Original Paper

Phonetica 2014;71:4–21
DOI: 10.1159/000356237

Received: June 18, 2013
Accepted after revision: October 9, 2013
Published online: June 5, 2014

Perceptual Assimilation and Discrimination of Non-Native Vowel Contrasts

Michael D. Tyler[a,b] Catherine T. Best[b,c] Alice Faber[c,d]
Andrea G. Levitt[c,e]

[a]School of Social Sciences and Psychology, University of Western Sydney; [b]The MARCS Institute, University of Western Sydney, Penrith, NSW, Australia; [c]Haskins Laboratories, New Haven, Conn., [d]Department of Psychology, Wesleyan University, Middletown, Conn., and [e]French Department, Wellesley College, Wellesley, Mass., USA

Abstract

Research on language-specific tuning in speech perception has focused mainly on consonants, while that on non-native vowel perception has failed to address whether the same principles apply. Therefore, non-native vowel perception was investigated here in light of relevant theoretical models: the Perceptual Assimilation Model (PAM) and the Natural Referent Vowel (NRV) framework. American-English speakers completed discrimination and native language assimilation (categorization and goodness rating) tests on six non-native vowel contrasts. Discrimination was consistent with PAM assimilation types, but asymmetries predicted by NRV were only observed for single-category assimilations, suggesting that perceptual assimilation might modulate the effects of vowel peripherality on non-native vowel perception.

Humans are born with the capacity to acquire the language of their environment, but quickly become 'tuned in' to the specific phonetic categories used in their native language. Research on adult cross-language speech perception suggests that the benefits of this perceptual attunement to native speech are often associated with a cost to discrimination of certain pairs of phones that signal a non-native phonological contrast in a language the listener has not previously been exposed to. That is, there is a sort of 'tuning out' of non-native contrasts that are irrelevant in the native language. The extent to which specific non-native contrasts are discriminable varies considerably, however, ranging from poor near-chance performance to excellent near-native performance levels. In recognition of those contrast-specific differences in discrimination, a number of theoretical models have sought to address the causes of the variation in performance.

However, the majority of research on this issue has focused on discrimination of non-native consonant contrasts. Relatively little is known about the extent to which performance on non-native vowel contrasts exhibits the same range of variability, nor

KARGER

E-Mail karger@karger.com
www.karger.com/pho

0031–8388/14/0711–0004
$39.50

Karger Open access

Dr. Michael Tyler
School of Social Sciences and Psychology, Building 24
Bankstown Campus, University of Western Sydney
Locked Bag 1797, Penrith, NSW 2751 (Australia)
E-Mail m.tyler@uws.edu.au

whether perception of non-native vowel contrasts follows the same or different principles as non-native consonant contrasts. Given numerous articulatory, acoustic, phonological, and perceptual differences between the two major segmental classes, it is important to investigate the possibility that the range and causes of variability in discrimination across non-native vowel contrasts may differ in at least some ways from that reported for consonants. The purpose of the present study is to evaluate whether similar or different principles may underlie perception of non-native vowel contrasts than theory and evidence have suggested for non-native consonant contrasts.

Acoustically, vowels differ from the majority of consonants in that they are usually of higher acoustic intensity, are often more extended temporally, and are distinguished from each other primarily in the first three formant frequencies [Ladefoged, 2005]. The acoustics of consonants, on the other hand, vary markedly depending on consonant class – nasals and approximants can be described largely in terms of formant frequency transitions, whereas stops and fricatives include as well some aperiodic noise component (stop release burst; frication, which is usually temporally extended). These acoustic differences between vowels and consonants appear to be accompanied by differences in how they are perceived. In classic categorical perception, labelling functions are less steep for vowels than for consonants, suggesting that the boundaries between phonological categories may be less sharp, and within-category discrimination may be better, for vowels than consonants [Fry et al., 1962].

Given these characteristics on which consonants and vowels differ, there is good reason to suspect that they might also impact on how well the cross-language speech perception models apply to vowel contrasts as compared to what is known about consonant contrasts. The three most commonly cited general models of cross-language speech perception are the Speech Learning Model [Flege, 1995, 2002], the Native Language Magnet Model [Kuhl, 1991, 1992], and the Perceptual Assimilation Model [Best, 1993, 1994a, b, 1995]. As we are interested here in perception of non-native contrasts by naïve listeners, whereas the Speech Learning Model is primarily concerned with second language (L2) speech learning, focuses on individual phones rather than contrasts, and on production more so than perception, we will not consider it further here (nor the two newer L2 speech learning models): Second Language Linguistic Perception [Escudero and Boersma, 2004; Escudero et al., 2009; or PAM-L2, Best and Tyler, 2007]. As the data supporting the Language Magnet Model have been widely criticized [e.g., Frieda et al., 1999; Lively and Pisoni, 1997; Lotto et al., 1998; Sussman and Lauckner-Morano, 1995], we focus our theoretical discussion on the Perceptual Assimilation Model (PAM) and a more recent model, the Natural Referent Vowel framework [NRV; Polka and Bohn, 2003, 2011].[1] The principles guiding NRV focus

[1] Although we did not test it directly, it is important to be aware of the predictions regarding stimulus and task variables of another recent model, the Automatic Selective Perception model (ASP) [Strange, 2011]. According to ASP, listeners develop two types of selective perceptual routines (SPR) for speech perception. Phonological SPRs detect information that is minimally sufficient for detecting contrasting word forms, and phonetic SPRs are used to detect context-dependent allophonic details. Perception in the phonetic mode requires focused attention, and more cognitive effort, than perception in the phonological mode. The native language thus interferes with cross-language perception primarily when automatic phonological SPRs are inappropriately applied to non-native phones. The ASP model emphasizes the influence of task demands and stimulus complexity on the accessibility of phonetic details. When tasks require rapid processing or have a high memory load, or when stimuli are embedded in phrase or sentence contexts, listeners are less likely to detect fine-grained phonetic detail than in less demanding tasks with more phonetically simple stimuli. The ASP model applies to both consonants and vowels, although many of the studies that led to its development investigated cross-language perception of vowels.

specifically on the acoustic and articulatory properties of vowels whereas PAM predictions were not intended to be specific to either consonants or vowels.

According to PAM [Best, 1993, 1994a, b, 1995], speech perception is shaped by perceptual attunement to the physical consequences of the constellations of articulatory gestures that signal phonological contrasts in the ambient language environment. During language acquisition, native perceivers become highly efficient at detecting single-feature phonetic differences that signal a meaningful change in lexical items, that is, at recognizing *phonological distinctiveness* within the native language. They become equally efficient at tuning out within-category phonetic differences that do *not* signal a change of meaning, that is, at recognizing *phonological constancy* within that language [Best et al., 2009]. On this view, phonological categories are formed through processes of perceptual learning [in the sense of Gibson and Gibson, 1955], whereby perceivers come to detect and attend to those higher-order phonetic invariants that are necessary and sufficient to discriminate one category from other categories that contrast with it in the native-language phonological space (distinctiveness), and learn to largely shift attention away from irrelevant variations (constancy). If a given non-native phonological contrast is perceptually assimilated as phonetically similar to a phonological distinction, that is, a *contrast*, in the native language, then it should be easily discriminated. Conversely, if the contrasting non-native phones do not assimilate to a native contrast, then they may be difficult to discriminate, at least initially, depending on the extent to which the corresponding phonetic distinctions have been tuned out in the native language. Thus, to test whether PAM can account for cross-language speech perception, it is necessary to investigate categorization and discrimination of non-native *contrasts*, and not only perception of singular phonetic categories (which are the focus of the Speech Learning Model and the Language Magnet Model).

