicates that the first (left) instance of the contrast is as discriminable as the second (right), ‘>’ that it is more discriminable. The latter comparisons are italicized because they deviate from PAM’s and SLM’s predictions

<i>Experiment</i>	<i>[high]</i>	<i>Second feature</i>
Front Round	/ɣ:œ/ > /y:ø/	[tense] /ø:œ/ > /y:ɣ/
Tense Front	/i:e/ = /y:ø/	[round] /i:y/ = /e:ø/
Lax Front	/i:ɛ/ = /ɣ:œ/	[round] /i:y/ = /ɛ:œ/
Tense Round	/u:o/ = /y:ø/	[back] /o:ø/ > /u:y/
Lax Round	/ʊ:ɔ/ = /ɣ:œ/	[back] /ɔ:œ/ > /ʊ:y/

The Front Round experiment showed that the lax vowels /ɣ:œ/ were easier to discriminate for [high] than the tense vowels /y:ø/ and that the mid vowels /ø:œ/ were easier to discriminate for [tense] than the high vowels /y:ɣ/, at least for the ABC listeners. The Tense Front and Lax Front experiments showed that the front rounded vowels /y:ø, ɣ:œ/ were as easy to discriminate for [high] as the front unrounded vowels /i:e, i:ɛ/, despite the fact that the front rounded vowels are foreign categories and the front unrounded vowels are native. These experiments also showed that the high vowels /i:y, i:y/ were as easy to discriminate for [round] as the mid vowels /e:ø, ɛ:œ/. Finally, the Tense and Lax Round experiments showed that the front vowels /y:ø, ɣ:œ/ were as easy to discriminate for [high] as the back vowels /u:o, ʊ:ɔ/, again despite the difference in their foreignness. However, the mid vowels /o:ø, ɔ:œ/ were much easier to discriminate for [back] than the high vowels /u:y, ʊ:y/.

The results of the Tense and Lax Front experiments conform to PAM’s and SLM’s predictions for both [high] and [round] contrasts: both instances of each contrast were equally discriminable in these experiments. Listeners’ equally good ability to discriminate front as back vowels for [high] in the Tense and Lax Round experiments also confirms PAM’s and SLM’s predictions, but their markedly poorer ability to discriminate high than mid vowels for [back] does not. Nor does the listeners’ poorer ability to discriminate tense than lax vowels for [high] and high than mid vowels for [tense] in the Front Round experiment. It does not matter that here listeners’ ability to identify these vowels was assessed over multiple days of training while Best and others only assess their ability when first exposed to the foreign sounds because these differences between instances of a contrast were present from the very beginning of each experiment.

Alternative versions of PAM or SLM could be entertained, in which each foreign category assimilates individually to a native category but pairs of foreign categories that contrast for the same feature in the foreign language do not assimilate in parallel; see Best’s (1994) discussion of “category recognition” and Flege’s (1991) and Best & Strange’s (1992) discussions of the effects of phonetic similarity on foreign category assimilation. In this alternative PAM or SLM, foreign categories would assimilate to the phonetically most similar category in the listener’s native language. Listeners in

the Front Round experiment would be better at distinguishing the mid vowels /ø/ and /œ/ for [tense] than the high vowels /y/ and /ɤ/ because the mid tense vowel assimilates to American English /ɶ/ and the mid lax vowel /œ/ to American English /ʌ/, while both /y/ and /ɤ/ assimilate to American English /u/ or /ʊ/. Similarly, listeners in the Tense and Lax Round experiments would discriminate the high vowels /u:y, ʊ:ɤ/ for [back] less well than the mid vowels /o:ø, ɔ:œ/ because many American English speakers now pronounce their own high back vowels /u, ʊ/ fairly far forward on the palate (de Jong, 1995; Hillenbrand et al., 1995; cf. Peterson & Barney, 1952) and thus with F2 values high enough to bring these vowels close to the German front rounded vowels /y, ɤ/. But these alternative versions of PAM and SLM, in which the phonetic similarity between individual foreign and native categories determines assimilation, makes few if any testable predictions. In these alternative versions, assimilations are an accidental consequence of phonetic similarity rather than of systematic correspondences between phonological categories or their allophonic variants.

Sometimes, phonetic differences between the pronunciations of different instances of the same contrast are perceptually salient, but other times they are not. Polka (1995) reports that in an AXB discrimination task, English listeners were significantly worse at distinguishing the tense German vowels /u/ and /y/ than the lax ones /ʊ/ and /ɤ/. The discriminability of these two pairs of vowels by adult speakers of American English was also tested here, in the Tense and Lax Round experiments, respectively. Although their discriminability was not compared within a single group of listeners, the training data in Figures 7 versus 9 and the post-training generalization data in Figures 8 versus 10 both show that, if anything, these listeners were slightly better at distinguishing the tense vowels /u/ versus /y/ than the lax ones /ʊ/ versus /ɤ/ for [back] (training: tense mean d' value is 1.373 vs. lax 1.062; post-training: tense 1.853 vs. lax 1.300). However, repeated measure ANOVAs showed no significant differences between tense versus lax vowel pairs for the [back] contrast in either the training data ($F < 1$; Tense vs. Lax was a between-subjects variable and training Block a within-subjects variable) or the post-training generalization data ($F < 1$; Tense vs. Lax was a between-subjects variable, Old vs. New/Old vs. New Alone Stimuli a within-subjects variable). My results could differ from Polka's because the much larger irrelevant variation in my stimuli washed away any effect of the tense:lax difference on discriminating high rounded vowels for [back]. However, Polka and Werker (1994) tested the ability of 4, 6–8, and 10–12 month old infants from English speaking households to discriminate these German vowels with the same stimuli as Polka (1995) and found that infants of all three ages discriminated lax vowels for [back] no differently than tense vowels. The fragility of the perceptual effects of pronunciation differences between instances of the same contrast suggests that phonetic similarity probably does not predict how foreign sounds will consistently assimilate to native ones either.

Response biases did not consistently favor the front unrounded and back rounded vowels that correspond reasonably closely to American English vowels, nor did they favor more peripheral vowels. The more peripheral vowels are the same as the familiar vowels for the backness and rounding contrasts but not necessarily for the height contrast. In only 3 out of 12 comparisons did the 95% confidence intervals for mean c values not include 0. In the three cases where these intervals did not include 0, the familiar or more peripheral vowel response was preferred, but that preference was

not very strong. These results therefore do not support predictions one might make from results reported by either Kuhl et al. (1992) or Polka and Bohn (1996), where infant listeners discriminated neighbors of more familiar or more peripheral vowels less well.

3 Exemplar effects on the learning of foreign vowels⁷

Like Logan et al. (1991), I used variable stimuli to train listeners in the experiments reported in the previous section. They found that Japanese listeners learned to discriminate American English /ɪ/ from /ɪ/ better when pronounced by some speakers than others and in some contexts than others (see also Lively et al., 1993, 1994). These findings are compatible with their listeners having learned the particular exemplars they were trained on. To test the predictions of the exemplar model of category learning, I manipulated the extent of speaker and context variation in the training stimuli in another series of experiments. If listeners learned sets of exemplars, then greater variation in speaker or context in the training stimuli should impair generalization to novel stimuli.

Here, too, my methods differ from Best's in a way that more closely simulates natural category learning. Best et al. (2001) presented listeners with six different tokens of each category, but all six were spoken by the same speaker and the phonetic context was fixed. With such limited irrelevant variation, listeners would probably not generalize well to new instances of a category but their failure would tell us very little because they'd been given so little opportunity to learn which dimensions were relevant.

3.1

Methods

Three variants of the Front Round experiment were run to test this hypothesis. The first is the one described above in which the training stimuli varied across three speakers and 12 consonantal contexts. This is called the Mixed Speaker and Context experiment or the Mixed experiment for short. The relevant data here are those obtained from the four listeners trained with speakers A, B, and C, without any two-stimulus tasks during training (the ABC listeners in the description of the Front Round experiment above). In the second variant, the speaker was fixed to one of A, B, or C in some training blocks and all 18 consonantal contexts were used; other training blocks used all three speakers and just 12 consonantal contexts, as in the Mixed experiment. Training blocks of the two kinds alternated, and training blocks with just one speaker were rotated systematically through the three speakers. This is called the Fixed Speaker experiment. In the third variant, the consonant preceding the vowel could only be one of /b/, /g/, or /t/ in some training blocks; in others, all three of these consonants occurred before the target vowel. Training blocks of the two kinds alternated, and training blocks in which the preceding consonant was

⁷ The results reported in this section were reported earlier (Kingston, Bartels, Rice, Benki, Moore, Thorburn, & Macmillan, 1996).

fixed were rotated systematically through the three consonants. In blocks where the preceding consonant was fixed, stimuli were produced by all five speakers. This is called the Fixed Context experiment. Note that when variation was reduced in either speaker or context, it was maximized in the other source of irrelevant variation. Otherwise, the procedures used in the two Fixed experiments were exactly the same as those used in the Mixed experiment. Four listeners were run in each variant. Table 4 summarizes the differences between the three experiments.

TABLE 4

Number of speakers and consonantal contexts in the three experiments testing the predictions of the exemplar model of category learning

<i>Experiment</i>	<i>Number of speakers</i>	<i>Number of consonantal contexts</i>
Mixed Speaker and Context	3	12
Fixed Speaker	1	18
Fixed Context	5	3

Listeners were recruited in the same way as in the previous set of experiments and selected according to the same criteria. They were paid \$7 per hour for participating.

