Original Paper

Phonetica 2009;66:222–242
DOI: 10.1159/000298584

Received: August 10, 2009
Accepted: February 25, 2010

Vowel Discrimination by English, French and Turkish Speakers: Evidence for an Exemplar-Based Approach to Speech Perception

Marc Ettlinger[a] Keith Johnson[b]

[a]Department of Communication Sciences and Disorders, Northwestern University, Evanston, Ill., and [b]University of California, Berkeley, Calif., USA

Abstract

Recent studies of speech perception have shown that speakers retain significant amounts of the phonetic detail of heard speech providing strong evidence for an exemplar-based model of the representation of speech sounds. This does not preclude the existence of a feature-based model as well; indeed many theories of speech perception advocate a feature- or contrast-based model for the discrimination of speech sounds. In this study, we provide evidence that an exemplar-based model of speech perception makes more accurate predictions for the performance of English, French and Turkish speakers in a vowel discrimination task. Participants were asked to discriminate the four high front vowels of German, which differ by both the rounding and tense/lax contrasts. Crucially, English has only the tense/lax contrast, while Turkish and French have only rounding. The results show that having one of the vowels in one's language facilitates discrimination more than having the featural contrast supporting an exemplar-based model. Furthermore, different discrimination tasks were used showing that the effects of native language on vowel discrimination are mediated by task and by psychoacoustic similarity.

Introduction

Background

A central question in the study of language concerns the mental representation of speech sounds. These representations crucially define how fundamental tasks such as speech production and speech perception operate and recent debate has centered around evidence for two different frameworks: feature-based and exemplar-based representations.

Feature theory is rooted in an Aristotelian theory of categorization and characterizes each speech sound category through a set of distinctive features [Chomsky and

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

0031–8388/09/0664–0222
$26.00/0
Accessible online at:
www.karger.com/pho

Marc Ettlinger
Department of Communication Sciences and Disorders,
Institute of Neuroscience, Northwestern University
2240 Campus Drive, Evanston, IL 60208 (USA)
Tel. +1 847 491 2430, Fax +1 847 491 2429
E-Mail marc@northwestern.edu

Halle, 1968; Fant, 1973; Jakobson et al., 1952]. Distinctive features differentiate a sound from others in the language and are the necessary and sufficient conditions defining the speech sound category. For example, the vowel /i/ may be defined by the features [+high], [–back], [–round] and [+tense] while /ɪ/ can be defined as [+high], [–back], [–round] and [–tense]. So, what differentiates the two vowels is their tenseness. In a feature-based approach to speech perception, relevant features are extracted from an acoustic signal, which is then categorized based on these extracted features [Liberman and Mattingly, 1989; Stevens, 2002]. So, if a listener were to hear an /i/, the perceptual system would extract the relevant cues (e.g. from the formants of the vowel), normalize the formant values relative to each other and the fundamental frequency and then translate this into a category-defining set of either articulatory or acoustic features[1].

An alternate approach, rooted in exemplar-based theories of human categorization [Nosofsky, 1986; Schacter et al., 1978], defines each speech sound as the collection of all of the heard tokens, or examples, of that sound. The representation of a sound is not an abstract category, but the sum total of all tokens of the sound the person has experienced with all of its acoustic, lexical, social and contextual information retained. Exemplars are associated with each other through these linguistic and extralinguistic properties at multiple levels of representation. This fundamentally changes the nature of the basic processes of speech production and perception. Speech is produced by reproducing the acoustic signal of the set of heard tokens that reflect desired contextual information. Similarly, speech perception does not involve the extraction of features; rather each speech sound is compared to the collection of stored exemplars for each category and the sound in question is assigned to the category with the greater collection of tokens most similar to it [Johnson, 1997; Pierrehumbert, 2001].

There is a significant body of evidence suggesting that speakers do retain the detailed phonetic information necessary for an exemplar model [see Goldinger, 1998 for a review]. To cite a couple of key examples: Hintzman et al. [1972] found that listeners could recall the particular voice with which they heard a particular word, and Goldinger [1996] showed that over the course of a day, listeners were better able to recall words that they heard when repeated in the same voice. More specific to speech perception, Norris et al. [2003] showed that the perceptual boundaries of sound categories can be shifted by recent perceptual experience or exemplars.

That speakers make use of detailed information in stored exemplars of words and sounds does not preclude a complementary feature-based theory of perception, however. Both representations can theoretically coexist since the evidence supporting each approach comes from disparate empirical phenomena. Evidence for exemplars generally comes from tasks explicitly involving the recall of detailed information, such as voice priming and speaker normalization, while the original evidence for featural representations comes from the higher-order phonological generalizations of a language [Chomsky and Halle, 1968]. Indeed, many models of speech perception make use of both of these memory representations [Goldinger, 2007; Massaro, 1975; McNeil and Lindig, 1973; McQueen et al., 2006; Poeppel et al., 2008] leaving open the question of which representation is primary in speech perception. Therefore, instead of trying

[1] We are characterizing the 'feature perception' view in terms of the distinctive features used in linguistic theory. Importantly though, the hypothesis that we test in this article is also relevant for gesture perception theories [e.g. Browman and Goldstein, 1986, 1992; Goldstein and Fowler, 2003]. Phonological feature theory is more explicit regarding the phonetic properties of vowels than is gesture theory, and therefore we frame our hypothesis in terms of these features.

to answer the question of whether the mental representation of speech sounds and the mental lexicon are exclusively exemplar- or feature-based, this article explores which model makes more accurate predictions for a particular task; in this case, vowel discrimination and non-native speech perception.

The Perceptual Assimilation Model [PAM; Best, 1995] is an example of a feature-based model for sound discrimination for non-native speech sounds since it makes use of native sound contrasts in what is referred to as two-category assimilation. In the PAM model, two novel sounds that differ by the same featural contrast as two native sounds may qualify as two-category assimilation and should be discriminated most easily. Evidence for this comes from the fact that English speakers easily discriminated a featural contrast between voiced and voiceless lateral fricatives (/ɮ/~/ɬ/) as compared to plosive vs. implosive voiced bilabial stops (/b/~/ɓ/) presumably because English makes use of a voicing contrast to differentiate fricatives (e.g. /z/~/s/, /ʒ/~/ʃ/) but does not have a plosive/implosive contrast [Best et al., 2001]. It is also worth noting that Best et al. [2001] have focused primarily on discriminating consonants, which may have more discrete, noncontinuous mental representations as compared to vowels. This contrast-based approach to discrimination is articulated more overtly in Flege's [1992] Speech Learning Model, which explicitly predicts that a phonetic difference that distinguishes contrasting foreign sounds but does not contrast native sounds will be poorly discriminated. PAM does not make a priori predictions as to when two-category assimilation and featural contrast will apply, as opposed to category goodness discrimination, when two non-native sounds are compared to the same native sound. As such, PAM is discussed here in the context of having the most explicit implementation of a feature-based theory of non-native sound discrimination in two-category assimilation. The remaining aspects of PAM are not discussed.

These models make predictions as to how one's native language influences the discrimination of two non-native sounds. What is crucial in the discrimination of non-native speech sounds in these models is whether one's native language makes use of the same gestural contrast or distinctive feature. If the native language makes use of a particular featural contrast then native language experience should facilitate discrimination between sounds that differ according to the same featural contrast. On the other hand, if the native language does not make use of a particular featural contrast, discriminating non-native sounds along this featural dimension should be more difficult.

Models of sound discrimination using an exemplar-based theory of sound categories make a different set of predictions. Instead of featural contrasts being crucial, in an exemplar model experience with a sound is the most important factor in perception and discrimination. For example, a set of studies exploring Japanese listeners' ability to discriminate the notorious /l/~/r/ contrast in English showed that training and exposure to these sounds improved subjects' discrimination of the pair [Lively et al., 1994]. Discrimination in certain positions was learned better than in others and this difference persisted for 3–6 months after training. This supports an exemplar-based theory of discrimination since the ability to discriminate these sounds depended on context, which is retained in exemplar models, but is eliminated in abstractionist feature-based models. Complementing these studies is the finding that speakers' voice matters in discriminating and identifying speech as well. Pisoni [1992] asked participants to identify words spoken by either the same voice throughout for one group of participants or by 15 different voices for the other group. Participants were faster and more accurate when only a single voice was used, suggesting that voice is also involved in speech perception, as suggested by exemplar theory, and is not abstracted away, as suggested by feature-based approaches. Exemplar theory makes the more general prediction that experience with a particular sound in its myriad contexts is what facilitates discrimination more than the featural contrasts in one's language.