PAM predicts that discrimination performance on non-native contrasts will vary from poor to excellent depending on how the contrasting non-native phones are assimilated (categorized and goodness-rated) to native phonological categories. The perceiver may detect features of a non-native phone that are perfectly consistent with a native phonological category, in which case it would be assimilated as an excellent version of the category, or it may be partially consistent and assimilated as a moderate-to-poor version of the category. In all of those cases, the non-native phone is said to be *categorized* within the listener's native phonological system. If the features of the non-native phone are not consistent with any one native category, then it is *uncategorized*, and if it is not heard as speech, then it is *non-assimilable* [Best et al., 1988]. Discrimination of contrasting non-native stimuli will be most accurate when the two non-native phones have been assimilated to different native phonological categories, a *two-category assimilation (TC)*, and less accurate when both are assimilated to the same phonological category. If both are heard as equally good or poor versions of the native category, a *single-category assimilation (SC)*, then discrimination will be poorer than if one is assimilated as a good version and the other as a poor version, a *category-goodness assimilation (CG)*. Discrimination of contrasts involving one uncategorized and one categorized phone, an *uncategorized-categorized assimilation (UC)*, is expected to be very good, and those involving two uncategorized phones, an *uncategorized-uncategorized assimilation (UU)* can vary from poor to excellent, depending on the phonetic similarity of the two phones to each other and to native phonological categories. Analogously, a non-native contrast between *two non-assimilable* phones can vary from fair to excellent, depending on how similar the two items

DOI: 10.1159/000356237

are to each other in their non-speech auditory properties [for a more detailed discussion, see Best, 1995].

These assimilation and discrimination predictions were conceived in PAM as generally applicable to segmental contrasts, that is, both consonants and vowels, but in fact PAM illustrations and empirical tests have focused largely on consonant contrasts. That observation is a core motivation for the present investigation on variations in performance among non-native vowel contrasts. For example, given the shallower category boundaries and higher within-category discrimination for vowels than consonants in categorical perception [Fry et al., 1962], SC assimilations might not occur for vowel contrasts, and absolute levels of discrimination could be higher for vowels than for consonants. Indeed, it is possible that even the differences in discrimination performance that PAM predicts among SC, CG, TC, UC and UU assimilation pairs could be masked or overridden by the less categorical perception of vowel than consonant contrasts. Moreover, vowels that are non-assimilable to speech (perceived as non-speech sounds) are unlikely to exist at all.

Prior research on PAM that has focused on non-native consonant contrasts has examined the full range of assimilation types from SC to CG to TC to non-assimilable, and their predicted relationships to discrimination levels in the same participants (e.g., excellent performance for TC assimilations, fair to very good performance for CG, and poor performance for SC assimilations), for both naturally produced contrasts [Best et al., 2001; see also Harnsberger, 2001; Polka, 1991, 1992; Polka et al., 2001] and computer-synthesized continua [Best and Strange, 1992; Bohn and Best, 2012; Hallé et al., 1999]. Support for PAM predictions also extends to non-native consonant clusters [Best and Hallé, 2010; Hallé and Best, 2007; Hallé et al., 1998]. Some studies have also been conducted with non-native vowels that were interpreted by those authors according to certain PAM principles, but they have most often looked only at categorization [e.g., Escudero and Boersma, 2004; Escudero and Williams, 2011; Strange et al., 2001, 2004, 2009] or only at discrimination [e.g., Levy and Strange, 2008]. They have not administered both tasks to the same listeners except for a few studies.

Two of the latter studies examined perception of two German vowel contrasts, /u/-/y/ and /ʊ/-/ʏ/, by English-speaking monolinguals [Polka, 1995; Polka and Bohn, 1996]. Our interpretation of their categorization results is that /u/-/y/ was a PAM CG assimilation, and /ʊ/-/ʏ/ was a PAM UC assimilation. Those authors found that discrimination was excellent (98–100% correct) for the former contrast, and very good (87%) for the latter contrast. The near-ceiling discrimination of the /u/-/y/ contrast that we have interpreted as a CG assimilation, based on their categorization results, suggests that there may indeed be some difference between vowels and consonants in the levels of discrimination accuracy across assimilation types, but it is crucial to compare the full range of PAM assimilation types.

In another pair of studies on the perceptual assimilation and discrimination of eight French vowel contrasts, naïve English-speaking listeners performed more poorly on discrimination of contrasts in which both phones were assimilated to the same native category (SC or CG) than on TC assimilations [Levy, 2009a, b]. Interestingly, some vowels were assimilated to different native categories depending on the consonantal context (bilabial or alveolar). Although this also provides some support for PAM predictions with vowels, the full range of assimilation types was not assessed in those studies either. Therefore, here we tested non-native vowel discrimination across a range of non-native contrasts that we anticipated would cover a full array of assimilation types (SC, CG, TC, UC, and UU), and that would allow us to systematically compare the two most germane theoretical models, PAM and NRV.

The NRV framework [Polka and Bohn, 2003, 2011] seeks to explain asymmetries observed in detection of a change from one vowel category to another by infants and adults. It appears that detecting a change from a repeating vowel that is acoustically and articulatorily more peripheral in the vowel space, to one that is less peripheral, is more difficult than detecting a change in the opposite direction. The perceptual asymmetry is observed in monolingually raised infants for both native and non-native vowel contrasts at 6 months of age, but only for non-native contrasts at 12 months and in adulthood. For example, the asymmetry was observed for Danish-learning infants' discrimination of Southern British English /ɒ/-/ʌ/ at both 6 and 12 months of age, but the asymmetry observed at 6 months for the native Danish /e/-/ɛ/ contrast was no longer apparent at 12 months of age [Polka and Bohn, 2011]. The asymmetry for discrimination of the non-native Southern British English /ɒ/-/ʌ/ continues into adulthood for Danish listeners.

On the basis of this, Polka and Bohn [2011] argue that more peripheral vowels act as perceptual anchors to guide the development of native vowel categories [see also Schwartz et al., 2005]. According to NRV, directional asymmetries should fade or disappear for native contrasts, but be maintained for non-native contrasts [see Polka and Bohn, 2011, for a review of studies with results that are consistent with NRV]. It is not yet clear whether or how NRV might apply to the perception of consonant contrasts, but these vowel-specific asymmetries would need to be accounted for by any general *or* vowel-focused model of cross-language speech perception.

Although it is not yet clear how PAM might account for directional asymmetries, the claim that directional asymmetries should be maintained for all non-native contrasts is not consistent with PAM. If each non-native phone in a contrast is assimilated to a different native category (TC assimilation) or one phone is assimilated and the other is not assimilated (UC assimilation), it is native-language phonological attunement that is responsible for the accurate discrimination. According to PAM, discrimination of a TC assimilation, in particular, would be tantamount to discrimination of a native contrast. The lack of perceptual asymmetry observed for native contrasts should therefore also be apparent for non-native contrasts that are assimilated as TC or UC, regardless of where they fall with respect to the centre versus the edges (periphery) of the vowel space. Asymmetries might only be observed under more limited circumstances, that is, when listeners are forced to rely on phonetic rather than phonological information as the basis for discrimination (SC, CG, or UU assimilations).