The most direct test of the extent to which irrelevant variation in training stimuli influences category learning is to compare how much better listeners are between the different variants at identifying the stimuli after training than they were before training. For each listener, d' values obtained in pretraining were subtracted from those obtained after training for the old stimuli that the listeners had been trained on. Pretraining d' values were also subtracted from those obtained to the new stimuli that they heard for the first time after training. These differences between new and pretraining d' values estimate the extent to which training affects generalization. The new stimuli were presented together with the old stimuli and alone, so differences were calculated separately for each presentation. In the repeated measure ANOVAs, the dependent variable is these post-minus pretraining differences in d' values, and the independent variables are the between-subjects variable, training Variant (Mixed, Fixed Speaker, vs. Fixed Context), and the within-subjects variables, Comparison (Old Post–Old Pre, New /Old Post–Old Pre, vs. New Alone–Old Pre) and vowel Pair (high vs. mid for vowels contrasting for [tense] and tense versus lax for vowels contrasting for [high]). As in the previous section, separate ANOVAs were carried out for the two contrasts.

3.2

Results

Listeners' improvement in identifying vowels as [high] was unaffected by training Variant alone [$F < 1$; Figure 12a]; however Variant did interact significantly with Comparison, $F(4, 18) = 5.901, p = .003$. As can be seen by comparing the left panel of Figure 12a with the other two panels, this interaction probably came from new stimuli being identified no better than the pretraining stimuli when they were presented

together with old stimuli in the Mixed Variant. Otherwise, listeners improved about equally much for [high] in all three Variants in all comparisons between post- and pre-training. There was no significant effect of vowel Pair [$F < 1$], but the interaction between it and Variant was marginally significant, $F(2, 9) = 3.496, p = .075$, because listeners improved more for the lax vowels /ɪ/ and /æ/ than the tense vowels /y/ and /ø/ in the Fixed Context Variant but more for the tense than the lax vowels in the Mixed and Fixed Speaker Variants (Fig. 13).

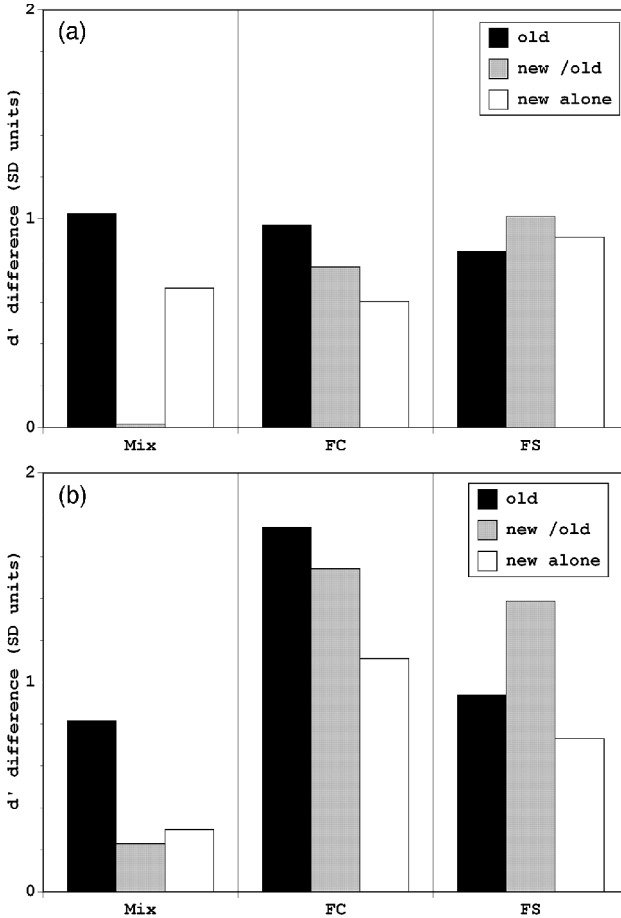


Figure 12

Differences in d' values (after–before training) averaged across listeners in Mixed Speaker and Context (Mix), Fixed Context (FC), and Fixed Speaker (FS) Variants of the Front Round experiment. (a) vowel Pairs contrasting for [high] /y:ø, y:æ/ (b) vowel Pairs contrasting for [tense] /y:y, ø:æ/. Black bars represent responses to training stimuli (“old” stimuli), gray bars to new stimuli presented together with old stimuli (“new /old”), and white bars to new stimuli presented alone (“new alone”)

Listeners improved more in distinguishing vowels for [tense] in the Fixed Context and Speaker Variants than the Mixed Variant, $F(2, 9) = 4.174, p = .052$; Figure 12b, but only the difference between the Fixed Context and Mixed Variants approached significance in the planned comparison ($p = .054$). In this comparison, too, training Variant interacted significantly with Comparison, because listeners identified the new stimuli little better than the pretraining stimuli but did improve on those stimuli, that is, the old stimuli, in the Mixed Variant (left panel of Figure 12b). They also identified the new stimuli better than the pretraining stimuli to the

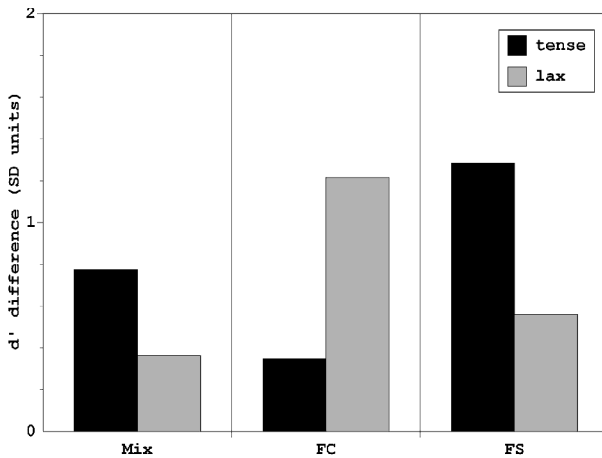


Figure 13

Difference in d' values averaged across listeners between tense /y:ø/ vowels (black bars) versus lax /ɜ:œ/ vowels (grey bars) contrasting for [high] across Mixed Speaker (Mix), Fixed Context (FC), and Fixed Speaker (FS) Variants of the Front Round experiment

same extent as they improved in identifying the old stimuli in both Fixed Variants (middle and right panels of Figure 12b). Listeners also improved significantly more at distinguishing the mid vowels /ø/ and /œ/ for [tense] than the high vowels /ɜ/ and /ɜ/, $F(1, 9) = 17.579, p = .002$. Vowel Pair interacted significantly with Comparison, $F(2, 18) = 3.576, p = .049$, because the mid vowels' advantage was best for old vowels and diminished for new vowels presented together with the old vowels and then again for new vowels presented alone (Fig. 14).

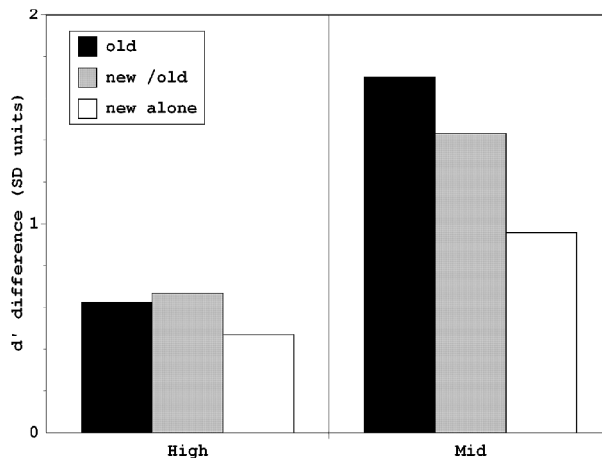


Figure 14

Differences in post versus pre-training d' values averaged across listeners for high /y:ɜ/ versus mid /ø:œ/ vowels contrasting for tense as a function of which post-training vowels are compared with those heard before training. Black bars represent responses to training stimuli ("old" stimuli), gray bars to new stimuli presented together with old stimuli ("new /old"), and white bars to new stimuli presented alone ("new alone")

3.3

Discussion

Constraining irrelevant variation in speaker or context did not noticeably affect how much listeners improved at identifying old stimuli or generalized to novel ones for [high] as a result of training; however, constraining variation in consonantal context

did significantly increase improvement for [tense]. Constraining variation in speaker increased improvement for [tense], too, just not significantly. The effects of constraining variability on learning to identify vowels contrasting for [tense] support the hypothesis that listeners learned sets of exemplars in training, as did the Japanese listeners' in Logan et al.'s experiments. However, the absence of any effect of constraining either context or speaker variability on how much listeners improved at distinguishing vowels for [high] disconfirms this hypothesis, as a predicted effect fails to emerge. This difference does not obviously depend on differences in the overall discriminability of these two contrasts in the Front Round experiment. The members of both [high] and [tense] contrasts are discriminated more or less equally well during training (Fig. 1). Absolute rather than relative post-training performance is also more or less the same for both contrasts (Fig. 2).

Listeners also responded differently to [high] than [tense] contrasts in the other experiments reported in the previous section. There, they discriminated back versus front, rounded versus unrounded, and tense versus lax vowels for [high] equally well or poorly, but they discriminated high vowels for [tense] differently from mid vowels. That is, those results show that listeners can separate a vowel's value for [high] from its values for any other distinctive feature but not its value for [tense]. Here, listeners could categorize vowels for [high] regardless of how much consonantal context or speaker varied irrelevantly but could not do so for [tense]. We'll see that [high] and [tense] also behave differently in the next set of experiments.

The results obtained with both features are in fact problematic for Nosofsky's (1986) model of category learning, in which variation along dimensions that are irrelevant to the categorization teaches observers what the relevant dimensions are. Such irrelevant variation is greatest in the Mixed Variant, at least in that both speaker and consonantal context vary simultaneously, while in both the Fixed Variants, variation is limited in either speaker or consonantal context even though it is maximized in the other irrelevant dimension. If irrelevant variation draws observers' attention to the relevant dimensions, then the listeners in this experiment should have learned more in the Mixed than either of the Fixed Variants. They did not do so for [high] and they plainly did the reverse for [tense]. The Japanese listeners tested by Pisoni and colleagues also did not behave strictly in accordance with the predictions of Nosofsky's model in that even three and six months after training they still identified /l/ and /ɹ/ better in some contexts and when spoken by some speakers than others. For them, too, irrelevant variation remained a salient property of the stimuli. At the same time, comparison of Pisoni and colleagues' results with Strange and Dittman's (1984) shows clearly that listeners learn more when stimuli vary irrelevantly than when they do not. This comparison together with the results reported here suggests a Goldilocks moral: some variation is better than none but too much is harmful. Taken together, these results also show that observers' attention to irrelevant variation may not diminish as they learn the categories in the way Nosofsky's model predicts. They show instead that both Japanese and American English listeners have learned contextual and idiolectal allophones. In the next section, I turn to a final set of experiments that directly pit Nosofsky's attention-weighted exemplar model of category learning against a competing feature learning model.