Finally, the results in Best et al. [2001], which are presented as evidence for a feature-based facilitation of discrimination, are also interpretable in exemplar theory. The greater discriminability of /ɮ/~/ɬ/ over /b/~/ɓ/, argued to be based on the existing voicing contrast in English, may be due to psychoacoustic differences between the pairs, which are not controlled. Therefore, we cannot assess whether this greater discriminability is due to native language effects or basic psychoacoustic properties of the stimuli, and so the results are compatible with either approach. Furthermore, [ɓ] may be interpreted as an allophone of /b/ [Purnell, pers. comment], impairing discrimination.

Present Study

This article reports on a set of experiments testing these two theories of speech perception by comparing performance on the discrimination of novel or foreign vowel sounds by listeners with different native language experience.

English has two high front vowels, /i/ and /ɪ/. These two sounds are differentiated by the feature [±tense] with /i/ defined as tense and /ɪ/ lax [Jakobson et al., 1952, and in many subsequent updates to feature theory]. Feature theory suggests that knowledge of English should facilitate the discrimination of non-native sound pairs that contrast along this same dimension. So, even though English has neither /y/ nor /ʏ/, the rounded equivalents of /i/ and /ɪ/, in its inventory, these two sounds are tense and lax, respectively, and should therefore be discriminated relatively easily, presumably because English listeners are attuned to the spectral and temporal difference associated with the tense-lax distinction. On the other hand, /ɪ/ and /ʏ/ are differentiated by the feature [±round], which is a featural contrast not found for English front vowels nor is it a distinctive feature in the sense that no two vowels of English are minimally differentiated by just the feature [±round].[2] Therefore knowledge of English should not facilitate the discrimination of this pair of sounds according to feature theory. Expressed in terms of PAM, /y/~/ʏ/ would be two-category assimilation with /y/ assimilating to /i/ or /u/ and /ʏ/ assimilating to /ɪ/ or /ʊ/, and therefore most easily discriminable, while /ɪ/~/ʏ/ would be a category goodness assessment and therefore not as discriminable.

Exemplar theory makes the opposite prediction. The existence of /ɪ/ in the inventory of English should facilitate discrimination of the /ɪ/~/ʏ/ vowel pair, while the lack of both /y/ and /ʏ/ vowels in English should make discrimination of this pair relatively difficult.

The discriminability of these two sound pairs cannot simply be compared to each other, however. As with the comparison in Best et al. [2001] discussed above, the psychoacoustic difference between these sounds may be different, which also plays a role. As one might expect and as the results below show the acoustic difference between

[2] There are round vowels in English but the specification of [round] is argued to be redundant and not distinctive within the phonology of the language [Archangeli, 1994] because there are no front unround vowels with a round counterpart. In addition to being predictable based on other qualities of the vowel, rounding is also a secondary cue as compared to openness and backness in vowel perception [Bohn, 1995; Lang and Ohala, 1996; Singh and Woods, 1971].

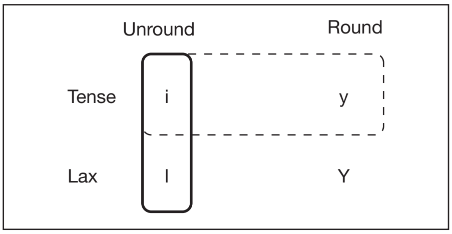

Fig. 1. Inventory of high front vowels in English (solid) and French and Turkish (dotted).

/y/ and /ʏ/ is greater than the difference between /ɪ/ and /ʏ/. So, /y/~/ʏ/ should be more easily discriminated than /ɪ/~/ʏ/ independent of whatever theory of representation one adopts and independent of native language. A direct comparison would therefore not be conclusive; psychoacoustic differences must be controlled.

Psychoacoustic differences can be controlled in a cross-linguistic comparison by observing the effect that different native language inventories have on the discrimination of the same vowel stimuli. Indeed, juxtaposing the high front vowels of French and Turkish to those of English, we find a useful distinction. French and Turkish have /i/ and /y/ in their phonemic inventory which differs by the feature [±round] (/i/ is unrounded, /y/ is rounded), but no /ɪ/. So, the vowel inventories of these two sets of languages are different (English /ɪ/ vs. Turkish and French /y/), but also overlap (/i/); none of these languages has /ʏ/. Similarly, the featural contrasts used in each language are different: English makes use of the tense/lax distinction but not the rounded/unrounded contrast while French and Turkish make use of the rounded/unrounded distinction, but not the tense/lax contrast. These differences are schematized in figure 1.

According to a contrast-based theory, French and Turkish speakers should be better at discriminating /ɪ/~/ʏ/ than English speakers because those two languages make use of the feature [±round] while English speakers should be better at discrimination /y/~/ʏ/ because English makes use of the feature [±tense]. Again, an inventory-based theory makes the opposite prediction: English speakers should be better at discriminating /ɪ/~/ʏ/ vowel pairs because English has /ɪ/ in its inventory and Turkish and French do not, while Turkish and French speakers should be better at discriminating /y/~/ʏ/ vowel pairs because Turkish and French have /y/ in their inventory and English does not. Since the vowel pairs being assessed are the same, psychoacoustic differences are controlled. These predictions are summarized in tables 1 and 2.

To test these predictions we conducted a set of different discrimination tasks with English, Turkish and French speakers. The first task was a speeded fixed AX discrimination task, the second was a similarity-rating task, and the third was a roving AX discrimination task. While the two crucial vowel pairs for testing the two hypotheses are the /ɪ/~/ʏ/ and /y/~/ʏ/ pairs, all six were used (/i/~/y/, /i/~/ɪ/, /i/~/ʏ/, /y/~/ɪ/). Each participant took all three tests in one session separated by time for directions for each test. The order was counterbalanced across participants for each language.[3] At the

[3] Despite the potential for a training effect based on the order of tasks, there was no significant difference on performance on each task based on the order it was presented [linear mixed model with order as fixed effect and subject as random effect: RT: t(47) = 0.52, p = 0.60, rating: t(47) = 0.41, p = 0.68; logistic lmm for % corr: z(47) = 0.76, p = 0.45].

Table 1. Predictions made by feature theory and exemplar theory on the discrimination of different vowel pairs for English, Turkish and French

	/ɪ/~/ʏ/	/y/~/ʏ/
Feature theory	T, F < E	E < T, F
Exemplar theory	E < T, F	T, F < E

< indicates 'is better able to discriminate', e.g. T, F < E indicates Turkish and French speakers are better able to discriminate this than English speakers.

Table 2. Predictions made by feature theory and exemplar theory on the discrimination of different vowel pairs for English, Turkish and French

Feature theory		Exemplar theory	
English	Turkish/French	English	Turkish/French
y	y	y	y
↕			↕
ɪ ʏ	ɪ↔ʏ	ɪ↔ʏ	ɪ ʏ

Arrows indicate improved discriminability.

conclusion of all three tasks, the individuals completed a questionnaire regarding their demographic and linguistic backgrounds.

Experiment 1: Speeded Discrimination

We began by exploring the psychoacoustic, hypothetically language-independent, perception of these vowels using a speeded fixed AX discrimination task. Previous research has suggested that a speeded task has the potential to eliminate native-language effects on perception as compared to a slower discrimination task or a rating or categorization task [Fox, 1984; Pisoni and Sawusch, 1975]. Instead, it should reflect the psychoacoustic differences among these speech sounds providing a baseline measure of discriminability. If these results show no influence of language, it would further support the practice of using a speeded fixed discrimination task to get at the psychoacoustic difference between speech sounds. It is also possible; however, that language does influence performance on this task.

Method

Participants

There were 49 participants: 16 native English speakers (7 male; 9 female) with an average age of 20.6, 16 native French speakers (4 male; 12 female) with an average age of 31.1, and 17 native Turkish speakers (8 male; 9 female) with an average age of 25.6. None of the participants came to the US prior to

Table 3. Vowel formants for the four stimulus vowels (in Hz)

	i	ɪ	y	ʏ
F0	200	200	200	200
F1	300	350	280	360
F2	2,560	2,250	2,220	1,750
F3	3,400	3,200	2,530	2,570

the age of 18; the French speakers had been in the US an average of 5.1 years (range: 0–13 years) years and the Turkish speakers had been in the US an average of 2.8 years (range: 0–6 years). One Turkish participant's results were excluded because of an apparent miscommunication of instructions. Participants were recruited at University of California, Berkeley and included undergraduate and graduate students as well as members of the local Berkeley, California, community. Participants were given USD 10 for approximately 1 h of their time and none reported any history of speech or hearing problems.