Testing between PAM and NRV is complicated by methodological differences among the studies cited in support of each model. Reports addressing the principles of PAM have tested discrimination using a categorial AXB task, in which three non-native phones are presented and the participant is required to indicate whether the second phone (X) is from the same category as the first (A) or third (B), whereas studies in support of NRV have employed various change detection paradigms. As highlighted by the Automatic Selective Perception model of non-native speech perception [ASP; Strange, 2011], those differences in stimulus task demand may affect listeners' ability to detect phonetic differences in non-native contrasts, so it is important to select a common task for our investigation. For consistency with previous PAM investigations, and to control for response bias [see Strange and Shafer, 2008, for a review of this issue for different discrimination tasks], we employed categorial AXB discrimination, used in tandem with categorization and goodness ratings of the stimulus tokens as the basis for determining the participants' assimilation patterns for each non-native vowel contrast. The four AXB trial types are AAB, ABB, BBA, and BAA. For the purpose of evaluating NRV predictions, we defined A as the more peripheral vowel in each contrast, and B as the less peripheral vowel of the contrast. Thus, AAB and ABB trials present a within-trial change from a more peripheral to a less peripheral vowel, and the BBA and BAA trials present a change in the opposite direction. If participants are able to detect a change in stimulus more often in the latter than the former case, consistent with the NRV hypothesis about the contribution of vowel peripherality to discrimination asymmetries, then this should be reflected in higher discrimination accuracy on BBA and BAA trials than on AAB and ABB trials.

Method

Participants

Thirteen university students from the northeast USA received a standard hourly payment for their participation in the experiment. They were all native speakers of American English with normal hearing and no history of speech or hearing problems. Selection criteria stipulated that volunteers should not have studied French, Norwegian, or Thai, or any of a range of languages that are known to have any of the target vowel contrasts.

Stimuli

The six contrasts chosen for this study were selected to provide maximal opportunity for observing SC, CG, and TC assimilations (and possibly also UC and/or UU assimilations) by American-English listeners unfamiliar with the target languages and contrasts. As no single language was likely to offer the full range of possible assimilation patterns for non-native, non-English vowel contrasts, we selected target contrasts from three languages that are members of different language groups (respectively from the Germanic, Romance and Tai language groups). We used two Norwegian contrasts, close front unrounded versus out-rounded /ki/-/ky/ and close front unrounded versus in-rounded (centralized) /ki/-/kʉ/, one Thai contrast, close back unrounded versus close-mid back unrounded /bɯ/-/bɤ/, and three French contrasts, oral versus nasal close-mid back rounded /bo/-/bõ/, close-mid front rounded versus open-mid front rounded /dø/-/dœ/, and close front rounded versus close-mid front rounded /sy/-/sø/. All were either minimal-pair words or were the beginning syllables of real words in the respective languages, to assure natural vowel pronunciations by the speakers. All contrasts were recorded by a female native speaker of the standard dialect of the source language (from Bergen in Norway, Brittany in France, and Bangkok in Thailand, respectively), digitized at 44.1 kHz using an Audiomedia board, and normalized for RMS amplitude. Six tokens of each syllable were chosen such that there was no overall difference between contrast pairs in duration or mean f0, and minimal variation in vowel intensity among the tokens. The selected tokens were verified by native speakers as acceptable exemplars of the target categories. Mean acoustic measurements across the six tokens of each syllable are reported in table 1.

To define A and B in the discrimination task in order to evaluate the predictions of NRV [Polka and Bohn, 2003, 2011], it is necessary to examine which of the vowels in each contrast is more peripheral than the other. Following Polka and Bohn [2011], the present stimulus vowels have been plotted on F1 and F2 axes in figure 1, with arrows pointing towards the more peripheral vowel (i.e., the one more proximal to the closest edge of vowel space). For both Norwegian contrasts, /i/ appears to be the more peripheral vowel,[2] whereas for the French contrasts /y/ is more peripheral than /ø/ in /sy/-/sø/, /ø/ is more peripheral than /œ/ in /dø/-/dœ/, and /o/ is more peripheral than /õ/. In the Thai contrast, /ɯ/ is more peripheral than /ɤ/.

[2] But see also the dispersion-focalization hypothesis, Schwartz et al. [2005], by which /y/ is essentially as focalized as /i/, even if not 'equally peripheral' in F1/F2 space. From this perspective, the in-rounded /y/ in Norwegian is less focal that the rounded /y/ in French [Schwartz et al., 1993], and because of scaling differences in the front and back part of the vocal tract for female and male speakers, /y/ is also less focal in French for female than male speakers [Schwartz et al., 1993].

Phonetica 2014;71:4–21
DOI: 10.1159/000356237

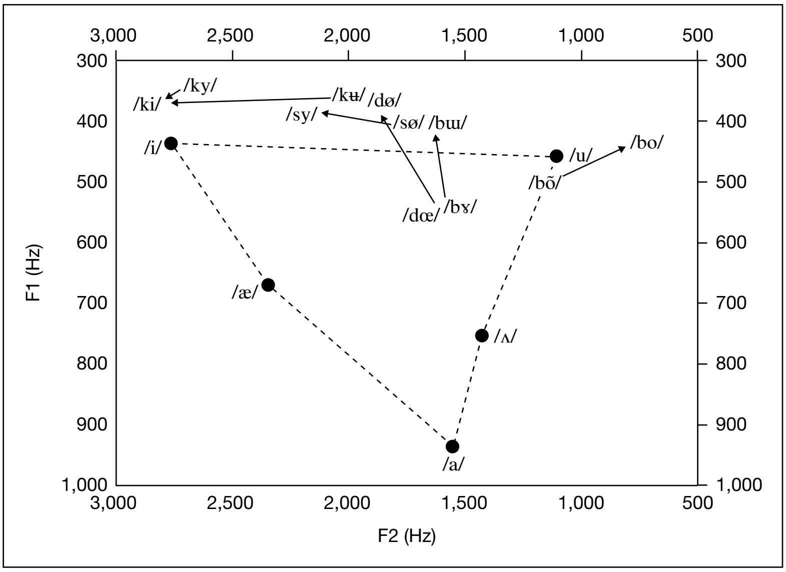

Fig. 1. Plot of mean F1/F2 frequencies for vowels in each of the syllables used in the experiment. Arrows point toward the hypothesized referent vowel for each contrast, i.e., the more peripheral vowel of the pair (more proximal to the closest edge of the vowel space), according to the NRV framework [Polka and Bohn, 2003, 2011]. For reference the American English corner vowels and the central vowel /ʌ/ are presented (connected by dashed lines), using the average formant frequencies of the 48 women reported in Hillenbrand et al. [1995].

Table 1. Mean acoustic measurements across the six tokens of each syllable used in the experiment

Syllable	Consonant duration ms	Vowel					
		duration ms	intensity dB	f0 Hz	F1 Hz	F2 Hz	F3 Hz
French							
/bo/	59	337	81	190	430	725	2,718
/bõ/	73	362	78	192	512	1,140	2,059
/dø/	102	316	83	185	381	1,859	2,593
/dœ/	93	307	81	183	547	1,619	2,786
/sy/	230	311	81	216	382	2,186	2,723
/sø/	231	310	83	205	402	1,754	2,573
Norwegian							
/ki/	71	255	84	259	373	2,833	3,423
/ky/	69	248	86	258	340	2,671	3,129
/kʉ/	71	256	86	255	364	2,012	2,844
Thai							
/bɯ/	58	243	88	202	411	1,626	2,829
/bɤ/	55	237	85	202	535	1,580	2,835

Formant values were measured at 50% of the duration of the vowel.