4 Distinguishing exemplar from feature learning⁸

In Nosofsky's attention-weighted exemplar model, category learning is a secondary process that depends on which dimensions the observers' attention is selectively focused. In such a model, a natural class is the set of all tokens of that category which the listener has ever experienced. The evidence for this is that listeners retain and consult information about individual stimulus tokens that is irrelevant to a phonetic task (Goldinger, 1996). Most pertinent here is Logan et al.'s (1991) demonstration that speaker and phonetic context influenced the ability of Japanese listeners to learn to identify American English /1/ versus /ɪ/. The experiments reported in the previous section showed that high context and perhaps also speaker variability impaired learning and generalization of the feature [tense] by American English listeners. Such evidence contradicts traditional linguistic models in which listeners recognize a phonemic category by extracting an abstract distinctive feature value from the physical, phonetic content of the speech signal (Iverson & Kuhl, 1995; Jakobson et al., 1953; Johnson, 1997; Marcus, Vijayan, Bandi Rao, & Vishton, 1999; Trubetzkoy, 1939/1969; cf. Bybee, 2001; Pierrehumbert, 2002, Saffran, Aslin, & Newport, 1996). In the traditional model, extraction is a process of stripping away and discarding irrelevant variation that is due to the context in which the phoneme is uttered, the identity of the speaker, and so forth. The experiments reported in the previous section did not, however, test variability across the different phonemes belonging to the natural class defined by a distinctive feature value. In those experiments, listeners learned to categorize tokens as individual phonemes, for example, /y/ versus /ø/, against the backdrop of contextual and speaker variation. In the experiments reported in this section, they learn to categorize tokens of several vowel phonemes into the larger natural classes defined by distinctive features, for example, [+high] versus [-high]. Other linguistic information (e.g., [back]) thus becomes task-irrelevant, like speaker and phonetic context.

4.1

Methods

The number of vowel phonemes in each natural class was manipulated in the training stimuli (henceforth more vowels, for short). If each class is a disjunctive set of exemplars, more training vowels should, like more contexts or speakers, impair generalization to a new vowel pair. If learning a natural class is instead discovering and attending selectively to the class's defining properties (Nosofsky, 1986), more training vowels should always improve generalization. Finally, if learning a natural class is discovering its defining distinctive feature value, adding more training vowels should improve learning only so long as doing so narrows down the choice between the class's defining feature value versus the other values that vary irrelevantly within the natural class.

Stimuli. Stimuli were the same two-syllable German words or pseudowords used in the other experiments reported here. The stimuli used for these experiments were spoken by speakers A, C, D, and E. Throughout the experiment, stimuli were presented in blocks differing in speakers and the consonants flanking the target vowel:

⁸ The experiments reported in this section of the paper have been reported in two earlier papers, coauthored with Elliott Moreton (Kingston & Moreton, 1998, 2001).

(1) Speakers D and E; Contexts [b_p, t_t, g_k, t_k] and (2) Speakers A and C; Contexts [p_p, d_t, k_k, d_p]. A fourth condition in the experiment is thus the Speaker-Context combination, which measures the extent to which listeners learn and generalize equally to different stimulus sets. For brevity, the two Speaker-Context combinations are referred to by the letters identifying the speakers, DE versus AC. Peak amplitudes, which always occurred within the target vowel, were equalized in the stimuli. They were presented binaurally through TDH-49 headphones at 20 kHz, low-pass filtered with an 11-pole filter at 8.133 kHz (energy was down at least 80 dB at the Nyquist frequency).

Design and conditions. The experiment had two phases: training and training-testing. American English listeners were first trained with feedback to categorize one, two, or three pairs of German rounded vowels according to their values for a single distinctive feature. They were then tested with a new pair of German rounded vowels. Listeners were assigned to one of the 18 conditions produced by orthogonal combination of three variables. First, training stimuli were drawn from the ends of parallel edges of the cube in Figure 15—the “Edge” rule—or from the ends of diagonals through the body of the cube—the “Body” rule. Each face of the cube represents a natural class, for example, the foreground face is the class of [–back] (rounded) vowels, the right hand face is the [–tense] vowels, and the top face is the [+high] vowels. The pair of vowels connected by the heavy line in this figure, /y:ø/, are drawn according to the Edge rule and differ in their value for just a single distinctive feature, [high], being [+high] versus [–high], respectively. This one pair is thus enough to identify those natural classes.⁹ The pair of vowels connected by the dashed line through the body of the cube, /y:ɔ/, are drawn according to the Body rule and differ instead in all three features, being [+high, –back, +tense] versus [–high, +back, –tense], respectively. This pair cannot by itself identify any of these natural classes. The second condition is the number of pairs of training vowels, which varied from One to Two to Three. The third condition is the feature by which the listeners were trained to categorize the vowels: [back] versus [high] versus [tense]. No listeners were trained with Two pairs of vowels for the feature [tense] with

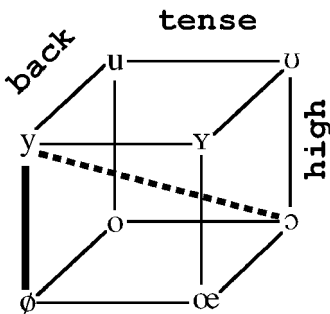


Figure 15

Rounded German vowels, by their values for [back], [tense], and [high]. The heavy solid line connects a pair of vowels /y/ versus /ø/ differing just in [high], the heavy dotted line a pair of vowels /y/ versus /ɔ/ differing in all three features

⁹ Neither this pair of vowels nor any other used in the experiments reported in this section differ in their value for [low]; all were [–low]. This analysis is confirmed by the low-to-mid F1 values for these vowels listed in the Appendix. Listeners could therefore conclude that the values of this feature were irrelevant to categorizing the vowels in all these experiments.

either the Body or the Edge rules, so results are available only for 16 out of the 18 possible conditions (2 Rules × 3 Numbers of training vowel pairs × 3 Features). Table 5 lists the training and test vowels used in each of the 16 conditions; the vowel pairs that would be used in Two Pair training for the feature [tense] are listed for completeness.

TABLE 5

Pairs of training and test vowels chosen according to Edge (top) and Body (bottom) rules for the features [back], [high], and [tense], when 1, 2, or 3 three pairs of vowels were used in training. Rows labeled 'feature?' using the Body rule list the features by which the training vowels in the cell immediately above them could be classified for each number of training vowel pairs. No data have been collected using two pairs of training vowels for the feature [tense]; the vowels that would have been used in this condition are printed in italics

<i>Rule</i>	<i>Feature</i>	<i>1 pair</i>	<i>2 pairs</i>	<i>3 pairs</i>	<i>Test pair</i>
Edge	back	ɔ:œ	ɔ:œ o:ø	ɔ:œ o:ø u:y	u:y
	high	ʊ:ɔ	ʊ:ɔ y:œ	ʊ:ɔ y:œ u:o	y:ø
	tense	o:ɔ	o:ɔ ø:œ	o:ɔ ø:œ u:ʊ	y:y
Body	back	ʊ:ø	ʊ:ø ɔ:y	ʊ:ø ɔ:y u:œ	o:y
	feature?	back tense high	back tense	back	back
	high	ʊ:ø	ʊ:ø u:œ	ʊ:ø u:œ y:o	y:ɔ
	feature?	high back tense	high back	high	high
	tense	u:œ	u:œ y:ɔ	u:œ y:ɔ o:y	ø:ʊ
feature?	tense high back	<i>tense</i> <i>high</i>	tense	tense	

For each pair of vowels in the table, the left one was assigned to one response category, the right one to the other. For the Body rule, the features that could be relevant for categorizing the vowels for each number of pairs of training vowels are listed below the pairs of vowels themselves, in the rows labeled ‘feature?’. As shown in these rows of the table, adding more pairs of training vowels with this rule progressively narrows down the choice between possibly relevant features. Furthermore, for all numbers of pairs of training vowels with this rule, adding the test pair identifies the relevant feature. No comparable list is needed for the Edge rule because the relevant feature can be identified with just one pair of training vowels.

Listeners and procedures. Listeners were recruited in the same way as in the two previous sets of experiments and selected according to the same criteria. They were paid \$7 per hour for participating.

Listeners first heard four training blocks, two with each combination of speakers and consonantal contexts. Following 16 ordered but unscored orientation trials, 96 test trials were presented in each training block, 48 for each category. With One training pair, each vowel was presented 48 times per block (each token 6 times), with Two training pairs, each vowel was presented 24 times (each token 3 times), and with Three training pairs, each vowel was presented 16 times (each token 2 times). After training on the same day, listeners heard four training-testing blocks, again two with each combination of speakers and contexts. In the training-testing blocks, 48 trials with the test vowels, 24 for each category (each token 3 times), were added to the 96 trials with the training stimuli, for a total of 144 trials. Order of stimulus presentation was random within a block and speaker-context block order was counterbalanced across groups of listeners.

A single stimulus was presented on each trial, and the listeners identified the category to which its vowel belonged by pressing one of two buttons. Following their response, a light came on above the button the listeners should have pressed. This feedback was given on all trials, throughout the initial training and the subsequent training-testing. Listeners had 2000 ms after the end of the stimulus to respond, the feedback light came on for 500 ms after their response, and a 500 ms pause followed before the next stimulus was presented.