Stimuli

The stimuli consisted of four vowel stimuli, [i], [ɪ], [y] and [ʏ], produced by a female native speaker of German from Hamburg, who spoke a Norddeutsche (or Hamburger) dialect. German was used for two reasons. First, neither French, nor English, nor Turkish have all four test vowels in their inventory (by design) yet we wanted to use a single speaker so that voice could not serve as a cue for discrimination. German does have all four. Second, the acoustic correlate of the tense/lax distinction in German is primarily spectral in nature (table 3) [Fischer-Jørgensen and Jørgensen, 1969; Hall, 2003; Jessen, 1998], as opposed to being primarily distinguished by vowel length, as is the case in a language like Dutch. German makes use of similar spectral and length cues as English for the tense/lax contrast (except German /i/ is slightly more peripheral) [Hall, 2003] and length can be normalized and the tense/lax contrast can still be discriminated by German or English listeners [Condax and Krones, 1976].

The German speaker was recorded saying minimally different words containing the four vowels in the carrier sentence 'Der erste Vokal im Wort ___ wird ausgesprochen als ___' (*The first vowel in the word ___ is pronounced ___*) several times. The words used were *bieten* [i], *bitten* [ɪ], *hüten* [y] and *Hütten* [ʏ]. The steady-state portion of the vowels was extracted from the full mention of the near-minimal pair words. The onset of the segmented portion was based on the termination of any formant transition from the initial consonant and the offset was when the amplitude of the vowel waveform began to decrement with the exact cut at a zero crossing. This was to avoid the contextual effects observed in Levy [2009], where consonantal context impacts on judgment. The vowel sounds were normalized with respect to pitch, amplitude and length [using Praat; Boersma and Weenink, 2005] so that spectral quality was the only cue for discrimination. The normalized length, 86 ms, was determined by taking the average length of the four vowels; the amplitude envelope was established so as to make the vowel sound as natural as possible, despite being extracted from a word, and with as little clipping as possible; the pitch contour was based on the vowel whose length was closest in length to the new vowel tokens.[4] We opted for the single best token of each vowel because multiple tokens would have required excessive testing to permute all possible pairings of the tokens and because there were only minimal differences between tokens of the same vowel after the extensive normalization.

The vowel formants at the midpoint of each vowel are shown in table 3. These measurements show that the four vowels are indeed distinguishable by spectral quality, with laxness reflecting an increase in the first formant frequency and roundedness reflecting a relative depression of the second and/or third formant frequencies.

[4] To ensure that these signal manipulations did not alter the phonetic identity of the tokens, we ran a small vowel identification experiment with a different set of participants. Four native English-speaking participants identified the /i/ and /ɪ/ tokens with 90 and 85% accuracy, respectively; 2 native French-speaking subjects identified the /i/ and /y/ tokens with 92 and 86% accuracy, respectively; 2 Turkish-speaking participants identified the /i/ and /y/ tokens with 95 and 92% accuracy, respectively.

Procedure

Participants were asked to assess whether two sounds were the same or different in a fixed AX discrimination task. The experiment was subdivided into one practice block followed by 12 test blocks in randomized order, each consisting of a comparison of two of the vowel tokens. Each block consisted of 16 trials, half of which had the same vowel token presented twice and half of which had different tokens. The two vowel sounds were separated by an interstimulus interval of 100 ms and listeners had 1,000 ms to respond with a button press of either (1) 'same' or (5) 'different'. Feedback (reaction time and correctness) was shown after each trial and a running total of percent correct was shown for motivation. The next trial started 2,500 ms after a response and 1,000 ms after the feedback screen.

Results

The percent correct and reaction time for all 12 vowel comparisons for each language are shown in figure 2. Half of the combinations are mirror images of the other half (e.g. [i]~[ɪ] and [ɪ]~[i]) and are merged for six order-independent pairs in the figure. The percent correct only show the results for vowel pairs that were different, while the reaction time reflects the results only for correct responses to different vowel pairs.

The reaction times and percent correct were analyzed using a repeated measures analysis of variance with the between-listeners factor native language (three levels) and the within-listeners factor vowel pair (six levels). There was no effect of language [$F(2, 267) = 0.81$, $p = 0.45$] on percent correct and the French speakers were slower, which may be an age effect as the French participants were older [$t(82) = 3.1$, $p = 0.003$]. Crucially, there was no interaction of language and vowel pair for reaction time [$F(10, 251) = 0.35$, $p > 0.9$] or percent correct [$F(10, 252) = 0.20$, $p > 0.9$]. This is also true when restricted to vowel pairs of interest for this study, [ɪ]~[ʏ] and [y]~[ʏ] [perent correct: $F(2, 84) = 0.05$, $p > 0.9$; RT: $F(2, 83) = 0.03$, $p > 0.9$].

Both percent correct [$F(5, 210) = 10.8$, $p < 0.001$] and reaction time [$F(5, 209) = 2.7$, $p = 0.02$] do reflect an effect of vowel pair. In particular, post-hoc two-sample t tests show that tense/lax contrasts facilitate discrimination more than the rounded/unrounded contrasts [table 4; RT: $t(175) = 2.4$, $p = 0.018$; percent correct: $t(150) = 3.7$, $p < 0.001$]. For the two crucial vowel pairs, listeners were slowest and got more incorrect responses for the [ɪ]~[ʏ] pair as compared to the [y]~[ʏ] comparison [post-hoc t test for percent correct: $t(75) = 3.5$, $p < 0.001$; RT: $t(87) = 2.0$, $p = 0.049$]. Even excluding the [ɪ]~[ʏ] pair, tense/lax pairs were discriminated faster and more accurately than rounded/unrounded (452 vs. 467 ms., 93 vs. 91%, respectively), which is marginally significant [$t(79) = 1.68$, $p = 0.10$; $t(87) = 1.50$, $p = 0.14$].

Fox's [1984] study of the time course of the Ganong effect suggests that lexical effects are more observable in trials with slower response times (above 800 ms). This suggests the possibility that we would see a language effect for slow responses in the experiment because participants had up to 1 s to respond. No effect is observable, however, when only including responses slower than 800 ms [$F(10, 23) = 0.65$, $p = 0.75$].

Discussion

There was no observable effect of native language on participants' ability to discriminate different vowel sounds across all vowel-pair comparisons for both reaction time and percent correct. This suggests that the speeded fixed AX discrimination task

