Perception of familiar contrasts in unfamiliar positions[a)]

Mirjam Broersma[b)]
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

(Received 24 August 2004; revised 21 March 2005; accepted 22 March 2005)

This paper investigates the perception of non-native phoneme contrasts which exist in the native language, but not in the position tested. Like English, Dutch contrasts voiced and voiceless obstruents. Unlike English, Dutch allows only voiceless obstruents in word-final position. Dutch and English listeners' accuracy on English final voicing contrasts and their use of preceding vowel duration as a voicing cue were tested. The phonetic structure of Dutch should provide the necessary experience for a native-like use of this cue. Experiment 1 showed that Dutch listeners categorized English final /z/–/s/, /v/–/f/, /b/–/p/, and /d/–/t/ contrasts in nonwords as accurately as initial contrasts, and as accurately as English listeners did, even when release bursts were removed. In experiment 2, English listeners used vowel duration as a cue for one final contrast, although it was uninformative and sometimes mismatched other voicing characteristics, whereas Dutch listeners did not. Although it should be relatively easy for them, Dutch listeners did not use vowel duration. Nevertheless, they attained native-like accuracy, and sometimes even outperformed the native listeners who were liable to be misled by uninformative vowel duration information. Thus, native-like use of cues for non-native but familiar contrasts in unfamiliar positions may hardly ever be attained. [DOI: 10.1121/1.1906060]

PACS numbers: 43.71.Hw, 43.71.Es [ARB] Pages: 3890–3901

I. INTRODUCTION

In 1939, Trubetzkoy (reprinted as Trubetzkoy, 1977) observed that the sounds of a foreign language often get misinterpreted, because they go through the ''phonological sieve'' of the native language. Later research has proven Trubetzkoy right. The Perceptual Assimilation Model (PAM) (Best, 1994; Best, McRoberts, and Sithole, 1988) describes how listeners assimilate non-native speech sounds to the native category that is perceptually most similar. The PAM predicts which non-native speech sounds will be difficult to distinguish, based on the similarities and dissimilarities of the phonological structures of the native and the non-native language. The most difficult distinction is that between non-native speech sounds which match a single native category equally well. If the non-native language has two categories where the native language has only one in the same phonetic space, both non-native speech sounds will be assimilated to a single category. This is the case, for example, with Japanese listeners' perception of English /r/ and /l/ (Best and Strange, 1992). According to the PAM the easiest distinction is that between non-native speech sounds which are assimilated to two separate native categories. As the non-native contrast corresponds to a native contrast, it is easy to perceive.

However, languages not only have a phoneme inventory, they also have their own language-specific phonotactic constraints. The perception of non-native contrasts not only depends on the presence or absence of similar speech sounds in the native language, but also on native-language phonotactic constraints. This was demonstrated in a study of Chinese listeners' perception of the English /d/–/t/ contrast in word-final position (Flege, 1989). Chinese has a /d/–/t/ contrast, but not in word-final position. Word-initial /d/ and /t/ are not distinguished by closure voicing in Chinese, but on the basis of information in the release burst (Flege, 1989). Flege (1989) found that Chinese learners of English categorized unedited tokens of English word-final /d/ and /t/ almost as accurately as the native English listeners did. Their performance hardly decreased when closure voicing was removed, but was strongly affected by removal of the release burst. Flege concluded that the Chinese listeners used Chinese word-initial cues to distinguish between English /d/ and /t/ in word-final position. Flege and Wang (1989) showed that not only experience with the contrast itself, but native-language experience with any word-final stops influenced the perception of the word-final stop voicing contrast. Neither Cantonese Chinese nor Mandarin Chinese has a word-final stop voicing contrast, but /p,t,k/ can occur word finally in Cantonese, whereas Mandarin does not permit any word-final obstruents. Flege and Wang (1989) found that native listeners of Cantonese distinguished the English final /d/–/t/ contrast more accurately than native listeners of Mandarin did, which they attributed to the Mandarin listeners' lack of native language experience with word-final obstruents.

Further, the perception of non-native contrasts may depend not only on the presence or absence of similar phonemes in the native language, but also on the presence of utterly different contrasts. Crowther and Mann (1992) showed that the use of perceptual cues for a particular non-native contrast may depend on the use of the same cues for other phoneme distinctions in the native language. Like Mandarin Chinese, Japanese has a /d/–/t/ contrast, and does not permit word-final stops. Whereas Japanese has long and short vowels, Mandarin Chinese does not have this distinction. Crowther and Mann tested the perception of the English

[a)]A partial report of this work was presented at the 147th Meeting of the Acoustical Society of America, New York, May 2004.
[b)]Electronic mail: mirjam.broersma@mpi.nl

word-final /d/–/t/ contrast by Japanese and Mandarin learners of English. The Japanese listeners showed a greater sensitivity to the duration of the vowel preceding the final consonant and categorized the English final /d/–/t/ contrast more accurately than the Mandarin listeners did.