Procedure

Following previous studies investigating PAM predictions with consonants [e.g., Best and Strange, 1992; Best et al., 1988], participants first completed a categorial AXB discrimination test for each of the six contrasts. The order of presentation was: /bo/-/bõ/, /ki/-/kʉ/, /sy/-/sø/, /bɯ/-/bɤ/, /dø/-/dœ/, and /ki/-/ky/.[3] In this procedure, A and B are tokens of the contrasting non-native phonemes, and listeners are asked to indicate on their answer sheet whether the middle item (X) is the same as the first (A) or third (B) item. The X token was always physically different from the categorially matching A or B item, so that listeners could not make a simple acoustic identity judgement [e.g., Best et al., 1988; Polka, 1991, 1992].

Each of the six AXB tests contained 48 trials in eight-trial blocks (interstimulus interval = 1 s; intertrial interval = 6 s). This was the interstimulus interval used in previous PAM reports [Best and Strange, 1992; Best et al., 1988, 2001]. The four AXB trial types, AAB, ABB, BBA, and BAA, were equally represented for each contrast, and within each test the trial order was randomized. Each of the six tokens per stimulus set occurred twice in each position (first, second, or third) in each trial type.

Following the discrimination tests, participants performed a keyword identification task for each of the six contrasts. In each test there were 40 trials consisting of four warm-up items (two from each contrast category), followed by 18 repetitions of each category in the contrast pair (three repetitions per token), presented in random order. On a given trial, the participants listened to a single token of each speech category, played twice. After the first presentation they circled an English keyword corresponding to the vowel that they heard in the item. The token was then repeated and they rated the similarity of the vowel in the auditory stimulus to the vowel in the chosen keyword (on a scale of 1 *unlike* to 5 *identical*). The keywords were chosen to provide not only the full range of vowel possibilities, but also contexts in which the participant could indicate phonetic properties such as nasality. The vowel categories for /i/, /ɪ/, /e/, /ɛ/, /æ/, /ɑ/, /ɔ/, /o/, /ʊ/, /ʌ/, /u/, and /ɝ/ (listeners were all from rhotic accents of American English) were represented by the keywords HEED, HID, AID, ED, AD, ODD, AWED, HOED, HOOD, DUD, FOOD, and HEARD, respectively. The keywords END, DONNED, and OWNED were also included to allow the participants to indicate similarity to the nasalized vowel allophones in /ɛn/, /ɑn/, and /on/, respectively. DUDE, as opposed to FOOD, was included as well because choice of this word might reflect sensitivity to the fronting of the close back rounded English vowel /u/ that occurs after coronal stops.

Stimuli were presented via an Otari MX5050 BQ-II reel-to-reel tape deck connected to a Kenwood amplifier, which fed to a Jamo compact loudspeaker. The speaker was centred on the opposite side of the table, facing the participant at a distance of approximately 1 m. Output from the loudspeaker was set to 70 ± 3 dB, as measured from the participants' location.

Results

Keyword Categorization Results

As there were no systematic differences between the choice of DUDE versus FOOD across participants, these keywords were combined and treated as /u/. The mean percentage of keyword selections for each stimulus syllable is presented in table 2, along with the mean category goodness ratings for those selections. Note that these values were obtained first by averaging each participant's ratings for that stimulus/keyword pairing, and then averaging the scores for all participants who chose that

[3] As it would have required a large number of presentation orders to counteract position effects and serial order effects among the six contrasts, we elected instead to keep the order constant. The order was chosen so that contrasts from the same language were separated by at least one contrast from a different language. The results do not appear to have been systematically affected by the fixed order. That is, from the percent correct discrimination scores in table 4, there appears to be no performance advantage (or special disadvantage) for the first contrast, and the poorer performance for the final contrast is unlikely to be due to a fatigue effect, as there is no evidence of a gradual decline in performance across the discrimination subtests, and no evidence of decreasing performance across the subsequent categorization subtests.

Table 2. Mean percent categorization and goodness rating (in parentheses) of vowel stimuli in terms of American English vowel categories and the response keywords

Contrast	Syllable	Vowel category															
		/i/ HEED	/ɪ/ HID	/e/ AID	/ɛ/ ED	/æ/ AD	/ɑ/ ODD	/ɔ/ AWED	/o/ HOED	/ʊ/ HOOD	/ʌ/ DUD	/u/ FOOD	/ɝ/ HEARD	/ɛn/ END	/on/ OWNED	/ɑn/ DONNED	missing
/dø/-/dœ/																	
	/dø/			11.1 (2.7)		1.3 (3.3)		1.3 (2.8)		*16.2* *(2.8)*	3.4 (3.1)	**50.9** **(3.6)**	15.4 (4.1)				0.4
	/dœ/			3.0 (2.3)	0.4 (3.0)	1.7 (3.0)				3.9 (3.6)	**85.5** **(3.8)**	1.7 (3.0)	1.3 (3.3)			0.9 (2.5)	1.7
/bo/-/bõ/																	
	/bo/							1.7 (3.8)	**44.0** **(3.6)**	6.0 (3.0)		*38.0* *(3.3)*			6.8 (3.9)	2.1 (2.1)	1.3
	/bõ/						6.0 (3.4)	5.1 (3.2)	6.4 (3.1)	3.9 (3.4)	2.6 (3.1)	*22.7* *(2.8)*			19.2 (3.0)	**33.8** **(3.2)**	0.4
/sy/-/sø/																	
	/sy/	**51.3** **(3.3)**	1.2 (3.5)							6.4 (2.9)		*32.5* *(3.3)*	7.3 (1.6)	0.4 (2.0)			0.9
	/sø/	0.4 (4.0)	0.9 (2.0)	11.5 (3.3)	9.8 (3.6)					*20.1* *(2.9)*	12.8 (2.8)	13.7 (2.6)	**21.4** **(3.1)**	8.1 (2.0)			1.3
/ki/-/kʉ/																	
	/ki/	**87.2** **(4.0)**	8.6 (2.5)							0.4 (3.0)		0.4 (3.0)		0.9 (2.0)		0.4 (1.0)	2.1
	/kʉ/	5.6 (2.8)	1.7 (1.7)	1.7 (2.0)	3.0 (2.7)					*17.1* *(3.5)*	1.3 (1.7)	**61.1** **(3.4)**		4.7 (2.9)	0.9 (2.5)	1.3 (2.0)	1.7
/ki/-/ky/																	
	/ki/	**95.7** **(3.9)**	3.9 (2.9)														0.4
	/ky/	**82.9** **(3.6)**	9.8 (2.9)									6.8 (3.1)					0.4
/bɯ/-/bɤ/																	
	/bɯ/		3.0 (1.1)		8.1 (2.6)	0.9 (2.5)	0.9 (2.5)	0.4 (3.0)	15.4 (2.7)	*21.8* *(2.8)*	**29.1** **(3.1)**	9.8 (2.1)	1.7 (3.0)		5.6 (2.4)	1.3 (2.3)	2.1
	/bɤ/		1.28 (1.33)		*12.0* *(2.0)*	4.3 (3.4)	2.1 (2.0)	6.4 (3.2)	3.0 (3.5)	9.8 (2.5)	**55.6** **(3.1)**	0.4 (1.0)	0.4 (4.0)		2.1 (2.3)	0.4 (1.0)	2.1

Boldfaced values indicate the most frequently chosen identification response per target; italicized values indicate next most frequent choices when the most frequent choice did not reach 70%. The goodness ratings are based on a scale that ranged from 1 (unlike) to 5 (identical).

stimulus/keyword pairing. A non-native vowel was deemed to be categorized if the same native vowel category was chosen for a given target vowel more than 70% of the time [following Antoniou et al., 2012, and Bundgaard-Nielsen et al., 2011; see Harnsberger, 2001, for a discussion of the relative merits of higher and lower categorization criteria]. We also analysed our results using a 50% categorization criterion and the same pattern of results was obtained.