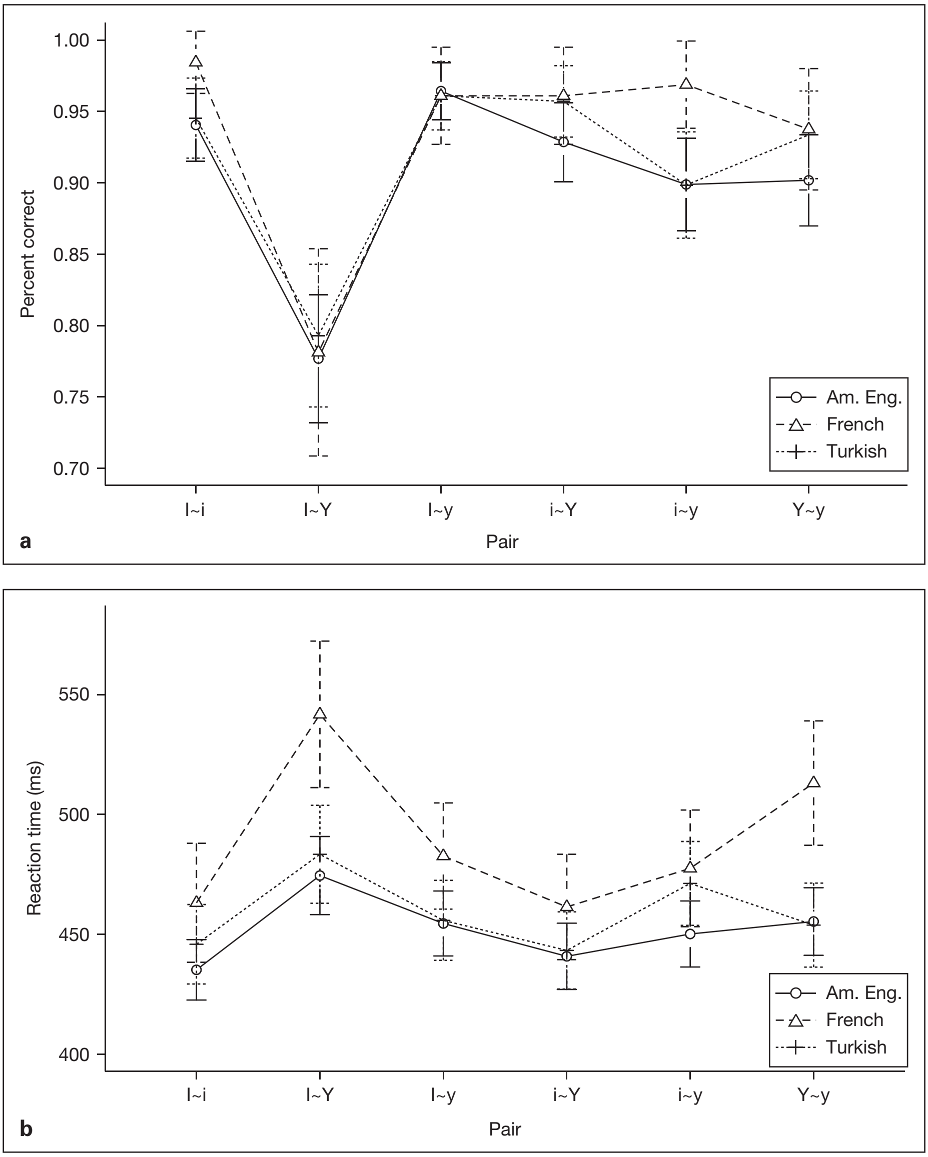

Fig. 2. Percent correct (**a**) and reaction time (**b**) for all vowel pairs.

can be taken to reflect a language-independent measure of the baseline discriminability of different vowels.

The main finding is the greater discriminability of the tense/lax contrast relative to rounding. An explanation of why this may be the case is that tenseness can be simplified as being based on F1 and F2, while the rounding contrast is primarily reflected in F2 and F3. While there are myriad acoustic cues signaling vowel quality, a gross

Table 4. Mean reaction time (proportion correct) for tense/lax vs. rounded/unrounded vowels pairs for English, French and Turkish speakers

Vowel pair	Language			
	English	Turkish	French	total
Tense/lax	445 (0.92%)	450 (0.94%)	483 (0.96%)	455 (0.93%)
Rounded/unrounded	461 (0.84%)	477 (0.85%)	487 (0.88%)	475 (0.85%)

approximation is that F1 and F2 serve as the primary determinants of vowel quality and are particularly salient. Therefore, we expect that the vowel contrasts that make use of these acoustic dimensions (tense/lax) are easier to discriminate than those that use less salient acoustic dimensions such as F3 (rounding).

This finding accords with, and can explain, the larger cross-linguistic distribution of the tense/lax contrast given that languages often have vowel inventories that are maximally discriminable to facilitate perceptibility or acquisition [De Boer, 2001; Ettlinger, 2007; Lindblom and Liljencrantz, 1972; Oudeyer, 2005]. A survey of the UPSID database [Maddieson and Precoda, 1990] shows that of the 451 languages, 74 (16.4%) make use of the tense/lax contrast for high front vowels, while only 24 (5.3%) use the rounded/unrounded contrast.

Experiment 2: Rating

In experiment 2, our aim was to study the role of native language – native language inventory and native language featural contrasts – on the perception and discrimination of native and non-native speech sounds. Previous research [Boomershine et al., 2008; Huang, 2004; Johnson, 2004] has shown that the results of a perceptual similarity judgment task reflect the influence of native language. So, to compare the two theories of how native language affects speech sound discrimination, we compared the similarity ratings for English, French and Turkish speakers for vowel sounds differing by rounding and tenseness. To review the predictions: If featural contrast serves as the main principle for the perception of speech sounds, then vowel pairs that reflect a contrast that is in the listener's native language should be judged as most different. However, if an inventory of exemplars defines the categories of language used for discrimination, then vowel pairs where one of the vowels exists in the listener's native inventory should be judged more distinct.

Method

As stated earlier, the same participants and vowels were used across all three experiments and the order of the three experiments was counterbalanced across participants for each language and procedure. In this task, participants were asked to give subjective ratings of the similarity of two speech sounds they heard. Each trial consisted of a comparison of two of the vowels; there were 4 × 4 comparisons possible and each comparison was tested 8 times for a total of 128 trials. Participants began with a practice block consisting of 5 random trials, and the 128 test trials were separated into two blocks of 64 to provide a short self-timed break. The two vowel sounds were separated by an interstimulus

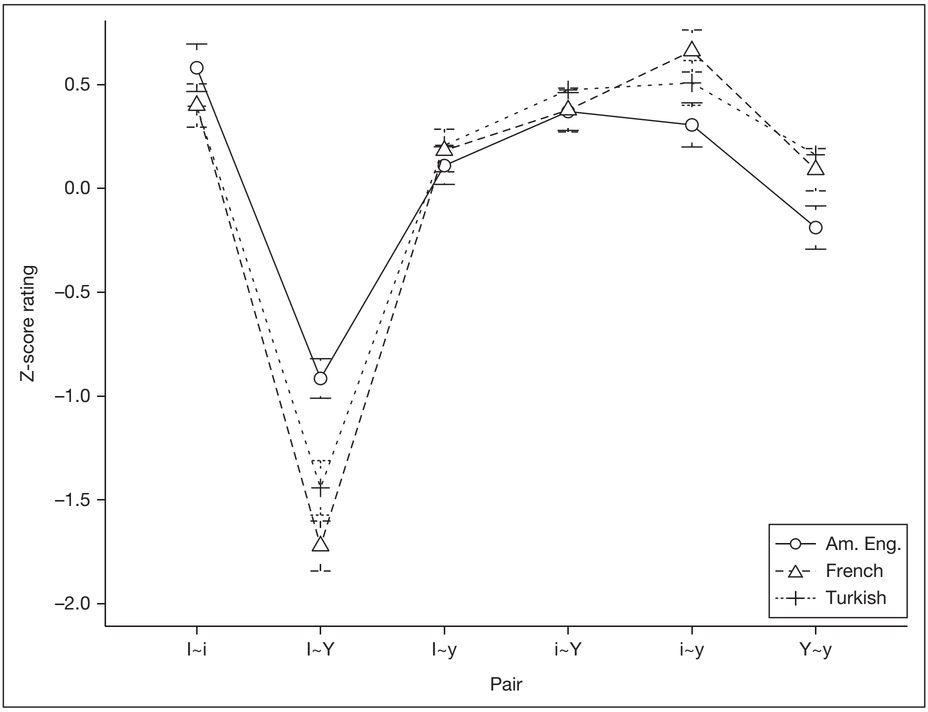

Fig. 3. Normalized similarity score of vowel comparisons for Turkish, French and English listeners.

interval of 500 ms and listeners had an unlimited amount of time to respond with a button press rating the pair on a 5-point scale from 'very similar' (1) to 'very different' (5). Their response was displayed on a screen for 1,500 ms and the next trial started 1,000 ms after the response screen.

Results

The mean rating for all vowel pairs was 3.86 (SD = 1.19) out of 5. English listeners' average response was 3.58 (SD = 1.17), Turkish 4.13 (SD = 1.14) and French 4.02 (SD = 1.19). A graph of the z-score normalized ratings (normalized separately for each listener) for all vowel pairs for all three listener groups is shown in figure 3.

Of note is the fact that vowel pairs in a native language were judged relatively more distinct by that group as one would expect (regardless of theory): English listeners judged [i]~[ɪ] more distinct [z score of 0.69 vs. 0.40 for non-English; nonparametric t test: $t(29) = 3.07$, $p = 0.0046$] and French and Turkish listeners judged [i]~[y] relatively more distinct [z score of 0.67 vs. 0.23 for English; $t(29) = 2.4$, $p = 0.025$].