Each listener was assigned to one of the 16 conditions described above. Each one categorized the vowels for just one distinctive feature, with One or Two or Three pairs of training vowels, drawn according to the Body or Edge rule. Four to six listeners were run in each condition (except Two training pairs for the feature [high] with the Body rule, in which only 1 listener has been run).

4.2 **Predictions**

If reducing phonological uncertainty matters, then listeners should categorize the test vowels more accurately with the Body but not the Edge rule when they’ve been trained with more pairs of vowels: Body $1 < 2 < 3$ versus Edge $1 = 2 = 3$. If this outcome is obtained, the categories are defined by distinctive feature values. For brevity, I will refer to this as the “feature” learning model.

Although phonological uncertainty is not reduced by adding pairs of training vowels with the Edge rule, those additional pairs may still make it easier for listeners to categorize the vowels by distinguishing the relevant from the irrelevant physical dimensions along which the contrasting categories differ. Nosofsky (1986) showed that observers categorize a stimulus on the basis of its aggregate perceptual similarity to all members of its category rather than its similarities to the category's individual members, and that selective attention stretches perceptual distances — the inverse of similarity — along dimensions relevant to that categorization and shrinks them along irrelevant dimensions. Selective attention should be effective for the natural classes to which speech sounds, even foreign ones like these, belong by virtue of their distinctive feature specifications, because distinctive features have reliable phonetic correlates in a particular context and language that wax in salience in proportion to the number of training pairs. Selective attention should be particularly effective for these German vowels, because they all contrast for features that are distinctive in English vowels, too. If so, then with the Edge as well as the Body rule: $1 < 2 < 3$. If this outcome is obtained, categories are sets of exemplars, in which selective attention increases perceived similarity between exemplars within a category and likewise perceived dissimilarity between exemplars of contrasting categories. More training pairs also increases each vowel's neighborhood density (Vitevitch & Luce, 1999), which should facilitate grouping the new vowels with the old because they share properties with the old vowels. This will be referred to as the “selective attention” model.

Finally, it's possible that adding more training pairs may instead make it harder to learn how to categorize the test vowels with both the Body and Edge rules, because the differences between the phonemes that make up each category may be difficult to ignore. That is, listeners may learn each category as an arbitrary disjunctive set rather than discovering the relevant physical similarity among its members or the distinctive feature value that defines that natural class. If so, then with both the Body and Edge rules: $1 > 2 > 3$. This outcome might be expected from the results of the experiments reported in the previous section, where too much irrelevant variation in consonantal context and speaker impaired learning and generalization for the feature [tense] (but not [high]). These new experiments thus test whether variation between the phonemes composing a natural class has the same effects on category learning as contextual and speaker variation. This will be referred to as the “disjunctive sets” model.

4.3 Results

Responses from the two presentations of a Speaker-Context combination were combined within each of the two phases of the experiment, training and training-testing, to produce three sets or Stages of responses per combination: First Training, Second Training during the training-testing phase, and Testing. Listeners' ability to categorize the stimuli was measured by calculating d' values for each of the three sets of responses, treating their task as yes-no classification. The First Training d' values were then subtracted from the Second Training d' values, and the Second Training d' values were subtracted from the Testing d' values. The Second – First Training difference represents the extent to which listeners performance improved with training,

and the Testing–Second Training difference the extent to which they could generalize their training to new categories of vowels contrasting for the same distinctive feature. The effects of the various experimental conditions on these differences were then evaluated in separate repeated measures ANOVAs for each distinctive feature. The within-subjects variables in these analyses were these Differences, Second–First Training and Testing–Second Training, and Speaker-Context combination, DE versus AC. The between-subjects variables were the Rule (Body vs. Edge) and Number of pairs of training vowels (One vs. Two vs. Three). The predictions of the different models were tested by the interaction between the Number of training vowels and the training Rule. Learning features predicts a significant interaction because Testing performance relative to Second Training is predicted to improve as the Number of training vowels increases with the Body but not the Edge Rule. Learning to attend selectively to the relevant dimension predicts no significant interaction between these two variables but instead improvement with Number of training vowels for both Rules. Learning disjunctive sets also predicts no interaction, but a decline in performance with the Number of training vowels for both Rules.

Figure 16a shows for the feature [back] that listeners improved relatively little between First and Second Training when trained with either Rule or any Number of vowel pairs. Figure 17a shows, however, that Testing performance got better relative to Second Training as the Number of training vowel pairs increased, especially for listeners trained with the Body rule, who actually did better with the new vowels in Testing than the old vowels in Second Training when trained with Three pairs of vowels. Performance was also markedly better during Testing relative to Second Training for Edge rule listeners trained with Two rather than just One pair of vowels, but listeners trained with Three pairs did not identify the new vowels better than those trained with just Two. Improvement in classifying the old vowels between the two stages of training was largely unaffected by either the training Rule or the Number of training vowels, perhaps because relatively little improvement occurred, but generalization to new vowels was affected by both. The main effects of Body versus Edge Rules and Number of training pairs were both highly significant, $F(1, 24) = 29.836, p < .001$; $F(2, 24) = 11.655, p < .001$, and their interaction nearly was, $F(2, 24) = 3.262, p = .056$. Both Rule and Number also interacted significantly with the Difference between stages of the experiment, $F(1, 24) = 20.767, p < .001$; $F(2, 24) = 3.786, p = .037$.

The difference caused by the Number of training vowels between listeners trained with the Body rather than the Edge Rules was larger and more consistent across comparisons for the feature [high]. Listeners trained with the Body rule not only classified the new vowels relative to the old better as the Number of training vowels increased (Fig. 17b, p. 338) but also classified the old vowels better during the Second stage of training relative to the First stage (Fig. 16b). Both learning and generalization by listeners trained with the Edge Rule improved far less as the Number of training vowels increased. The main effect of the Body versus Edge Rule was not significant [$F < 1$] but that of the Number of training vowels was, $F(2, 19) = 17.777, p < .001$, as was the interaction between these two variables, $F(2, 19) = 5.192, p = .016$. Rule did not interact significantly with Difference [$F < 1$], but Number did, $F(2, 19) = 4.965, p = .018$.

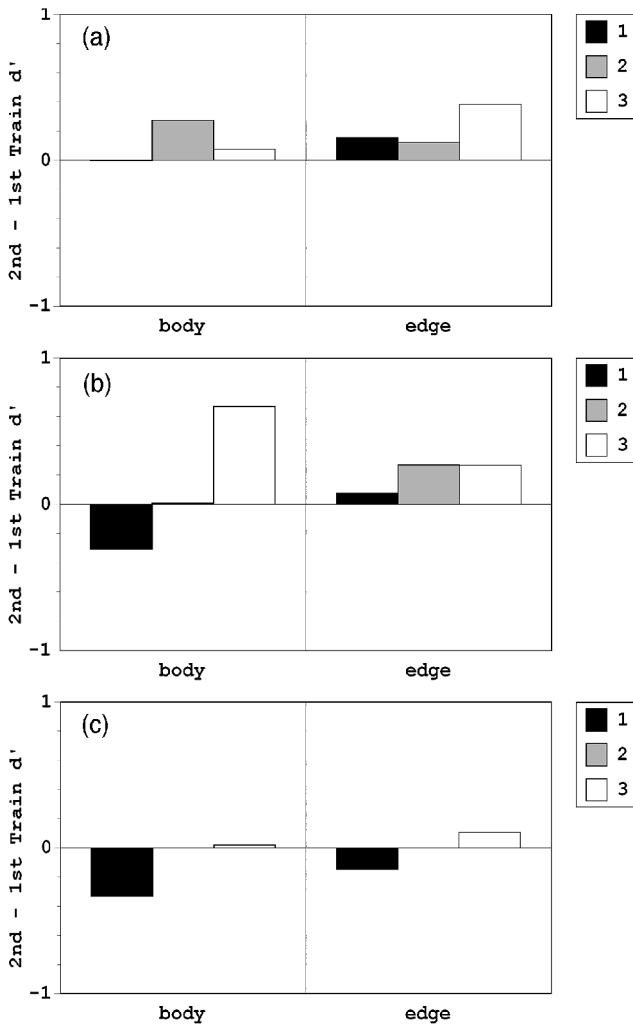


Figure 16

Difference in d' values between Second and First Training averaged across listeners trained with the Body versus Edge Rules and one versus two versus three Pairs of vowels for the features (a) [back] (b) [high], and (c) [tense]. Bars of 0 height are given for two pairs of vowels for the feature [tense] because no listeners have been run in that condition

In sharp contrast to the differences caused by Number of training vowels between Body and Edge Rules for the features [back] and [high], listeners did better with the Edge as well as the Body rules for [tense] during Second Training relative to First (Fig. 16c) and during Testing relative to Second Training (Fig. 17c) when they're trained with Three rather than just One pair of vowels. The main effect of Number of training vowels was highly significant, $F(1, 15) = 17.715, p = .001$, but neither the main effect of Body versus Edge Rule nor its interaction with Number was [both $F < 1$]. Rule did not interact with Difference [$F < 1$] and the interaction between Number and Difference was at best marginally significant, $F(1, 15) = 3.356, p = .087$. The main effect of Difference itself was highly significant, $F(1, 15) = 11.432, p = .004$, because differences were smaller between Second and First Training than Testing and Second Training.