Thus, the categorization of a non-native contrast which exists in the native language, but in a position where it does not occur in the native language, seems to benefit from native-language experience with one of the phonemes of the contrast in the relevant position and from experience with relevant perceptual cues. These findings suggest that the potential for accurate and native-like categorization of a non-native but familiar contrast in an unfamiliar position is highest for native listeners of a language which provides such experience. Of all languages that contrast voiced and voiceless obstruents but not in word-final position, those languages which allow for either voiced or voiceless obstruents in word-final position, and in which vowel duration is used as a cue (for any phoneme contrast), offer the best preparation for accurate categorization of the word-final obstruent voicing contrast and for the use of vowel duration as a cue. As Dutch has a distinction between voiced and voiceless obstruents in word-initial and -medial position, allows for voiceless obstruents in word-final position, and also provides experience with the use of vowel duration as a cue for several phoneme distinctions, native listeners of Dutch should be well prepared to learn to distinguish English voiced and voiceless word-final obstruents as a familiar contrast in an unfamiliar position, and to use vowel duration as a cue. Especially advanced learners of English can be expected to have learned to do this, through combining their native and non-native language experience. Therefore, this paper investigates whether Dutch listeners with a high level of proficiency in English categorize English final obstruent voicing contrasts with a native-like level of accuracy and with a native-like use of the vowel duration cue. It provides a test of the perception of a non-native but familiar contrast in an unfamiliar position by listeners with a language background that is most suitable for the task.

Dutch and English share four pairs of voiced and voiceless obstruents: the alveolar and labiodental fricatives /z/, /s/, /v/, and /f/, and the bilabial and alveolar stops /b/, /p/, /d/, and /t/. Unlike English, Dutch neutralizes voicing distinctions in syllable-final, prepausal position (Booij, 1995). Thus, although in Dutch obstruent voicing is a relevant contrast in word-initial and-medial position, Dutch has no word-final voicing contrasts. Dutch does allow for /s,f,p,t/ in word-final position. Further, Dutch distinguishes between long and short vowels (Booij, 1995). As part of the difference between long and short vowels is phonetic vowel duration, Dutch listeners are familiar with the assessment of this cue. Dutch listeners even have native-language experience with the use of vowel duration as a cue to word-medial obstruent voicing. In Dutch, vowels preceding a medial voiced consonant are slightly longer than vowels preceding a medial voiceless consonant. According to Slis and Cohen (1969a), the average difference is 30 ms before stops and 40 ms before fricatives. Van den Berg (1989) found that Dutch listeners used vowel duration to decide on the voicing of intervocalic two-obstruent sequences, although it was not among the most important perceptual cues. A study by Jongman *et al.* (1992) suggests that Dutch listeners may be able to generalize their knowledge about the relationship between vowel duration and word-medial obstruent voicing to the case of word-final obstruents. In this study, Dutch listeners categorized vowels from a vowel length continuum as long or short. Stimuli corresponded to the Dutch words /stad/ and /sta:t/, and /zat/ and /za:d/, in which vowel length and underlying voicing of the final consonant are crossed. The surface word-final consonant was always voiceless. The location of the phoneme boundary differed between the two continua, suggesting that the perception of ambiguous vowel duration depended on the underlying voicing of the word-final stop.

In English, the difference in vowel duration before voiced and voiceless obstruents is larger than in Dutch. Peterson and Lehiste (1960) found a difference of 96 ms before word-final stops and 148 ms before word-final fricatives. There is extensive evidence for the great importance of preceding vowel duration for the perception of voicing of word-final obstruents in English (e.g., Raphael, 1972). Although the role of vowel duration as a cue to voicing seems to be smaller in Dutch than in English, Dutch listeners' familiarity with the cue in word-medial position may facilitate its use in word-final position in English.