The normalized ratings for the two critical vowel pairs, [y]~[ʏ] and [ɪ]~[ʏ], are shown in figure 4, and a repeated measures analysis of variance of the interaction of language and vowel pair is significant [$F(10, 210) = 4.11$, $p < 0.001$]. Planned post-hoc unpaired t tests show that Turkish and French speakers judge the [ʏ]~[y] pair as more different [$t(39) = 2.6$, $p = 0.012$], while English speakers judge the [ʏ]~[ɪ] vowel

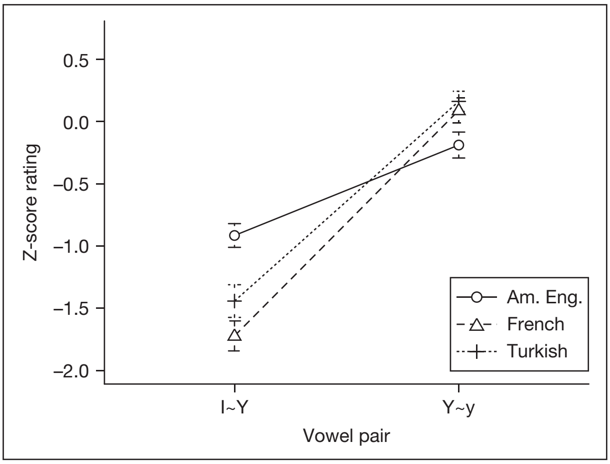

Fig. 4. Normalized similarity scores for vowel pairs of interest.

pair as more different [t(38) = 8.2, p < 0.001]. Neither of the other two comparisons, [i]~[ʏ] and [ɪ]~[y], which involve both featural contrasts, reflect any significant language differences.

Discussion

Both the [ɪ]~[ʏ] and [y]~[ʏ] ratings provide evidence for the exemplar-based theory of vowel perception. English speakers judge [ɪ]~[ʏ] as more distinct than Turkish and French speakers. This vowel pair reflects a rounded/unrounded contrast, a contrast that is present in the Turkish and French vowel inventories, and includes the English vowel /ɪ/. The [y]~[ʏ] vowel pair reflects the opposite featural contrast and inventory characteristics – an English featural contrast, but a Turkish/French inventory sound – and the opposite results obtain: Turkish and French speakers rate it as more distinct than English listeners. Therefore, the presence of one of the vowels in the listener's inventory facilitates discrimination more than the effect, if any, of the presence of a featural contrast.

The results of the other comparisons do not provide evidence for one theory over the other, but provide insight into the interaction between native-language effects and psychoacoustic differences. French and Turkish listeners rated the [i]~[y] pair as more distinct than English listeners, as one would expect from any theory given that these two sounds represent distinct phonemes in French and Turkish, but not in English. Similarly, English listeners judged [i]~[ɪ] as more distinct.

Experiment 3: Roving Discrimination

While the rating task showed an exemplar-based effect of language experience, the fixed task did not show any language effect. This, and the findings of Fox [1984] and Johnson and Babel [2010], suggests that the influence of language experience may

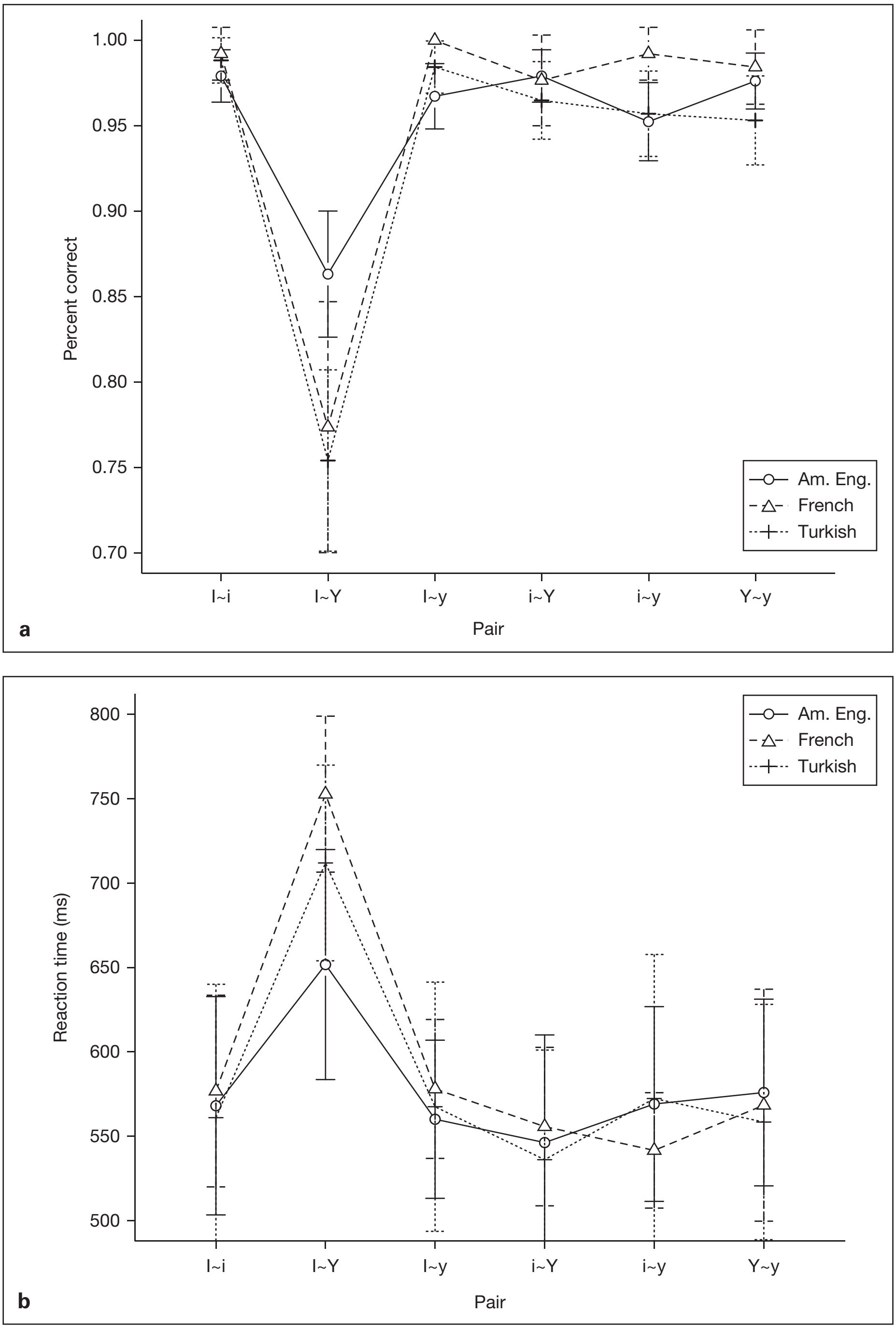

Fig. 5. **a** Percent correct on roving discrimination of all vowel pairs. **b** Reaction time on roving discrimination of all vowel pairs.

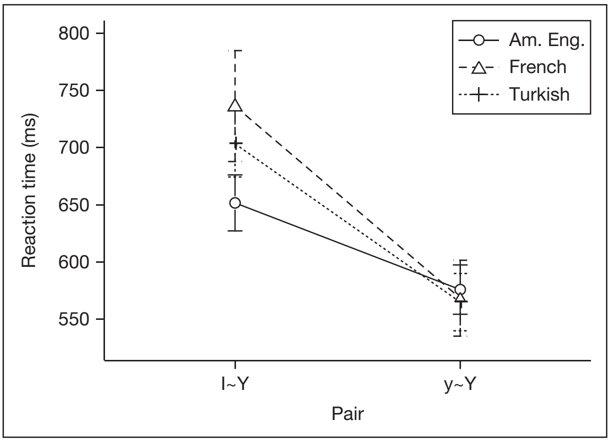

Fig. 6. Reaction time on roving discrimination of vowel pairs of interest.

be task-dependent and that a slower discrimination task with greater listener uncertainty may allow for native language influence. To explore this question, we conducted a roving AX discrimination task with longer interstimulus intervals using the same stimuli and participants as in experiments 1 and 2.

Method

Procedure. In this task, participants were asked to assess whether two sounds were the same or different. The experiment began with a practice block consisting of 10 random practice trials, 5 with the same two vowels and 5 with different vowels. The main portion of the experiment was subdivided into two blocks with a self-timed break in the middle. Half of the trials had the same two vowels, while the other half consisted of one of the 12 combinations of two different vowels repeated 8 times for a total of 12 × 8 × 2 = 192 trials in a different randomized order for each listener. The two vowel sounds were separated by an interstimulus interval of 500 ms and listeners had 2,000 ms to respond with a button press of either (1) ‘same’ or (5) ‘different’. Feedback (reaction time and correctness) was given after each trial and a running total of percent correct was shown.