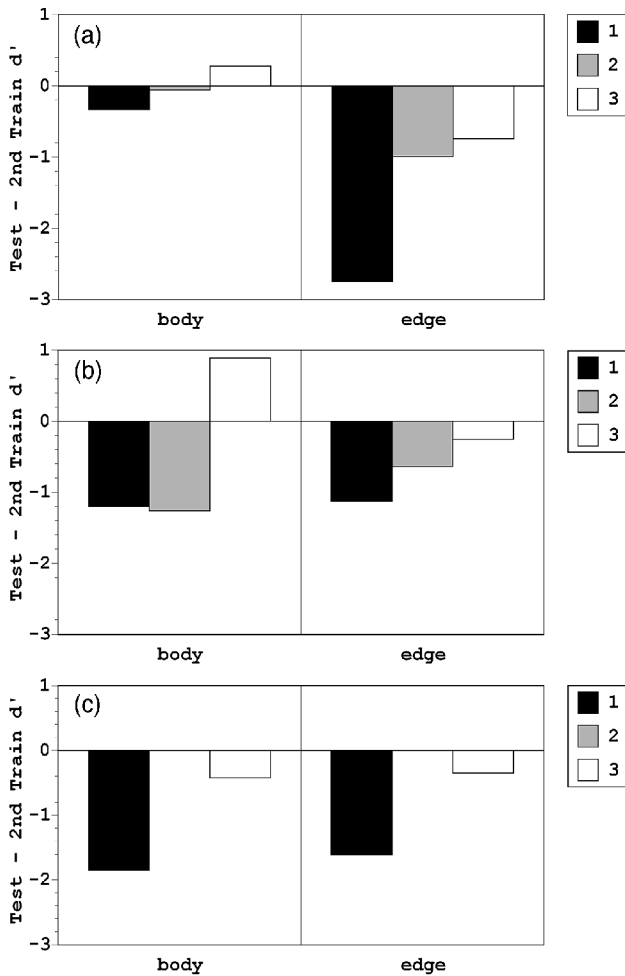


Figure 17

Difference in d' values between Testing and Second Training averaged across listeners trained with the Body versus Edge Rules and with one versus two versus three pairs of vowels for the features (a) [back] (b) [high], and (c) [tense]. Bars of 0 height are given for two pairs of vowels for the feature [tense] because no listeners have been run in that condition

4.4

Discussion

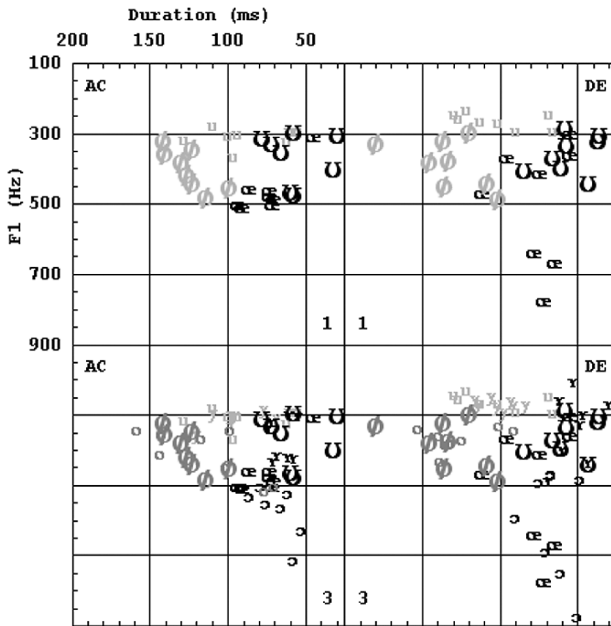
For two of the three features tested, [back] and especially [high], these results support the interpretation that listeners have learned features rather than sets of exemplars. With the Body but not the Edge rule, the new vowels in the Testing stage were categorized more accurately for [back] and [high] relative to the Training vowels when more pairs of training vowels were used. The significantly greater improvement with the Body than the Edge rule disconfirmed the prediction of the selective attention exemplar model that performance would improve equally with the number of training vowels with both rules. No difference in the effect of Number of training pairs was found between the Body and Edge rules for the feature [tense], however. With both Rules, increasing the Number of training vowels improved generalization to new vowels for [tense] by more or less the same amount. This result matches the predictions of the hypothesis that listeners learned to attend selectively to the physical dimensions that distinguished the sets of exemplars corresponding to each natural class

but is contrary to the predictions of the disjunctive exemplar model. For all three features, phonetic variation between vowel phonemes is different from speaker and context variation; it helps while they hurt. These results also show that increasing the new vowels' neighborhood densities by adding more pairs of training vowels facilitated classifying the new vowels with the old ones by increasing the number of similar sounding members in each class (Vitevitch & Luce, 1999).

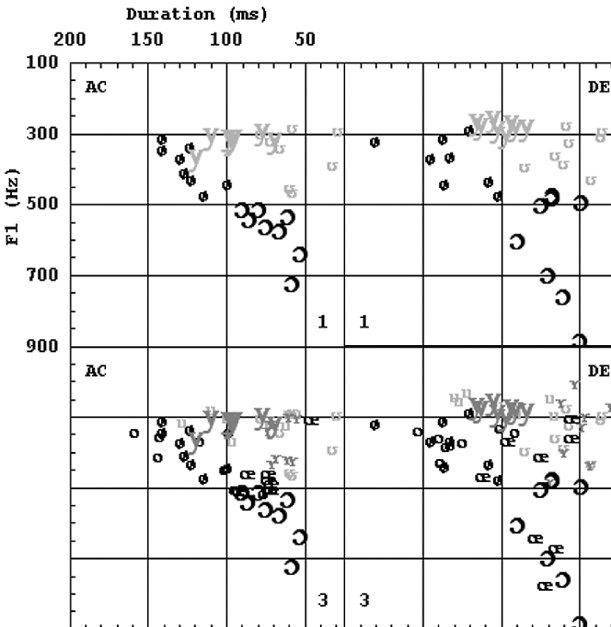
Carsten Eulitz (p.c.) suggests that [tense] is learned differently than [back] or [high] because it is principally a difference in quantity rather than quality: [+tense] vowels being long or consisting of two moras while [-tense] vowels are short or consist of just one mora. Alternatively, the difference is between vowels in open syllables, which have been called "tense" here, versus vowels in closed syllables, which have been called "lax" here. Either way, the contrast is by hypothesis in prosodic structure rather than distinctive feature values, and listeners may therefore have no way of abstracting natural classes from objects that differ prosodically. American English listeners might also find it difficult to map this contrast onto a difference in vowel quality because German tense versus lax vowels do not differ in the extent and direction of diphthongization in the way that American English tense versus lax vowels do.

This explanation must overcome two complementary problems. First, like American English lax vowels, German lax vowels are pronounced with lower and more centralized tongue positions and thus with higher F1 and less extreme F2 values than their tense counterparts (Fischer-Jørgensen, 1990, and other studies cited by her; Bohn & Polka, 2001; Strange & Bohn, 1998). The values for these formants in Tables A1–5 in the *Appendix* and those displayed in Figure 18 (overleaf) shows that the lax vowels used in this experiment also had higher F1 values than their tense counterparts. Second, the contrast for the feature [high] is probably also conveyed in part by differences in duration, the mid vowel stimuli being systematically longer than high vowel stimuli (Fig. 19, overleaf). Thus, lax vowels differ in spectra as well as duration from tense vowels, and high vowels differ in duration as well as in spectra from mid vowels. These facts indicate that some other explanation must be sought for the different behavior of [tense].

Irrelevant variation also affected listeners' categorization of vowels for [tense] but not [high] in the experiments reported in the preceding section. There, listeners' learning of [tense] but not [high] was impaired by irrelevant variation in consonantal context and speaker. Here, irrelevant variation in the other feature values of the vowels in the natural classes improved learning of [tense] but not [high]: generalization improved with Number of training vowels for the Edge as well as the Body rule for [tense] but only with the Body rule for [high]. In the first set of experiments, these two features also behaved differently: listeners discriminated vowels for [high] equally well or poorly regardless of what their other distinctive feature values were, but they discriminated vowels for [tense] differently depending on what those other values were. The results of these three sets of experiments suggest for [tense] that phonologically irrelevant dimensions should be distinguished from phonetically irrelevant ones for this feature. Irrelevant variation in other distinctive features does teach the learner that the relevant feature is [tense] and not one of the other features,

**Figure 18**

F1 and duration values of training and testing stimuli to be categorized as [+tense] (gray) versus [-tense] (black) with the Body rule for the AC (left) and DE (right) Speaker-Context combinations. The two vowels used in One pair training are displayed together with the testing vowels at the top; the six vowels used in Three pair training at the bottom. The superimposed symbols for the testing vowels are larger than those for the training vowels

**Figure 19**

F1 and duration values of training and testing stimuli to be categorized as [+high] (gray) versus [-high] (black) with the Body rule for the AC (left) and DE (right) Speaker-Context combinations. The two vowels used in One pair training are displayed at the top; the six vowels used in Three pair training at the bottom. The superimposed symbols for the testing vowels are larger than those for the training vowels

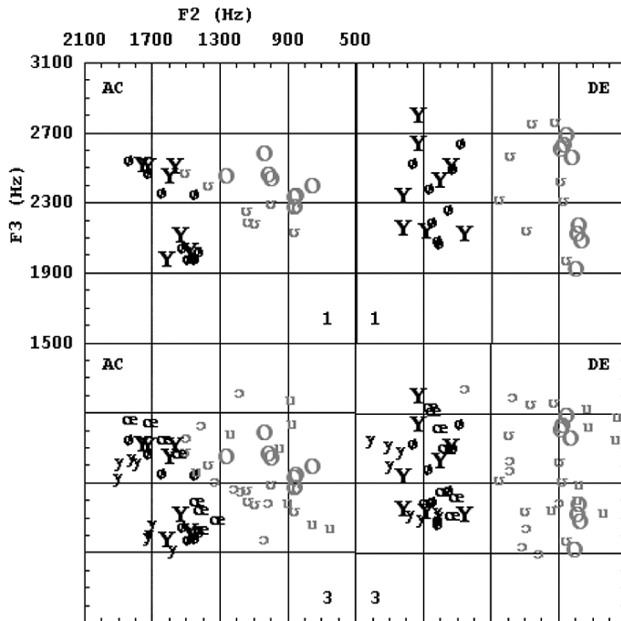
but irrelevant variation in consonantal context or speaker does not. In terms of Nosofsky's model, listeners learn to attend selectively to [tense] rather than differences in other distinctive features, but they do not learn to attend selectively to phonological as opposed to phonetic differences. For [high], the results of these

three sets of experiments instead converge on the conclusion that listeners could extract this feature's value from the signal regardless of irrelevant phonetic variation, once phonological uncertainty is removed that it was the defining feature. These interpretations must be considered tentative until further testing is done; moreover, they leave open why [tense] and [high] behave differently in this way.