Examination of individual participants' data reveals large individual differences. In some cases, individual participants consistently chose a native vowel category that was different from the rest of the cohort, such that when averaged across participants the mean percent for that category was low even though it was high for some individual participants. Using the data averaged across participants, shown in table 2, may therefore cloud the analysis of AXB discrimination and the testing of PAM hypotheses. The only viable analysis was to determine each individual participant's assimilation pattern for each contrast. As participants categorized and rated each syllable 18 times, there were sufficient data points to perform this analysis. If no single native vowel category was chosen above 70% of the time then the non-native vowel was considered to be uncategorized. If both members of the contrast were categorized to the same native vowel category, we compared the participant's category goodness ratings using a t test to determine assimilation type. A significant difference in goodness ratings in such cases indicates a CG assimilation, otherwise the contrast is an SC assimilation.

The analysis of individual participants' vowel contrast assimilation patterns is presented in table 3. It can be seen that SC assimilations were found only for the /ki/-/ky/ contrast (n = 7) and CG assimilations were spread among /ki/-/ky/, /bɯ/-/bɤ/, and /bo/-/bõ/ (n = 6). The remainder were TC (n = 18), UC (n = 26), and UU (n = 21).

AXB Discrimination Analyses

Overall mean percent correct responses for each contrast are presented in table 4. Participants performed above chance for all contrasts (see table 4 for results of one-sample t tests against a chance score of 50%), and achieved perfect discrimination of /ki/-/kʉ/. In fact, discrimination was excellent (>90%) for all contrasts except /ki/-/ky/, for which discrimination was fairly poor (~70%). Given the individual differences observed in PAM assimilation types, it is not possible to evaluate PAM predictions based on the overall mean discrimination accuracy for each contrast. Instead, we grouped the mean percent discrimination scores based on each individual's assimilation type rather than on vowel contrast. For example, the data for all of the individual CG assimilations were grouped together, regardless of the contrast for which they were observed.

As PAM predicts very good or excellent discrimination for the two assimilation types that involve a native phonological contrast, TC and UC, we grouped them together, which yielded three categories of assimilation: (1) SC, (2) CG, and (3) those assimilations that cross a phonological boundary, that is, TC and UC. UU assimilations were not included in the analysis because PAM does not make a clear prediction about accuracy of discriminating those contrasts (it could range from poor to very good, depending on the phonetic distance or the proximity to native categories). Discrimination of UU contrasts was close to ceiling (mean = 97.22%) and there was no effect of peripherality (as we defined it earlier for the AXB trial types).

The results for the remaining assimilation types were analysed using non-orthogonal planned contrasts. Two planned contrasts were used to compare differences among

Table 3. Individual contrast assimilation types for each participant on each contrast at a 70% assimilation criterion

	Contrast					
	/ki/-/ky/	/dø/-/dœ/	/bɯ/-/bɤ/	/sy/-/sø/	/ki/-/kʉ/	/bo/-/bõ/
Participant						
1	SC	UC	UC	TC	TC	UU
2	SC	UC	UC	TC	UC	UC
3	CG	UC	UU	UC	UC	UU
4	CG	TC	UU	UC	UC	UU
5	CG	TC	UU	UU	TC	UU
6	CG	TC	UC	TC	UC	UC
7	UC	TC	CG	UC	UU	UC
8	TC	UU	UC	UU	TC	TC
9	SC	TC	UC	UU	UU	UU
10	SC	TC	UC	UC	TC	UU
11	SC	UU	UU	UC	UU	UU
12	SC	TC	TC	TC	UU	CG
13	SC	UC	UC	UC	UC	UU
Frequency of individual assimilation types observed per target contrast (%)						
SC	**54**	0	0	0	0	0.0
CG	31	0	8	0	0	8
TC	8	**54**	8	31	31	8
UC	8	31	**54**	**46**	**39**	23
UU	0	15	31	23	31	**62**

Boldfaced values in the frequency distribution subtable at the bottom indicate the most frequent assimilation pattern per target contrast (each column sums to 100%).

Table 4. Mean percent correct discrimination for each contrast in the experiment, listed in presentation order, and results of one-sample t tests against a chance score of 50%

	Discrimination		Test versus chance	
	mean percent correct	SE	t(12)	95% CI
/bo/-/bõ/	96.31	1.95	23.78	92.07–100.56
/ki/-/kʉ/	100.00	0.00	*	
/sy/-/sø/	99.20	0.44	110.88	98.23–100.17
/bɯ/-/bɤ/	95.51	1.24	36.56	92.80–98.22
/dø/-/dœ/	96.47	0.93	50.23	94.46–98.49
/ki/-/ky/	72.76	3.12	7.28	65.95–79.56

All p values <0.001. No t test could be performed for /ki/-/kʉ/ because all participants obtained perfect scores. SE = Standard error of the mean.

the assimilation types. The *phonological status* planned contrast compared cross-boundary assimilation types (TC and UC) against the within-category assimilation type that was predicted by PAM to have the highest discrimination accuracy (CG), while the *phonetic goodness* planned contrast compared the two within-category assimilation types (CG and SC) predicted by PAM to show different levels of discrimination.

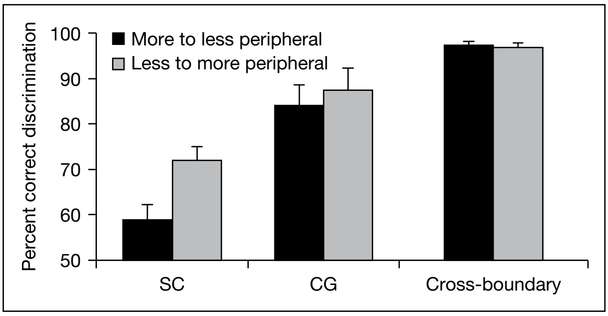

Fig. 2. Mean percent correct discrimination scores for contrasts falling into SC, CG, and cross-boundary (TC and UC) assimilation types. Error bars represent standard error of the mean.

The effect of vowel peripherality was assessed using an additional repeated-measures contrast, *direction of change*, comparing AXB trials involving a change from a more peripheral to less peripheral vowel (AAB and ABB) and those involving a change from a less peripheral to a more peripheral vowel (BAA and BBA). A Bonferroni-adjusted α rate of 0.025 was used. The mean percent correct discrimination scores, by assimilation type, are presented in figure 2.