Results

Results for the different vowel pairs are shown in figure 5a (percent correct) and b (reaction time). There is a marginally significant interaction between language and vowel pair for reaction time [$F(10, 210) = 1.655$, $p = 0.09$], whereas there is no significant interaction for percent correct.

Isolating the two crucial vowel pairs, [ɪ]~[ʏ] and [y]~[ʏ], reveals a significant interaction for reaction time between vowel pair and language via a repeated measures analysis of variance [$F(2, 42) = 3.508$, $p = 0.045$]. Post-hoc t tests reveal this is driven primarily by the [ɪ]~[ʏ] contrast [two-sample t test: $t(35) = 2.95$, $p = 0.0057$]), with Turkish and French speakers responding slower, while the /y/~/ʏ/ showed no significant difference [$t(35) = 0.40$, $p = 0.69$] as shown in figure 6. Examining order of presentation reveals a significant order effect for [ɪ]~[ʏ] when comparing English to

Turkish and French speakers. When the /ɪ/ is first, English listeners show improved reaction time as compared to when it is second, while Turkish and French speakers are slower [$F(2, 81) = 6.0$, $p = 0.0038$].

Discussion

The results reflect an amalgam of the effects observed in the language-influenced rating task and the language-independent fixed discrimination task. In particular, reaction time on this task shows a significant interaction of language and vowel pair for the diagnostic pairs of interest such that, just as with the rating task, listeners' performance was affected by native language in accordance with an exemplar theory of sound representations. When comparing vowels where at least one is foreign, subjects react more quickly when the other vowel is native as opposed to when the featural contrast is native. The results on this task were less unequivocal than the rating task, and were more statistically ambiguous. This suggests a couple of interesting caveats to the idea that this task reflects native-language experience.

First, there seems to be a threshold of performance, partly dictated by the psychoacoustic difference between the vowels. The significant reaction time results for the crucial pair is mainly driven by the [ɪ]~[ʏ] results but not the [y]~[ʏ] results. In other words, Turkish and French listeners did not receive much of a boost in performance on the [y]~[ʏ] discrimination task because of native-language inventory as compared to the significant native-language inventory effect observed for English listeners on the [ɪ]~[ʏ] discrimination task. This may reflect performance limitations as the reaction times on the [y]~[ʏ] discrimination are near peak discriminability for each of the listeners. So, English listeners' average was 578 ms and French and Turkish listeners were slightly faster (569 and 567 ms, respectively) with the fastest performance for each language group being 539 and 555 ms, respectively. Thus, it is reasonable to posit that significant effects of native language in discrimination are observable when the psychoacoustic difference between sounds is small enough to be perceptually challenging. It remains to be seen whether this lack of native language effect is present on the [y]~[ʏ] pair in more adverse listening conditions.

Second, for the [ɪ]~[ʏ] findings, we find that the performance improvement in reaction time for English listeners primarily comes when the vowel in the native language inventory occurs first, i.e. the pair order [ɪ]~[ʏ] is more discriminable than [ʏ]~[ɪ]. This finding complements work such as Polka and Bohn [1996] where what matters in order of presentation is peripherality. They found that both German and English infants were better at discriminating vowels in their language (/u/~/y/ and /æ/~/ɛ/, respectively) when the more peripheral vowel is presented second. Similarly, myriad studies with adults have shown that discrimination is poorer for a vowel change from a high, front vowel to a lower or more central vowel compared to a change in the reverse direction [Cowan and Morse, 1986; Iverson and Kuhl, 1995; Repp and Crowder, 1990; Sussman and Lauckner-Morano, 1995]. Our results suggest the finding can be generalized to an observation concerning order effects and general prominence, be it nativeness or peripherality.

There is a crucial distinction, however. In Polka and Bohn [1996] as well as in other studies involving infants [Polka and Werker, 1994] and adults [Grieser and Kuhl, 1989; Iverson and Kuhl, 1995], greater discriminability comes from having the peripheral vowel *second*. This may stem from the adult studies using /i/ and /ɪ/, where /ɪ/ is actually more frequent in English by type and token [type: 37,000:27,000; token: 172,000:96,000; extracted from *CMU Pronunciation Dictionary*, Weide, 1994]. An exemplar-based account of vowel categorization of the sort presented in Johnson [1997] may account for this observation. For English listeners, when the first vowel is [ɪ], this sound activates a large cloud of exemplars associated with the native vowel category. On hearing the second, different vowel, discrimination is faster because of the robust boundaries and stronger memory trace of the initial vowel category articulated by the exemplars. When [ʏ] is heard first, however, only a small set of exemplars are activated, leading to difficulty in categorizing the sound as a member of the /ɪ/, /ʌ/, /ə/ or /ʊ/ categories. When the second vowel is heard, discrimination is more difficult because of the lack of clear definition of what precisely the original [ʏ] sound was or due to a rapidly deteriorating memory trace. This is reminiscent of frequency effects on wordhood judgment tasks where wordhood is more quickly assessed for more frequent words, presumably because of the greater number of exemplars activated [Balota and Chumbley, 1984; Whaley, 1978]. Here, the larger number of nearby exemplars of the first sound facilitates its ability to serve as an anchor for the discrimination of a following sound.

General Discussion

Overall, these results can make a strong case for an exemplar-based model of speech discrimination. Native-language inventory facilitates distinguishing vowels in a rating task and a roving discrimination task more than the possible effect of native-language featural contrast. This can be expressed graphically using multidimensional scaling [MDS; Francis and Nusbaum, 2002; Johnson, 2008] of the rating scores for the vowels as shown in figure 7 for each listener group. Larger distances correspond to larger difference scores given by listeners and smaller distances correspond to greater similarity. The figures show a stretching of the perceptual space around /ɪ/ for English speakers, and around /y/ for Turkish and French speakers instead of a stretching along featural dimensions. (This two-dimensional metric MDS analysis of four categories is not meant to suggest any conclusive findings regarding the dimensions differentiating these vowels, rather is only included for illustrative purposes to present a graphic representation of the inventory effect.)

This result does not obviate feature-based analyses of sound patterns as evidenced by the success of feature theory in explaining phonological processes and natural classes [but see Mielke, 2008]. The conjecture in Stevens [2002] is that the representations governing the patterning of sounds are the same as those used for speech perception. The results here suggest that this is not the case. Instead, there is converging evidence that exemplar-based representations are used for lower-level tasks like sound discrimination as well as speaker normalization (see references above), which are crucial to speech perception, and that featural contrasts best account for the higher-order patterning of language, perhaps through emergent features [Mielke, 2008; Wedel, 2004]. Along these lines, it may be that featural contrast also plays a role in discrimination, albeit a lesser role than inventory. Indeed, this approach would be able to incorporate the present findings with studies presenting evidence for feature-based discrimination [Iverson et al., 2003].