One other difficulty also remains unresolved. The first set of experiments showed that listeners discriminated high vowels for [back] less well than mid vowels. Here, we observed that listeners generalize better to novel vowels contrasting for [back] when trained with more vowels chosen according to the Body but not the Edge Rule. In this set of experiments, [back] is learned like [high] but in the first set of experiments it was learned like [tense]. Unfortunately, learning of [back] contrasts was not tested in the second set of experiments, so we have no idea of how irrelevant context or speaker variation affects its learning.

5 Summary and Conclusions

The first set of experiments reported here shows that pairs of vowels contrasting minimally for the same feature in German often will not assimilate in the same way to English vowels. The front:back contrast assimilates differently for high than mid vowels, as does the tense:lax contrast for mid than high front rounded vowels. The high:mid contrast, on the other hand, assimilates in much the same way for tense as lax, rounded as unrounded, and back as front vowels, and the rounded:unrounded contrast assimilates in the same way for high as mid vowels. All the cases where parallel contrasts fail to assimilate disconfirm the predictions of PAM or SLM equally. They are, however, compatible with an alternative version in which individual vowels assimilate to the phonetically most similar categories in the listeners' native language, what Best calls "category recognition." The second set of experiments also supported the conclusion that listeners learned concrete or phonetic rather than phonological categories for tense versus lax vowels in that their ability to generalize to novel tokens was impaired by too much irrelevant variation in the training stimuli, particularly in the consonantal context flanking the target vowels, but also in the speaker who pronounced them. For the feature [high], however, limiting the extent of this irrelevant variation did not increase the extent to which listeners could generalize their training to novel vowels contrasting for this feature. At that stage in the paper, the evidence from the [tense] contrast pointed to listeners having learned the categories of German vowels as individual sets of exemplars. The evidence from the [high] contrast did not, and instead pointed to listeners' having separated that distinctive feature from other features and from contextual and speaker variation. The point of the last set of experiments was to test the competing models of category learning indicated by these conflicting results, the accumulation of exemplars versus extraction of abstract distinctive feature values. For [back] and [high], the results of that third set of experiments support the interpretation that listeners could in fact learn these features, while for [tense], the results showed instead that they could again have learned a set of exemplars. Even for [tense], however, listeners learned to attend selectively to the physical dimensions that are relevant for classifying vowels as tense versus lax, and did not just cobble together disjunctive sets of tense and lax vowels.

**Figure 20**

F2 and F3 values of training and testing stimuli to be categorized as [+back] (gray) versus [-back] (black) with the Body rule for the AC (left) and DE (right) Speaker-Context combinations. The two vowels used in One pair training are displayed at the top; the six vowels used in Three pair training at the bottom. The superimposed symbols for the testing vowels are larger than those for the training vowels.

Given this array of results, what characterizes the categories listeners learned in these experiments? The first part of the answer is that it obviously depends on what contrast one is talking about. The evidence consistently shows that listeners learned a quite different kind of category for [high] than [tense]; this evidence is instead equivocal and incomplete for [back]. That is, listeners can extract abstract values for the features [high] and perhaps [back] but not [tense], for which they instead learn sets of exemplars. Perhaps, listeners responded differently to the three contrasts because the stimuli gave them high quality information for [high] but low quality information for [tense], and intermediate quality information for [back].

However, the displays of the acoustic differences between the stimuli used in the third set of experiments suggest instead that equally good information was available for all three contrasts. Figures 18 and 19, together with the similar display in Figure 20, all show that the categories could be distinguished by deterministic, perhaps even linear rules (Ashby & Gott, 1988). Vowels with longer durations and lower F1 values would be categorized as [+tense], those with shorter durations and higher F1 values as [-tense] (Fig. 18). Vowels with shorter durations and lower F1 values would be categorized as [+high], those with longer durations and higher F1 values as [-high] (Fig. 19). Finally, vowels with lower F2 values would be categorized as [+back], those with higher F2 values as [-back] (Fig. 20). The success with which the natural classes defined by all three contrasts can be separated by these deterministic rules is surprising given how differently they have behaved in the three sets of experiments reported here.

An alternative and perhaps more successful explanation is that the categories defined by these features are polymorphous in the sense of Kluender (1994), and that these three contrasts differ in how polymorphous they are. Each individual vowel category differs in more than one phonetic property from every category it contrasts

minimally with, and the phonetic differences that distinguish it from some minimally contrasting category are not entirely the same as those that distinguish another pair of vowels that contrast minimally in the same way. In other words, all members of a natural class defined by a distinctive feature value will not differ phonetically in all the same properties from all the members of the minimally contrasting natural class. I suggest that listeners differ in the extent to which they can extract abstract features for these contrasts because the categories defined by [high] are the least polymorphous, those defined by [tense] are the most, and those defined by [back] are in between. Displays like those in Figures 18–20 are necessarily incomplete in that they will not show the full extent to which any of these categories is polymorphous. Despite their polymorphous character, natural classes defined by distinctive feature values are or can become perceptually coherent to listeners, even if they are formed from foreign speech sounds, because listeners can eventually learn deterministic rules to discriminate the members of minimally contrasting natural classes. They are simply slower to do so when the categories are more polymorphous.

One piece of evidence from the first set of experiments challenges this interpretation. Differences in listeners' ability to discriminate different instances of a contrast, for example, their far better ability to discriminate the mid vowels, / θ / versus / α /, for [tense] than the high vowels, /y/ versus / γ /, persist undiminished across multiple days of training. This persistence could simply indicate a defect in the training procedures, but it also suggests that the natural classes defined by some features remain polymorphous, and the rules distinguishing them remain less deterministic. In short, foreign categories are and remain perceptually unequal.

These results still challenge PAM and SLM as well as exemplar models of category learning. These challenges indicate that it is time to start studying how infants learn the kinds of categories that generalize across speakers, contexts, and ultimately across allophonic differences that are contrastive in other languages, attract foreign sounds to themselves, and define natural classes. Testing infants' ability to discriminate minimally contrasting sounds that do not vary irrelevantly will not inform us about how categories are learned that are polymorphous in the way phonemic categories are. The second and third sets of experiments presented here are intended to be models of what kind of experiments need to be done, although neither's design would work unchanged with infants. Moreover, infants' learning of different contrasts within the large classes of vowels and consonants needs to be carefully compared, to determine whether they, like the adults in these experiments, learn to abstract features more readily for some contrasts than others.

References

- ASHBY, F. G., & GOTT, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *J. Experimental Psychology: Learning, Memory, and Cognition*, **14**, 33–53.
- ASLIN, R. N., JUSCZYK, P. W., & PISONI, D. B. (1998). Speech and auditory processing during infancy: Constraints on and precursors to language. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology: Cognition, perception, and language* (pp. 147–254). New York: Wiley.
- BEST, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation hypothesis. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception* (pp. 167–224). Cambridge, MA: MIT Press.
- BEST, C. T., McROBERTS, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech*, **46**, 183–216.
- BEST, C. T., McROBERTS, G. W., & GOODELL, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *J. Acoustical Society of America*, **109**, 775–794.
- BEST, C. T., McROBERTS, G. W., LAFLEUR, R., & SILVER-ISENSTADT, J. (1995). Divergent developmental patterns for infants' perception of two non-native speech contrasts. *Infant Behavior and Development*, **18**, 339–350.
- BEST, C. T., & STRANGE, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, **20**, 305–330.
- BOHN O. -S & POLKA, L. (2001). Target spectral, dynamic spectral, and duration cues in infant perception of German vowels. *J. Acoustical Society of America*, **110**, 504–515.
- BROWMAN, C. P., & GOLDSTEIN, L. M. (1986). Towards an articulatory phonology. *Phonology Yearbook*, **3**, 219–252.
- BROWMAN, C. P., & GOLDSTEIN, L. M. (1989). Articulatory gestures as phonological units. *Phonology*, **6**, 201–251.
- BROWMAN, C. P., & GOLDSTEIN, L. M. (1990a). Gestural specification using dynamically-defined articulatory structures. *Journal of Phonetics*, **18**, 299–320.
- BROWMAN, C. P., & GOLDSTEIN, L. M. (1990b). Representation and reality: Physical systems and phonological structure. *Journal of Phonetics*, **18**, 411–424.
- BROWMAN, C. P., & GOLDSTEIN, L. M. (1990c). Tiers in articulatory phonology, with some implications for casual speech. In J. Kingston & M. E. Beckman (Eds.), *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech* (pp. 341–376). Cambridge University Press.
- BROWMAN, C. P., & GOLDSTEIN, L. M. (1992). Articulatory phonology: An overview. *Phonetica*, **49**, 155–180.
- BYBEE, J. (2001). *Phonology and language use*. Cambridge, U.K.: Cambridge University Press.
- De JONG, K. (1995). On the status of redundant features. In B. Connell & A. Arvaniti (Eds.), *Papers in laboratory phonology IV* (pp. 68–86). Cambridge, U.K.: Cambridge University Press.
- DUPOUX, E., PALLIER, C., SEBASTIAN, N., & MEHLER, J. A. (1997). A destressing deafness in French? *J. Memory and Language* **36**, 406–421.
- DUPOUX, E., PEPERKAMP, S., & SEBASTIAN GALLES, N. (2001). A robust method to study stress “deafness.” *J. Acoustical Society of America*, **110**, 1606–1618.
- FISCHER-JØRGENSEN, E. (1990). Intrinsic F0 in tense and lax vowels with special reference to German. *Phonetica*, **47**, 99–140.
- FLEGE, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In C. Ferguson & T. Huebner (Eds.), *Crosscurrents in second language acquisition and linguistic theories*, Philadelphia, PA: John Benjamins.