The phonological status planned contrast was significant, $F(1, 54) = 14.77$, $mean_{difference} = 11.25\%$, SE = 2.93%, 97.5% confidence interval (CI): 4.50–18.01%, showing that cross-boundary assimilation types were discriminated more accurately than CG assimilations. The phonetic goodness contrast was also significant, $F(1, 54) = 29.37$, $mean_{difference} = 20.29\%$, SE = 3.74%, 97.5% CI: 12.78–27.79%, such that CG assimilations were discriminated significantly more accurately than SC assimilations. The repeated measures contrast showed that responses on trials in which a more peripheral vowel changed to a less peripheral vowel (AAB and ABB trials) were significantly less accurate, overall, than on trials in which a less peripheral vowel changed to a more peripheral vowel (BBA and BAA trials), $F(1, 54) = 25.78$, $mean_{difference} = 5.37\%$, SE = 1.06%, 97.5% CI: 2.93–7.80%. A significant interaction between phonetic goodness and direction of change, $F(1, 54) = 9.89$, contrast mean = 9.62%, SE = 3.06%, 97.5% CI: 2.57–16.78%, reflected a greater influence of the direction of change on accuracy for SC than CG assimilations. There was no interaction between the phonological status contrast and direction of change, suggesting that there was little effect of vowel peripherality on CG and TC/UC contrasts. To confirm this we conducted separate paired t tests on the discrimination data for SC, CG, and TC/UC contrasts. Less-to-more peripheral vowel changes were discriminated significantly more accurately than more-to-less changes for SC contrasts, $t(6) = 3.67$, $p = 0.01$, but not for CG, $t(5) = 0.58$, or TC/UC, $t(64) = 0.39$.

Discussion

This project examined American English monolinguals' perception of six cross-language vowel contrasts, chosen to yield a wide range of assimilation patterns with respect to the English vowel system. Our primary purpose was to determine whether perception of non-native vowel contrasts appears to follow the same principles as have been observed with non-native consonant contrasts, and in particular to evaluate predictions from the PAM and NRV models.

The results revealed high levels of interindividual variability in the assimilation patterns for non-native vowels. In light of this variability, we took the novel approach of analysing according to each individual participant's contrast assimilations. The majority of vowel contrasts were assimilated as TC, UC, or UU. SC assimilations were only found for the /ki/-/ky/ contrast. To test the PAM prediction that discrimination would be most accurate for TC assimilations, followed by CG, and then SC, we grouped discrimination data based on assimilation type rather than on vowel contrast. That analysis allowed us to compare discrimination of cross-boundary assimilation types (TC and UC) to that for within-category assimilation types (CG and SC). The results confirmed our extrapolation of PAM predictions about relations between assimilation patterns and differences in discrimination performance, as applied to non-native vowel perception (i.e., TC/UC > CG > SC). Furthermore, the results indicate that TC, CG, SC, UU, and UC assimilation types can all be obtained for non-native vowel contrasts, but assimilation type varies among individuals for any given non-native vowel contrast.

The analysis of AXB discrimination also tested whether the NRV peripherality prediction, that a change from a less to a more peripheral non-native vowel is easier to detect than a change in the opposite direction, holds true for all PAM assimilation types. We had predicted that asymmetries would not be observed for TC and UC assimilations, and that they might only be observed for SC, CG, or UU assimilations. We did not include UU contrasts in our main analysis, however, because the PAM discrimination predictions are dependent on phonetic distance between the assimilated phones, which could not be determined for the UU assimilation types as determined for individual listeners across the six non-native contrasts. It should be noted, nonetheless, that discrimination for UU was at ceiling and thus no perceptual asymmetries were observed for that assimilation type.

Consistent with our predictions, discrimination of TC/UC assimilations was at ceiling and there was no effect of peripherality. There was a clear effect of peripherality in the direction predicted by the NRV framework for SC assimilations, but there was no significant effect of peripherality for CG assimilations. Therefore, although we predicted that peripherality effects might occur for both SC and CG assimilations, it appears that they only occurred here for SC assimilations, which in this study were observed only for the /ki/-/ky/ contrast, for a slim majority of participants (~54%, table 3).

One limitation of these findings is the number of observations that were included in the analysis of SC versus CG assimilations (7 data points for SC and 6 data points for CG). The effect for SC appears to have been large enough to overcome the small number of individual cases of this assimilation type, but it is possible that there remains a more subtle effect of peripherality for CG assimilations that we did not have sufficient power to detect. This is supported by the observation that Polka and Bohn [2011] reported an asymmetry for German /u/-/y/ and /ʊ/-/ʏ/, using a change detection task, which are assimilated as either CG or UC by English listeners. Although our participant sample size was consistent with previous studies on cross-language vowel perception [e.g., Beddor and Strange, 1982; Flege et al., 1997; Levy, 2009a], it appears that much larger participant samples may be required to overcome the interindividual differences in assimilation types per vowel contrast that we have observed in this study. Future research should therefore target non-native contrasts that are likely to result in SC and CG assimilations. Based on our results, we tentatively suggest that if natural referent vowels are perceptual anchors for adult listeners, as proposed by Polka and Bohn

[2011], then the important factor limiting their influence is not only whether the vowel contrast is non-native, but specifically whether the contrasting non-native phones are assimilated to a single native vowel, either equally or with a category goodness difference. Discrimination asymmetries appear unlikely for non-native vowel contrasts that are assimilated as a native phonological distinction (or as multiple native vowels, i.e., uncategorized with respect to any single native vowel).

PAM was devised to account for the influence of native-language attunement on speech perception in adults and developing infants. Vowel perception asymmetries could be considered to arise as a result of universal phonetic perception, and they may play an important role in the development of native vowel categories. According to Polka and Bohn [2011], the more peripheral referent vowels act as perceptual anchors to facilitate the development of less peripheral vowel categories. By this account, the 'corner vowels' /a/, /i/, and /u/ play an important role in native-language attunement because they are universally available across the world's languages [Polka and Bohn, 2011, discuss how lip-rounded vowels may also fit into the NRV framework]. We would note, as well, that those vowels reflect two intersecting edges of the vowel space. Just as peripherality effects appear to decline as the infant attunes to native-language contrasts, according to PAM they should also decline for non-native contrasts that are assimilated as TC or UC. That is, attunement to native phonological *contrasts* takes precedence over universal phonetic perception.

Discrimination was very accurate for five out of six of the non-native vowel contrasts, with all but /ki/-/ky/ at 95% correct or higher. Discrimination of /ki/-/ky/ was fair (73% correct), but the analysis by individual participant assimilations showed that this average value combined the data of participants who assimilated the contrast as SC, CG, UC, or TC. When the seven SC assimilations were analysed on their own, the mean discrimination accuracy was around 65% (significantly above a chance score of 50%, $t(6) = 5.94$, $p = 0.001$), which is consistent with the PAM prediction that discrimination of SC contrasts will be poor [but possibly somewhat above chance; Best, 1995]. Thus, while it may appear that the overall discrimination of vowel contrasts is generally higher than for consonants, this may be due, in part, to the individual differences observed in assimilation type. The discrimination accuracy observed in the individual participant analysis was consistent with PAM: SC was poor, CG was on the upper end of good to very good, and TC/UC was very good to excellent.