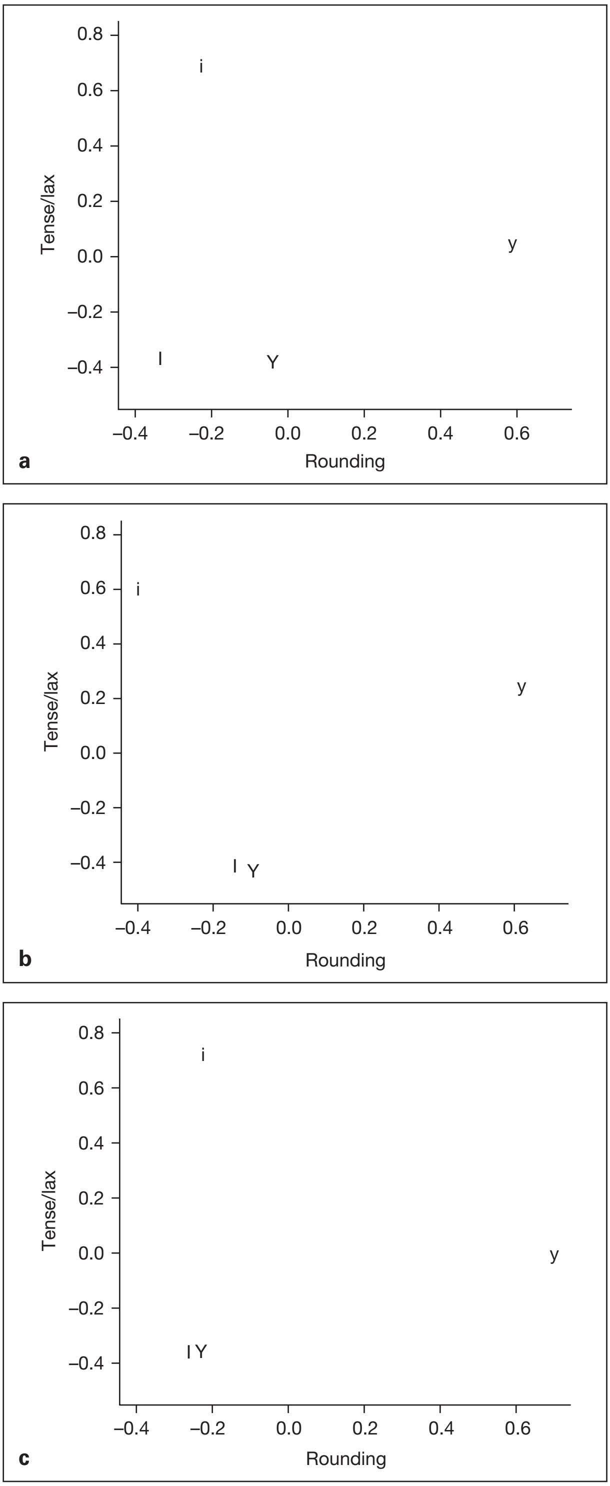

Fig. 7. Multidimensional scaling of vowel rating scores for English (**a**), French (**b**) and Turkish (**c**) listeners.

Table 5. Mean Bark values for formants for German /ʏ/ stimulus, French /ø/, Turkish /œ/ and English /u/

Vowel	F3-F2, Barks	F2-F1, Barks	F1, Barks
German stimulus ʏ	2.6	8.5	3.5
German stimulus y	0.86	10.9	2.8
French ø	3.3	7.3	4.5
French y	3.3	11.1	2.6
Turkish œ	3.2	6.4	4.9
Turkish y	2.8	9.2	2.8
English u	1.4	9.1	3.0

This study does not provide evidence for an overall coarse distinction of linguistic processes by representation, and so further work is needed to understand the parcellation of linguistic function by linguistic representation. Similarly, future experiments can also illuminate whether exemplar models can account for another property of language ascribed to features, the reuse of the same features across the vowel space. Here, we only examined the high front vowel, but computational models suggest that exemplars can account for vowel dispersion of language as well [Ettlinger, 2007; Oudeyer, 2005].

Ultimately, these claims must be tempered by the practical limitations of the generation of stimuli. One issue relates to the use of German vowels, particularly /i/, /ɪ/ and /y/, which differ slightly from their English, French and Turkish counterparts. These differences are minor, however, particularly in light of the individual variation speakers hear for their own language, and using a single speaker is preferable to using different stimuli for each speaker group with different speakers for non-native vs. native vowels, not to mention the need to include /ʏ/.

A second issue concerns the possibility that French and Turkish listeners are assimilating /ʏ/ to /ø/ and /œ/, respectively, and are therefore using the height feature that distinguishes these vowels from /y/ in their language to facilitate [y]~[ʏ] discrimination. The mean Bark differences for the key vowels in these languages and the stimuli are shown in table 5 [French data from Strange et al., 2007; Turkish from Kiliç and Ögüt, 2004]. Indeed, the similarity between German stimulus [ʏ] and French /ø/ suggests that this is possible for French speakers. This is unlikely for Turkish speakers, however, where the F2 of a typical /œ/ is much lower than that of the front vowels (reflected in the low F2-F1 Bark value) and the [ʏ] is much more likely to be compared to /y/ based on acoustic properties [Kiliç and Ögüt, 2004]. A similar issue arises due to the /u/ fronting present in many dialects of English. The data for one such dialect [Clopper and Pisoni, 2004] is presented in table 5, and shows that [y], not [ʏ], is more likely to assimilate to /u/ based on acoustic similarity. Ultimately, this reflects an inability to find languages with high front round vowels but not mid front round vowels. While alternate explanations are possible for certain results, such as French discrimination of [y]~[ʏ], these accounts are unable to explain the entirety of the data. We believe that exemplar theory accounts for all the observations most parsimoniously.

This study also explores one dimension of how psychoacoustic and language-dependent factors interact in speech sound discrimination. We observe that the [y]~[ʏ] psychoacoustic distinction is greater than that of [ɪ]~[ʏ] using a fixed discrimination task. This accords with typological evidence (tenseness a more frequent distinction for front vowels cross-linguistically as discussed above) and serves as a baseline for looking at performance on other tasks. We then find that language-specific effects only come into play on a slower roving discrimination task when two sounds are close enough together in the perceptual space (e.g. /ɪ/~/ʏ/), but are irrelevant when two sounds are very different. So, the tuning of the perceptual system for vowels seems to only have an effect at relatively small psychoacoustic distances that are not always present in vowel systems; some distinctions (e.g. /y/~/ʏ/) may therefore be universally discriminable, while other, more subtle vowel distinctions (e.g. /ɪ/~/ʏ/) may require perceptual tuning. Further confirmation of this idea will depend on assessing the discrimination of /y/~/ʏ/ in more adverse listening conditions.

These findings also relate to categorical perception, where sounds close to a native-language vowel prototype are more difficult to discriminate than are sounds at category boundaries [Eimas, 1963; Liberman et al., 1957]. The present findings may be interpreted such that the non-native vowel [ʏ] is near the category boundaries of /ɪ/ for English and of /y/ for Turkish and French. For English listeners, the presence of /ɪ/ in the inventory serves to delineate a boundary for the /ʏ/~/ɪ/ dimension and for Turkish and French listeners, the presence of /y/ in the inventory serves to define the boundary for the /ʏ/~/y/ dimension.

Finally, a minor point: These results suggest that the ordering effects in a speech discrimination task are not only related to peripherality [Polka and Bohn, 1996], but also depend on whether the sounds are native or non-native as well, with discrimination facilitated when the native sound is first. This may be accounted for via an exemplar-based theory with peripherality and nativeness both being associated with exemplar strength, but is significantly more challenging to explain if one assumes a feature-based approach.

References

Archangeli, D.B.: Grounded phonology (MIT Press, Cambridge 1994).

Balota, D.A.; Chumbley, J.I.: Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. J. exp. Psychol. hum. Percept. Perform. *10*: 340–357 (1984).

Best, C.T.: A direct realist perspective on cross-language speech perception; in Strange, Speech perception and linguistic experience: theoretical and methodological issues in cross-language speech research, pp. 167–200 (Timonium, York 1995).

Best, C.T.; McRoberts, G.W.; Goodell, E.: Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. J. acoust. Soc. Am. *109*: 775 (2001).

Boersma, P.; Weenink, D.: Praat: Doing phonetics by computer; version 4.2 (2005). *URL: www.praat.org*.

Bohn, O.S.: Cross-language speech perception in adults: first language transfer doesn't tell it all; in Strange, Speech perception and linguistic experience: issues in cross-language research, pp. 279–304 (York Press, Baltimore 1995).

Boomershine, A.; Currie Hall, K.; Hume, B.; Johnson, K.: The influence of allophony vs. contrast on perception: the case of Spanish and English; in Avery, Dresher, Rice, Phonological contrast (Mouton, The Hague 2008).

Browman, C.P.; Goldstein, L.: Towards an articulatory phonology; in Ewen, Anderson, Phonol. Yb., vol. 3, pp. 219–252 (Cambridge University Press, Cambridge 1986).

Browman, C.P.; Goldstein, L.: Articulatory phonology: an overview. Phonetica *49*: 155–180 (1992).

Chomsky, N.; Halle, M.: The sound pattern of English (Harper & Row, New York 1968).

Clopper, C.G.; Pisoni, D.B.: Effects of talker variability on perceptual learning of dialects. Lang. Speech *47*: 207–239 (2004).

Condax, J.D.; Krones, R.R.: Duration of four vowels in manually produced synthetic speech. J. Phonet. *4*: 151–171 (1976).

Cowan, N.; Morse, P.A.: The use of auditory and phonetic memory in vowel discrimination. J. acoust. Soc. Am. *79*: 500–507 (1986).

De Boer, B.: The origins of vowel systems (Oxford University Press, Oxford 2001).