- GOLDINGER, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *J. Experimental Psychology, Learning, Memory, and Cognition*, **22**, 1166–1183.
- GOLDINGER, S. D. (2000). The role of perceptual episodes in lexical processing. In A. Cutler, J. M. McQueen, & R. Zondervan (Eds.), *Proceedings of the Workshop on Spoken Word Access Processes (SWAP)* (pp. 155–159). Nijmegen: Max Plack Institute for Psycholinguistics.
- GOLDINGER, S. D., PISONI, D. B., & LOGAN, J. S. (1991). On the locus of talker variability effects in recall of spoken word lists. *J. Experimental Psychology, Learning, Memory, and Cognition*, **17**, 152–162.
- GUENTHER, F. H. (2000). An analytical error invalidates the “depolarization” of the perceptual magnet effect. *J. Acoustical Society of America*, **107**, 3576–3577.
- GUION, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals: An investigation into age of acquisition effects on the mutual influence of the first and second languages. *Phonetica*, **60**, 98–128.
- HILLENBRAND, J., GETTY, L. A., CLARK, M. J., & WHEELER, K. (1995). Acoustic characteristics of American English vowels. *J. Acoustical Society of America*, **97**, 3099–3111.
- IVERSON, P., & KUHL, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *J. Acoustical Society of America*, **97**, 553–562.
- IVERSON, P., & KUHL, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics*, **62**, 874–886.
- JAKOBSON, R., FANT, G., & HALLE, M. (1953). *Preliminaries to speech analysis*, Cambridge, MA: MIT Press.
- JOHNSON, K. J. (1997). Speech perception without speaker normalization. In K. J. Johnson., & J. W. Mullenix (Eds.), *Talker variability in speech processing*, (pp. 145–165). San Diego: Academic Press.
- KINGSTON, J., BARTELS, C., RICE, J., BENKÍ, J. R., MOORE, D., THORBURN, R., & MACMILLAN, N. A. (1996). Learning non-native vowel categories. In T. Bunnell & W. Idsardi (Eds.), *ICSLP 96: Proceedings Fourth International Conference Spoken Language Processing* (pp. 1553–1556). Philadelphia.
- KINGSTON, J., & MORETON, E. (1998). The discovery of natural classes by non-native listeners. *J. Acoustical Society of America*, **103**, 2986.
- KINGSTON, J., & MORETON, E. (2001). Do listeners learn foreign vowel categories as disjunctive sets, through selection attention, or as prototypes?, *Proceedings Workshop Speech Recognition as Pattern Classification (SPRAAC)* (pp. 31–36). Max Planck Institute for Psycholinguistics, Nijmegen.
- KLUENDER, K. R. (1994). Speech perception as a tractable problem in cognitive science. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 173–217). San Diego: Academic Press.
- KUHL, P. K. (1991). Human adults and human infants show a “perceptual magnet effect” for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, **50**, 93–107.
- KUHL, P. K. (1993). Innate predispositions and the effects of experience in speech perception: The native language magnet theory. In B. de-Boysson-Bardies, S. de Schonen, P. Juszczyk, P. McNeilage, & J. Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life* (pp. 259–274). Dordrecht, Netherlands: Kluwer Academic Publishers.
- KUHL, P. K. (2000). Language, mind, and brain: Experience alters perception. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 99–115). Cambridge, MA: MIT Press.

- KUHL, P. K., WILLIAMS, K. A., LACERDA, F., STEVENS, K. N., & LINDBLOM, B. (1992). Linguistic experience alters phonetic perception in infants by six months of age. *Science*, **255**, 606–608.
- LIVELY, S. E., LOGAN, J. S., & PISONI, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new speech categories. *J. Acoustical Society of America*, **94**, 1242–1255.
- LIVELY, S. E., LOGAN, J. S., PISONI, D. B., YAMADA, R. A., TOHKURA, T., & YAMADA, T. (1994). Training listeners to identify English /r/ and /l/: III. Long-term retention of new phonetic categories. *J. Acoustical Society of America*, **96**, 2076–2087.
- LOGAN, J. S., LIVELY, S. E., & PISONI, D. B. (1991). Training Japanese listeners to identify /r/ and /l/: A first report. *J. Acoustical Society of America*, **89**, 874–886.
- LOTTO, A. J. (2000). Reply to “An analytical error invalidates the ‘depolarization’ of the perceptual magnet effect.” *J. Acoustical Society of America*, **107**, 3578–3580.
- LOTTO, A. J., KLUENDER, K. R., & HOLT, L. L. (1998). The perceptual magnet effect depolarized. *J. Acoustical Society of America*, **103**, 3648–3655.
- MACMILLAN, N. A., & CREELMAN, C. D. (1991). *Detection theory: A user's guide*, Cambridge, U.K.: Cambridge University Press.
- MADDIESON, I., & PRECODA, K. (1992). UCLA phonological segment inventory database. Los Angeles, CA: UCLA Phonetics Laboratory.
- MARCUS, G. F., VIJAYAN, S., BANDI RAO, S., VISHTON, P. M. (1999). Rule learning by seven-month-old infants. *Science*, **283**, 77–80.
- MAYE, J., WERKER, J. F., & GERKEN, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, **82**, B101–111.
- MULLENNIX, J. W., & PISONI, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics*, **47**, 379–390.
- MULLENNIX, J. W., PISONI, D. B., & MARTIN, C. S. (1989). Some effects of talker variability on spoken word recognition. *J. Acoustical Society of America*, **85**, 365–378.
- NOSOFSKY, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *J. Experimental Psychology: General*, **115**, 39–57.
- PALLIER, C., BOSCH, L., SEBASTIAN-GALLÉS, N. (1997). A limit on behavioral plasticity in speech perception. *Cognition*, **64**(3), B9–B17.
- PALLIER, C., COLOMÉ, A., SEBASTIAN-GALLÉS, N. (2001). The influence of native-language phonology on lexical access: Concrete exemplar-based vs. abstract lexical entries. *Psychological Science*, **12**, 445–449.
- PEPERKAMP, S., & DUPOUX, E. (2002). A typological study of stress “deafness.” In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 203–240). Berlin: Mouton de Gruyter.
- PETERSON, G. E., & BARNEY, H. L. (1952). Control methods used in the study of vowels. *J. Acoustical Society of America*, **24**, 175–184.
- PIERREHUMBERT, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology 7* (pp. 101–139). Berlin: Mouton de Gruyter.
- PISONI, D. B., LIVELY, S. E., & LOGAN, J. S. (1994). Perceptual learning of non-native speech contrasts: Implications for theories of speech perception. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception* (pp. 121–166). Cambridge, MA: MIT Press.
- POLKA, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *J. Acoustical Society of America*, **89**, 2961–2977.
- POLKA, L. (1992). Characterizing the influence of native experience on adult speech perception. *Perception & Psychophysics*, **52**, 37–52.
- POLKA, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *J. Acoustical Society of America*, **97**, 1286–1296.

- POLKA, L., & BOHN, O. -S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *J. Acoustical Society of America*, **100**, 577–592.
- POLKA, L., & WERKER, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *J. Experimental Psychology, Human Perception, and Performance*, **20**, 421–435.
- SAFFRAN, J. R., ASLIN, R. N., & NEWPORT, E. L. (1996). Statistical learning by eight-month-old infants. *Science*, **274**, 1926–1928.
- STRANGE, W., & BOHN, O. -S. (1998). Dynamic specification of coarticulated German vowels: Perceptual and acoustical studies. *J. Acoustical Society of America*, **104**, 488–504.
- STRANGE, W., & DITTMAN, S. (1984). Effects of discrimination training on the perception of /r-/ by Japanese adults learning English. *Perception & Psychophysics*, **36**, 131–145.
- TRUBETZKOY, N. S. (1969). *Grundzüge der Phonologie*, Trans. by C. Baltaxe, *Principles of Phonology*, Los Angeles: University of California Press. (Original work published 1939)
- VITEVITCH, M. S., & LUCE, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *J. Memory and Language*, **40**, 374–408.
- WERKER, J. F., & LALONDE, C. E. (1989). Cross-language speech perception: Initial capabilities and developmental change. *Developmental Psychology*, **24**, 672–683.
- WERKER, J. F., & TEES, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization. *Infant Behavior and Development*, **7**, 49–63.

Appendix

For each speaker, means (standard deviations) across all consonantal contexts of fundamental frequency (F0), formant frequencies (F1–F4), and formant bandwidths (B1–B4) during a 30 ms interval (= three 10 ms frames) at the center of each vowel. Formant frequencies and bandwidths (in Hz) were obtained with LPC analysis, using 21 poles. F0 was obtained using autocorrelation. Durations (in ms) were measured from waveform and spectrogram displays using conventional acoustic criteria for identifying the vowel onset and offset.