Examination of table 2 revealed some interesting patterns of categorization across the vowel contrasts. The French /bo/-/bõ/ contrast was predominantly assimilated as UU, and overall discrimination was excellent, at over 96% correct. The categorization responses to /bo/ were largely split between English /o/ and /u/, and those for /bõ/ among /u/, /on/, and /ɑn/. This suggests that a non-native vowel that falls in an untuned region of the vowel space (i.e., where actual native vowel tokens are rarely encountered) may be perceived as an unusual-sounding version of more than one nearby native vowel category. Here we raise the possibility [following similar suggestions for L2 speech perception by Levy, 2009a, and Tyler, 2007] that the instability observed in categorization of these two non-native French vowels could be due to some overlap in the native vowel categories to which each was assimilated. When two non-native vowels are uncategorized, we might expect better discrimination if there is no overlap in the native vowel categories to which the two are assimilated, than if the two non-native vowels show some overlapping use of the same native category (or categories), as was seen in the use of /u/ as one of the English assimilation categories for both /bõ/

and /bo/. This is because, in the non-overlap case, listeners would be able to detect phonological differences between one group of categories and the other group, and thus discrimination should be excellent.

Some evidence for this possibility can be found with the /sy/-/sø/ and /bɯ/-/bɤ/ contrasts. As can be seen in table 2, /sy/-/sø/ was primarily non-overlapping whereas the /bɯ/-/bɤ/ contrast was highly overlapping, with /ʌ/ being the modal response choice for both vowels. Although the discrimination results should be interpreted with caution, given the near-ceiling performance, they are consistent with our logic about the effects of assimilation category overlap in the case of uncategorized non-native phones. Specifically, there were 11 UC/UU assimilations for /sy/-/sø/, which were discriminated significantly more accurately than the 9 UC/UU assimilations for /bɯ/-/bɤ/ (99.3 vs. 96.2%, respectively), as confirmed by a t test [unequal variances assumed; $t(13.815) = 2.61$, $p = 0.021$]. Compatible with this reasoning, better discrimination for non-overlapping than for overlapping native-language phonological categories in UC and UU assimilations has also been observed in Italian and Danish listeners' perception of several non-native English consonant contrasts [Bohn et al., 2011].[4]

Additional contributions of this report, which may guide future research, were the predictions generated from NRV for assessment in AXB tasks and our grouping of non-native contrasts based on categorization results for each individual participant. The finding that individuals differ in their assimilation of vowels is also important for studies of L2 learning from a PAM-L2 [Best and Tyler, 2007] perspective. As PAM-L2 predictions of L2 development are based on PAM assimilation types at the time of L2 immersion, individual assimilation patterns rather than group patterns may form a more solid basis for PAM-L2 investigations, especially with L2 vowel contrasts but possibly also with L2 consonants. Indeed, one other model of L2 speech perceptual learning, Second Language Linguistic Perception, has taken seriously the importance of evaluating individual differences in learning-related changes in L2 vowel categorizations [Escudero and Boersma, 2004; Escudero et al. 2009]. Furthermore, if the vowel peripherality asymmetry observed here for SC assimilations continues to be observed for other vowel contrasts, it suggests that L2 learners may benefit from a training paradigm in which they learn to detect the more peripheral vowel from among a repeating sequence of less-peripheral vowel tokens.

We conclude that PAM principles do extend beyond non-native consonants, to non-native vowels, and that perceptual asymmetries due to vowel peripherality do not occur for non-native vowel contrasts that cross a native phonological boundary (PAM TC and UC assimilation types). NRV predictions were upheld for SC assimilation types, and converging evidence from Polka and Bohn [2011] suggests that they may also be found for CG assimilation types. Regarding the latter, we may not have had sufficient statistical power to detect a weaker effect of peripherality on CG assimilations. Together, this suggests that the effect of peripheral vowels on adult non-native vowel discrimination may only be apparent when there is no influence of native phonological

[4] The difference in discrimination performance could be attributed to differences in auditory distance rather than category overlap. The F1 and F2 distances between /sy/ and /sø/ are 0.19 Bark and 1.48 Bark, respectively, whereas for /bɯ/ and /bɤ/ the distances are 1.10 Bark and 0.19 Bark, respectively. However, we favour our explanation based on perceptual assimilation because auditory distances could not account for our /ki/-/ky/ results, where individual differences in categorization correspond to systematic differences in discrimination performance. That is, participants who categorized /ki/-/ky/ as an SC contrast performed significantly more poorly than those who categorized it one of the other assimilation types [65 vs. 81%, $t(11) = 3.51$, $p = 0.005$].

distinctions in the perceptual assimilations to the native vowel system (as in CG, SC, and UU assimilations). With the recent theoretical developments of NRV and the ASP model [Strange, 2011], and our confirmation here that PAM's predictions also apply to vowels, it is essential to obtain additional data that will allow the models' predictions to be compared directly. For such comparisons, it is crucial to test both categorization and discrimination, and for NRV there must be sufficient deviation of discrimination from ceiling performance levels to allow for testing directional asymmetries. We look forward to future studies that will test the generalizability of both models across multiple languages, listener groups, and phonetic contexts.

Acknowledgements

This research was supported by NIH grant No. DC00403 (PI: C. Best) and Australian Research Council Discovery grant DP0880913 (CI: M. Tyler). We thank Janet Calderón and Mark Antoniou for research assistance, and Jean-Luc Schwartz and Linda Polka for helpful comments on a previous version of this article.

References

Antoniou, M.; Tyler, M.D.; Best, C.T.: Two ways to listen: do L2-dominant bilinguals perceive stop voicing according to language mode? J. Phonet. *40:* 582–594 (2012).

Beddor, P.S.; Strange, W.: Cross-language study of perception of the oral-nasal distinction. J. acoust. Soc. Am. *71:* 1551–1561 (1982).

Best, C.T.: Emergence of language-specific constraints in perception of non-native speech: a window on early phonological development; in Boysson-Bardies, Schonen, Jusczyk, Mac-Neilage, Morton, Developmental neurocognition: speech and face processing in the first year of life, pp. 289–304 (Kluwer Academic Publishers, Dordrecht 1993).

Best, C.T.: Learning to perceive the sound pattern of English; in Rovee-Collier, Lipsitt, Adv. Infancy Res., vol. 9, pp. 217–304 (Ablex, Norwood 1994a).

Best, C.T.: The emergence of native-language phonological influences in infants: a perceptual assimilation model; in Goodman, Nusbaum, The development of speech perception: the transition from speech sounds to spoken words, pp. 167–244 (MIT Press, Cambridge 1994b).

Best, C.T.: A direct realist view of cross-language speech perception; in Strange, Speech perception and linguistic experience: issues in cross-language research, pp. 171–204 (York Press, Baltimore 1995).

Best, C.T.; Hallé, P.A.: Perception of initial obstruent voicing is influenced by gestural organization. J. Phonet. *38:* 109–126 (2010).

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:* 775–794 (2001).

Best, C.T.; McRoberts, G.W.; Sithole, N.M.: Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. J. exp. Psychol. hum. Percept. Perform. *14:* 345–360 (1988).

Best, C.T.; Strange, W.: Effects of phonological and phonetic factors on cross-language perception of approximants. J. Phonet. *20:* 305–330 (1992).

Best, C.T.; Tyler, M.D.: Nonnative and second-language speech perception: commonalities and complementarities; in Munro, Bohn, Second language speech learning: the role of language experience in speech perception and production, pp. 13–34 (Benjamins, Amsterdam 2007).

Best, C.T.; Tyler, M.D.; Gooding, T.N.; Orlando, C.B.; Quann, C.A.: Development of phonological constancy: toddlers' perception of native- and Jamaican-accented words. Psychol. Sci. *20:* 539–542 (2009).

Bohn, O.S.; Best, C.T.: Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners. J. Phonet. *40:* 109–128 (2012).