Eimas, P.D.: The relation between identification and discrimination along speech and non-speech continua. Lang. Speech *6*: 206–217 (1963).
Ettlinger, M.: An exemplar-based model of chain shifts. Proc. 16th Int. Congr. Phonet. Sci., 2007, pp. 685–688.
Fant, G.: Speech sounds and features (MIT Press, Cambridge 1973).
Fischer-Jørgensen, E.; Jørgensen, H.P.: Close and loose contact (‘Anschluss’) with special reference to North German. Annu. Rep. Inst. Phonet. Univ. Copenh. *4:* 43–80 (1969).
Flege, J.E.: Speech learning in a second language; in Phonological development: models, research, implications, pp. 565–604 (York Press, Timonium 1992).
Fox, R.A.: Effect of lexical status on phonetic categorization. J. exp. Psychol. hum. Percept. Perform. *10*: 526–540 (1984).
Francis, A.L.; Nusbaum, H.C.: Selective attention and the acquisition of new phonetic categories. J. exp. Psychol. hum. Percept. Perform. *28*: 349–366 (2002).
Goldinger, S.D.: Words and voices: episodic traces in spoken word identification and recognition memory. J. exp. Psychol. Learn. Memory Cognit. *22*: 1166–1183 (1996).
Goldinger, S.D.: Echoes of echoes? An episodic theory of lexical access. Psychol. Rev. *105*: 251–279 (1998).
Goldinger, S.D.: A complementary-systems approach to abstract and episodic speech perception. 16th Int. Congr. Phonet. Sci., 2007.
Goldstein, L.; Fowler, C.A.: Articulatory phonology: a phonology for public language use; in Schiller, Meyer, Phonetics and phonology in language comprehension and production: differences and similarities, pp. 159–207 (Mouton de Gruyter, Berlin 2003).
Grieser, D.; Kuhl, P.K.: Categorization of speech by infants: support for speech-sound prototypes. Devl Psychol. *25*: 577–588 (1989).
Hall, C.: Modern German pronunciation: an introduction for speakers of English (Manchester University Press, Manchester 2003).
Hintzman, D.L.; Block, R.; Inskeep, N.: Memory for mode of input. J. verbal Learn. verbal Behav. *11*: 741–749 (1972).
Huang, T.: Language specificity in auditory perception of Chinese tones; PhD diss. Ohio State University (2004).
Iverson, P.; Kuhl, P.K.: Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. J. acoust. Soc. Am. *97*: 553–562 (1995).
Iverson, P.; Kuhl, P.K.; Akahane-Yamada, R.; Diesch, E.; Tohkura, Y.; Kettermann, A.; Siebert, C.: A perceptual interference account of acquisition difficulties for non-native phonemes. Cognition *87*: B47–B57 (2003).
Jakobson, R.; Fant, G.; Halle, M.: Preliminaries to speech analysis (MIT Press, Cambridge 1952).
Jessen, M.: Phonetics and phonology of tense and lax obstruents in German (Benjamins, Amsterdam 1998).
Johnson, K.: Speech perception without speaker normalization: an exemplar model; in Johnson, Mullennix, Talker variability in speech processing, pp. 145–165 (Academic Press, San Diego 1997).
Johnson, K.: Cross-linguistic perceptual differences emerge from the lexicon; in Agwuele, Warren, Park, Proc. 2003 Texas Linguistics Soc. Conf., pp. 26–41 (Cascadilla Press, Somerville 2004).
Johnson, K.: Quantitative methods in linguistics (Blackwell, Oxford 2008).
Johnson, K.; Babel, M.: On the perceptual basis of distinctive features: Evidence from the perception of fricatives by Dutch and English speakers. J. Phonet. *38*: 127–138 (2010).
Kiliç, M.A.; Ögüt, F.: A high unrounded vowel in Turkish: is it a central or back vowel? Speech Commun. *43*: 143–154 (2004).
Kiliç, M.A.; Ögüt, F.; Dursun, G.; Okur, E.; Yildirim, I.; Midilli, R.: The effects of vowels on voice perturbation measures. J. Voice *18*: 318–324 (2004).
Lang, C.E.; Ohala, J.J.: Temporal cues for vowels and universals of vowel inventories. ICSLP 96, Proc. 4th Int. Conf. on Spoken Language, vol. 1, 1996.
Levy, E.S.: Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. J. acoust. Soc. Am. *125*: 1138 (2009).
Liberman, A.M.; Mattingly, I.G.: A specialization for speech perception. Science *243*: 489–494 (1989).
Liberman, A.M., Harris, K.S., Hoffman, H.S., Griffith, B.C.: The discrimination of speech sounds within and across phoneme boundaries. J. exp. Psychol. *54*: 358–368 (1957).
Lindblom, B.; Liljencrantz, J.: Numerical simulation of vowel quality systems: the role of perceptual contrast. Language *48*: 839–862 (1972).
Lively, S.E.; Pisoni, D.B.; Yamada, R.A.; Yamada, T.: Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. J. acoust. Soc. Am. *96*: 2076 (1994).
Maddieson, I.; Precoda, K.: Updating UPSID. UCLA Working Papers Phonet. *74*: 104–111 (1990).
Massaro, D.W.: Experimental psychology and information processing (Rand McNally, Chicago 1975).
McNeil, D.; Lindig, K.: The perceptual reality of phonemes, syllables, words, and sentences. J. verbal Learn. verbal Behav. *12*: 419–430 (1973).
McQueen, J.M.; Cutler, A.; Norris, D.: Phonological abstraction in the mental lexicon. Cognitive Sci. *30*; 1113–1126 (2006).
Mielke, J.: The emergence of distinctive features (Oxford University Press, Oxford 2008).
Norris, D.; McQueen, J.M.; Cutler, A.: Perceptual learning in speech. Cognitive Psychol. *47*: 204–238 (2003).
Nosofsky, R.M.: Attention, similarity, and the identification-categorization relationship. J. exp. Psychol. *115*: 39–57 (1986).

Oudeyer, P.Y.: The self-organization of speech sounds. J. theor. Biol. *233*: 435–449 (2005).
Pierrehumbert, J.: Exemplar dynamics: word frequency, lenition, and contrast; in Bybee, Hopper, Frequency effects and the emergence of lexical structure, pp. 137–157 (Benjamins, Amsterdam 2001).
Pisoni, D.B.: Some comments on talker normalization in speech perception; in Tohkura, Vatikiotis-Bateson, Sagisaka, Speech perception, production and linguistic structure, pp. 143–151 (Ohmsha, Tokyo 1992).
Pisoni, D.B.; Sawusch, J.R.: Some stages of processing in speech perception; in Cohen, Nooteboom, Structure and process in speech perception, pp. 16–35 (Springer, New York 1975).
Poeppel, D.; Idsardi, W.; van Wassenhove, V.: Speech perception at the interface of neuroscience and linguistics. Phil. Trans. R. Soc. 363: 1071–1086 (2008).
Polka, L.; Bohn, O.S.: A cross language comparison of vowel perception in English learning and German learning infants. J. acoust. Soc. Am. *100*: 577 (1986).
Polka, L.; Werker, J.F.: Developmental changes in perception of nonnative vowel contrasts. J. exp. Psychol. *20*: 421–435 (1994).
Repp, B.H.; Crowder, R.G.: Stimulus order effects in vowel discrimination. J. acoust. Soc. Am. *88*: 2080–2090 (1990).
Schacter, D.L.; Eich, J.E.; Tulving, E.: Richard Semon's theory of memory. J. verbal Learn. verbal Behav. *17*: 721–743 (1978).
Singh, S.; Woods, D.R.: Perceptual structure of 12 American English vowels. J. acoust. Soc. Am. *49*: 1861 (1971).
Stevens, K.N.: Toward a model for lexical access based on acoustic landmarks and distinctive features. J. acoust. Soc. Am. *111*: 1872–1891 (2002).
Strange, W.; Weber, A.; Levy, E.S.; Shafiro, V.; Hisagi, M.; Nishi, K.: Acoustic variability within and across German, French, and American English vowels: phonetic context effects. J. acoust. Soc. Am. *122*: 1111–1129 (2007).
Sussman, J.E.; Lauckner Morano, V.J.: Further tests of the 'perceptual magnet effect' in the perception of [i]: identification and change/no change discrimination. J. acoust. Soc. Am. *97*: 539 (1995).
Wedel, A.B.: Self-organization and categorical behavior in phonology; PhD thesis University of California at Santa Cruz (2004).
Weide, R.L.: CMU pronouncing dictionary (1994). Downloaded from http://www.speech.cs.cmu.edu/cgi-bin/cmudict on February 1, 2010.
Whaley, C.P.: Word-nonword classification time. J. verbal Learn. verbal Behav. *17*: 143–154 (1978).