TABLE A1: SPEAKER A

Vowel	F0	F1	F2	F3	F4	B1	B2	B3	B4	Dur
i	242 (20)	290 (22)	2432 (86)	3188 (88)	3624 (111)	78 (24)	172 (75)	272 (141)	279 (101)	78 (12)
I	252 (23)	415 (44)	2089 (123)	2785 (163)	3706 (112)	78 (35)	232 (69)	304 (84)	263 (85)	45 (8)
e	234 (23)	443 (36)	2311 (100)	2992 (85)	3576 (118)	64 (24)	192 (67)	320 (79)	379 (188)	101 (13)
ɛ	243 (22)	565 (83)	1954 (103)	2655 (174)	3522 (151)	193 (75)	252 (85)	374 (108)	384 (138)	58 (9)
y	251 (16)	315 (26)	1883 (136)	2434 (55)	3556 (118)	100 (31)	194 (110)	196 (110)	208 (96)	75 (13)
Y	252 (27)	418 (24)	1659 (104)	2489 (106)	3595 (82)	81 (26)	186 (45)	242 (78)	222 (89)	48 (7)
ø	235 (16)	451 (26)	1636 (108)	2429 (73)	3543 (88)	47 (21)	130 (72)	202 (80)	257 (113)	99 (15)
œ	245 (22)	498 (33)	1619 (78)	2565 (164)	3603 (82)	142 (54)	202 (85)	272 (105)	281 (55)	59 (10)
u	251 (23)	332 (44)	1082 (173)	2464 (113)	3566 (105)	86 (33)	155 (69)	186 (87)	254 (121)	78 (15)
ū	252 (24)	435 (54)	1225 (159)	2459 (114)	3528 (118)	95 (40)	127 (53)	227 (75)	241 (78)	44 (12)
o	242 (21)	468 (42)	1043 (123)	2494 (143)	3492 (96)	48 (30)	116 (71)	193 (75)	232 (69)	98 (19)
ɔ	239 (20)	623 (85)	1297 (118)	2543 (187)	3481 (85)	236 (99)	145 (66)	226 (100)	341 (130)	59 (11)

TABLE A2: SPEAKER B

<i>Vowel</i>	<i>F0</i>	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>B1</i>	<i>B2</i>	<i>B3</i>	<i>B4</i>	<i>Dur</i>
i	197 (10)	229 (34)	2276 (468)	3199 (292)	3799 (237)	65 (33)	229 (144)	236 (83)	449 (198)	147 (20)
ɪ	194 (12)	336 (47)	2094 (144)	2796 (110)	3486 (371)	92 (44)	250 (110)	318 (156)	541 (222)	73 (11)
e	178 (5)	328 (48)	2364 (114)	3059 (198)	3862 (339)	113 (42)	278 (85)	299 (174)	511 (204)	175 (21)
ɛ	176 (9)	495 (141)	1922 (301)	2569 (262)	3356 (435)	221 (146)	262 (99)	399 (240)	711 (329)	91 (14)
y	195 (12)	259 (57)	1865 (107)	2305 (106)	3746 (221)	104 (42)	190 (128)	203 (116)	231 (148)	146 (17)
Y	190 (11)	333 (48)	1670 (184)	2277 (131)	3739 (200)	113 (47)	284 (174)	198 (86)	268 (162)	76 (12)
ø	179 (9)	354 (21)	1689 (151)	2234 (142)	3713 (277)	133 (60)	264 (134)	221 (100)	353 (228)	165 (10)
œ	174 (11)	446 (118)	1555 (152)	2329 (132)	3564 (378)	147 (67)	273 (122)	225 (128)	404 (238)	93 (21)
u	193 (7)	287 (37)	852 (169)	2358 (116)	3495 (359)	87 (25)	307 (198)	375 (146)	405 (235)	145 (21)
ū	192 (15)	363 (56)	1129 (197)	2279 (113)	3777 (209)	122 (60)	241 (137)	325 (126)	271 (164)	73 (16)
o	180 (7)	338 (37)	763 (121)	2233 (77)	3658 (315)	128 (39)	161 (56)	411 (134)	419 (240)	166 (20)
ɔ	172 (10)	492 (176)	1133 (109)	2245 (190)	3730 (197)	306 (150)	253 (108)	202 (96)	381 (216)	84 (16)

TABLE A3: SPEAKER C

<i>Vowel</i>	<i>F0</i>	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>B1</i>	<i>B2</i>	<i>B3</i>	<i>B4</i>	<i>Dur</i>
i	110 (14)	282 (27)	1936 (97)	2792 (159)	3276 (105)	79 (34)	133 (117)	182 (52)	210 (79)	109 (11)
ɪ	113 (7)	279 (29)	1903 (101)	2543 (109)	3188 (84)	97 (42)	195 (179)	180 (58)	179 (81)	59 (15)
e	103 (5)	307 (24)	1906 (69)	2579 (88)	3210 (66)	110 (33)	136 (95)	166 (61)	182 (66)	122 (16)
ɛ	106 (4)	477 (103)	1677 (78)	2351 (59)	3195 (110)	376 (186)	180 (51)	152 (38)	206 (111)	79 (10)
y	108 (5)	297 (12)	1657 (89)	2019 (54)	2992 (83)	61 (13)	137 (42)	115 (44)	98 (45)	106 (11)
Y	128 (66)	315 (38)	1531 (73)	2035 (58)	2978 (68)	89 (48)	152 (44)	141 (43)	134 (99)	66 (9)
ø	103 (5)	333 (24)	1448 (43)	1991 (52)	2937 (51)	124 (30)	111 (34)	114 (56)	69 (19)	121 (14)
œ	105 (5)	466 (65)	1386 (42)	2142 (77)	3013 (66)	256 (97)	145 (53)	115 (43)	122 (65)	83 (15)
u	116 (13)	307 (15)	1023 (538)	2326 (306)	3153 (396)	94 (35)	218 (104)	308 (106)	266 (214)	111 (13)
ū	114 (6)	366 (152)	1023 (273)	2226 (180)	3050 (357)	123 (69)	109 (35)	190 (87)	209 (87)	70 (12)
o	104 (4)	340 (25)	793 (82)	2277 (168)	2875 (78)	177 (56)	154 (52)	271 (132)	245 (177)	130 (17)
ɔ	104 (4)	521 (72)	1089 (70)	2203 (95)	2987 (73)	331 (152)	83 (18)	141 (65)	178 (110)	81 (11)

TABLE A4: SPEAKER D

<i>Vowel</i>	<i>F0</i>	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>B1</i>	<i>B2</i>	<i>B3</i>	<i>B4</i>	<i>Dur</i>
i	120 (14)	239 (21)	2083 (103)	3055 (150)	3466 (123)	41 (19)	174 (127)	262 (179)	266 (189)	97 (26)
ɪ	123 (14)	287 (35)	1986 (121)	2620 (132)	3544 (107)	72 (29)	185 (101)	199 (49)	205 (95)	53 (12)
e	106 (9)	308 (18)	2130 (87)	2732 (147)	3475 (221)	48 (19)	205 (131)	229 (62)	260 (156)	135 (33)
ɛ	119 (11)	408 (70)	1879 (87)	2544 (100)	3499 (166)	164 (106)	152 (37)	250 (122)	327 (145)	71 (12)
y	120 (17)	264 (20)	1685 (65)	2130 (53)	3236 (68)	46 (16)	117 (51)	102 (60)	172 (59)	98 (19)
ʏ	126 (16)	287 (38)	1635 (96)	2147 (86)	3235 (135)	71 (35)	114 (86)	169 (88)	164 (64)	56 (14)
ø	115 (11)	318 (27)	1579 (44)	2146 (98)	3218 (67)	63 (22)	71 (26)	163 (132)	134 (30)	122 (34)
œ	116 (11)	390 (38)	1569 (61)	2203 (90)	3283 (178)	130 (76)	92 (47)	174 (120)	292 (564)	71 (12)
u	118 (11)	268 (23)	767 (100)	2209 (77)	3321 (129)	89 (92)	275 (217)	220 (160)	228 (94)	92 (27)
ʊ	126 (17)	308 (24)	1024 (173)	2157 (125)	3311 (109)	96 (26)	136 (71)	205 (79)	201 (76)	56 (12)
o	109 (9)	343 (22)	781 (96)	2119 (69)	3105 (130)	106 (33)	148 (38)	170 (103)	198 (115)	116 (29)
ɔ	111 (11)	474 (49)	1084 (162)	2098 (151)	3336 (147)	203 (133)	138 (80)	188 (133)	241 (127)	71 (14)

TABLE A5: SPEAKER E

<i>Vowel</i>	<i>F0</i>	<i>F1</i>	<i>F2</i>	<i>F3</i>	<i>F4</i>	<i>B1</i>	<i>B2</i>	<i>B3</i>	<i>B4</i>	<i>Dur</i>
i	226 (43)	254 (22)	2479 (200)	3164 (284)	3726 (151)	37 (32)	337 (328)	305 (215)	267 (131)	105 (17)
ɪ	232 (25)	380 (54)	1965 (281)	2752 (143)	3686 (302)	91 (39)	228 (122)	223 (76)	323 (159)	49 (14)
e	220 (23)	406 (29)	2424 (160)	2955 (196)	3690 (274)	70 (46)	280 (206)	196 (55)	315 (148)	131 (16)
ɛ	210 (25)	653 (114)	1900 (264)	2698 (197)	3499 (430)	249 (53)	275 (110)	356 (235)	426 (109)	65 (14)
y	241 (23)	266 (16)	1881 (81)	2463 (67)	3647 (75)	42 (28)	137 (68)	148 (70)	114 (41)	98 (19)
ʏ	233 (37)	387 (53)	1602 (85)	2568 (106)	3632 (166)	157 (142)	151 (70)	237 (60)	334 (253)	57 (12)
ø	218 (26)	429 (78)	1547 (111)	2511 (143)	3634 (118)	81 (59)	151 (105)	254 (115)	195 (148)	123 (26)
œ	227 (21)	662 (119)	1635 (61)	2626 (152)	3760 (160)	248 (78)	159 (64)	286 (111)	383 (163)	71 (19)
u	242 (28)	260 (18)	765 (161)	2660 (104)	3657 (78)	43 (27)	187 (173)	278 (130)	155 (53)	107 (20)
ʊ	227 (35)	392 (40)	1101 (196)	2627 (143)	3625 (111)	133 (42)	178 (66)	206 (99)	224 (120)	57 (13)
o	219 (28)	392 (39)	824 (84)	2635 (120)	3532 (174)	107 (62)	263 (181)	222 (74)	151 (99)	133 (20)
ɔ	221 (22)	707 (116)	1274 (109)	2603 (161)	3812 (117)	298 (93)	213 (75)	211 (80)	337 (205)	67 (14)