Bohn, O.-S.; Best, C.T.; Avesani, C.; Vayra, M.: Perceiving through the lens of native phonetics: Italian and Danish listeners' perception of English consonant contrasts; in Lee, Zee, Proc. 17th Int. Congr. Phonet. Sci., pp. 336–339 (Department of Chinese, Translation and Linguistics, City University of Hong Kong 2011).

Bundgaard-Nielsen, R.L.; Best, C.T.; Tyler, M.D.: Vocabulary size is associated with second-language vowel perception performance in adult learners. Stud. second Lang. Acquis. *33:* 433–461 (2011).

Escudero, P.; Benders, T.; Lipski, S.C.: Native, non-native and L2 perceptual cue weighting for Dutch vowels: the case of Dutch, German, and Spanish listeners. J. Phonet. *37:* 452–465 (2009).

Escudero, P.; Boersma, P.: Bridging the gap between L2 speech perception research and phonological theory. Stud. second Lang. Acquis. *26:* 551–585 (2004).
Escudero, P.; Williams, D.: Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners. J. acoust. Soc. Am. *129:* EL1–EL7 (2011).
Flege, J.E.: Second language speech learning: theory, findings, and problems; in Strange, Speech perception and linguistic experience: issues in cross-language research, pp. 233–276 (York Press, Baltimore 1995).
Flege, J.E.: Interactions between the native and second-language phonetic systems; in Burmeister, Piske, Rohde, An integrated view of language development: papers in honor of Henning Wode, pp. 217–243 (Wissenschaftlicher Verlag, Trier 2002).
Flege, J.E.; Bohn, O.-S.; Jang, S.: Effects of experience on non-native speakers' production and perception of English vowels. J. Phonet. *25:* 437–470 (1997).
Frieda, E.M.; Walley, A.C.; Flege, J.E.; Sloane, M.E.: Adults' perception of native and nonnative vowels: implications for the perceptual magnet effect. Percept. Psychophys. *61:* 561–577 (1999).
Fry, D.B.; Abramson, A.S.; Eimas, P.D.; Liberman, A.M.: The identification and discrimination of synthetic vowels. Lang. Speech *5:* 171–189 (1962).
Gibson, J.J.; Gibson, E.J.: Perceptual learning: differentiation or enrichment? Psychol. Rev. *62:* 32–41 (1955).
Hallé, P.A.; Best, C.T.: Dental-to-velar perceptual assimilation: a cross-linguistic study of the perception of dental stop+/l/ clusters. J. acoust. Soc. Am. *121:* 2899–2914 (2007).
Hallé, P.A.; Best, C.T.; Levitt, A.: Phonetic vs. phonological influences on French listeners' perception of American English approximants. J. Phonet. *27:* 281–306 (1999).
Hallé, P.A.; Segui, J.; Frauenfelder, U.; Meunier, C.: Processing of illegal consonant clusters: a case of perceptual assimilation? J. exp. Psychol. hum. Percept. Perform. *24:* 592–608 (1998).
Harnsberger, J.D.: On the relationship between identification and discrimination of non-native nasal consonants. J. acoust. Soc. Am. *110:* 489–503 (2001).
Hillenbrand, J.; Getty, L.A.; Clark, M.J.; Wheeler, K.: Acoustic characteristics of American English vowels. J. acoust. Soc. Am. *97:* 3099–3111 (1995).
Kuhl, P.K.: Human adults and human infants show a 'perceptual magnet effect' for the prototypes of speech categories, monkeys do not. Percept. Psychophys. *50:* 93–107 (1991).
Kuhl, P.K.: Speech prototypes: studies on the nature, function, ontogeny and phylogeny of the 'centers' of speech categories; in Tohkura, Vatikiotis-Bateson, Sagisaka, Speech perception, production and linguistic structure, pp. 240–264 (IOS Press, Tokyo 1992).
Ladefoged, P.: Vowels and consonants: an introduction to the sounds of languages; 2nd ed. (Blackwell, Malden 2005).
Levy, E.S.: On the assimilation-discrimination relationship in American English adults' French vowel learning. J. acoust. Soc. Am. *126:* 2670–2682 (2009a).
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:* 1138–1152 (2009b).
Levy, E.S.; Strange, W.: Perception of French vowels by American English adults with and without French language experience. J. Phonet. *36:* 141–157 (2008).
Lively, S.E.; Pisoni, D.B.: On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception. J. exp. Psychol. hum. Percept. Perform. *23:* 1665–1679 (1997).
Lotto, A.J.; Kluender, K.R.; Holt, L.L.: Depolarizing the perceptual magnet effect. J. acoust. Soc. Am. *103:* 3648–3655 (1998).
Polka, L.: Cross-language speech perception in adults: phonemic, phonetic, and acoustic contributions. J. acoust. Soc. Am. *89:* 2961–2977 (1991).
Polka, L.: Characterizing the influence of native experience on adult speech perception. Percept. Psychophys. *52:* 37–52 (1992).
Polka, L.: Linguistic influences in adult perception of non-native vowel contrasts. J. acoust. Soc. Am. *97:* 1286–1296 (1995).
Polka, L.; Bohn, O.-S.: A cross-language comparison of vowel perception in English-learning and German-learning infants. J. acoust. Soc. Am. *100*: 577–592 (1996).
Polka, L.; Bohn, O.-S.: Asymmetries in vowel perception. Speech Commun. *41:* 221–231 (2003).
Polka, L.; Bohn, O.-S.: Natural Referent Vowel (NRV) framework: an emerging view of early phonetic development. J. Phonet. *39:* 467–478 (2011).
Polka, L.; Colantonio, C.; Sundara, M.: A cross-language comparison of /d/-/ð/ perception: evidence for a new developmental pattern. J. acoust. Soc. Am. *109:* 2190–2201 (2001).
Schwartz, J.-L.; Abry, C.; Boë, L.-J.; Vallée, N.: Asymmetries in vowel perception, in the context of the Dispersion-Focalisation Theory. Speech Commun. *45:* 425–434 (2005).
Schwartz, J.L.; Beautemps, D.; Abry, C.; Escudier, P.: Interindividual and cross-linguistic strategies for the production of the [i] vs [y] contrast. J. Phonet. *21:* 411–425 (1993).
Strange, W.: Automatic selective perception (ASP) of first and second language speech: a working model. J. Phonet. *39:* 456–466 (2011).
Strange, W.; Akahane-Yamada, R.; Kubo, R.; Trent, S.A.; Nishi, K.: Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners. J. acoust. Soc. Am. *109:* 1691–1704 (2001).

Strange, W.; Bohn, O.-S.; Trent, S.A.; Nishi, K.: Acoustic and perceptual similarity of North German and American English vowels. J. acoust. Soc. Am. *115:* 1791–1807 (2004).
Strange, W.; Levy, E.S.; Law II, F.F.: Cross-language categorization of French and German vowels by naive American listeners. J. acoust. Soc. Am. *126:* 1461–1476 (2009).
Strange, W.; Shafer, V.L.: Speech perception and second language learners: the re-education of selective perception; in Hansen Edwards, Zampini, Phonology and second language acquisition, pp. 159–198 (Benjamins, Amsterdam 2008).
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:* 539–552 (1995).
Tyler, M.D.: Perceptual assimilation of ambiguous /r/-/l/ speech tokens. New Sounds 2007: 5th Int. Symp. on Acquis. Second Lang. Speech, Florianopolis 2007.