As their native language has not provided them with any knowledge about the relevant acoustic cues for voicing in final position, Dutch listeners may try to identify the voicing of English final obstruents with the aid of the perceptual cues they rely on for Dutch initial and intervocalic voicing contrasts. This may be quite successful, as Dutch and English obstruents have a high degree of articulatory similarity, and the perceptual cues that signal the voicing distinctions overlap to some extent. Van Alphen and Smits (2004) showed that Voice Onset Time (VOT), specifically the presence or absence of prevoicing, is the strongest cue to initial stop voicing in Dutch. In the absence of prevoicing, voicing judgments for labials relied most strongly on the extent of $F0$ change into the following vowel, and for alveolars on the spectral center of gravity of the burst. Other significant cues were the duration and power of the burst. For intervocalic obstruents, presence or absence of vocal-fold vibration (Slis and Cohen, 1969b; Slis and Van Heugten, 1989), closure duration for stops (Kuijpers, 1996; Slis and Cohen, 1969a), and frication duration for fricatives (Slis and Van Heugten, 1989) have been shown to influence the perception of voicing. For intervocalic two-obstruent sequences, presence or absence of vocal-fold vibration during the closure of the two obstruents is the most important cue (Van den Berg, 1989). Closure duration of the second consonant, duration of the preceding vowel, and for fricatives the intensity of frication noise play a smaller but significant role (Van den Berg, 1989).

All of the above-mentioned cues have been found to be used by English listeners to distinguish voiced and voiceless obstruents in word-final position (see, e.g., Watson, 1983 for a review), and Dutch listeners may use their knowledge about Dutch voiced and voiceless obstruents to make the same distinction for final English obstruents. However, there

are differences between Dutch and English obstruents, and thus between the critical values of the cues, and the weight attributed to each cue for optimal identification.

For example, the two languages differ in the critical value of VOT for voicing of initial stops. English contrasts voiceless unaspirated and voiceless aspirated stops (Lisker and Abramson, 1964), and the duration of the voicing lag is a cue to voicing in English (e.g., Watson, 1983). Van Alphen and Smits (2004) found that 75% of Dutch voiced initial stops were produced with a voicing lead, and that the presence or absence of prevoicing was the strongest perceptual cue for initial stop consonant voicing in Dutch. Although initial stops without prevoicing were not automatically categorized as voiceless, but were assessed on the basis of other cues (as described above), initial stops without prevoicing were misperceived more often than prevoiced stops (37% vs 1%). In English, initial voiced stops are less often prevoiced than in Dutch. Smith (1978) found that bilabial voiced stops were prevoiced 56%, and alveolar stops 50% of the time in careful speech. Therefore, Dutch listeners may misperceive English initial voiced stops relatively often.

Another difference between Dutch and English is the importance of the duration of the preceding vowel as a cue to obstruent voicing. If Dutch listeners process English final obstruents in the same way they process Dutch obstruents, they may not attribute as much weight to vowel duration as English listeners do. This may not be a problem when enough other cues are available, but it may lead to less accurate categorization of unreleased stops. In English, final stops are often produced without a release burst (Byrd, 1993). English listeners have little difficulty identifying the voicing of stops without a release burst (e.g., Flege and Hillenbrand, 1987), which may be explained by the redundancy of information in the speech signal. However, if Dutch listeners use vowel duration as a voicing cue less than English listeners do, the Dutch listeners may have more difficulty identifying English final obstruents without a release burst.

In experiment 1, Dutch and English listeners' categorization of the British English obstruent voicing contrasts /z/–/s/, /v/–/f/, /b/–/p/, and /d/–/t/ was investigated in initial and final position in nonwords. It was investigated whether Dutch listeners had a preference for identifying English final obstruents as voiceless, as Dutch allows voiceless but not voiced obstruents in word-final position. The effect of removal of the release burst was investigated for the final stops. For reasons of comparison, a contrast which was expected to be difficult to distinguish for Dutch listeners was included in the experiment, namely the English /æ/–/ɛ/ contrast. The PAM predicts that this phoneme pair belongs to the set of most difficult English contrasts for Dutch listeners. Standard southern British English distinguishes two open midfront unrounded vowels, whereas Dutch has only one vowel in this part of the vowel space. Although the Dutch vowel is denoted as /ɛ/, it is lower than the English /ɛ/, so that it is located between English /ɛ/ and /æ/. As Dutch listeners will assimilate both English vowels to the single Dutch category, the distinction between the phonemes is expected to be difficult. Indeed, British English /æ/ and /ɛ/ have been found to be difficult to distinguish for Dutch listeners (Schouten, 1975).

TABLE I. Experiment 1 items.

/æ/–/ɛ/	/fæf/–/fɛf/
Initial /z/–/s/	/**z**i:f/–/**s**i:f/
Final /z/–/s/	/fu:**z**/–/fu:**s**/
Initial /v/–/f/	/**v**u:k/–/**f**u:k/
Final /v/–/f/	/ku:**v**/–/ku:**f**/
Initial /b/–/p/	/**b**o:f/–/**p**o:f/
Final /b/–/p/	/fi:**b**/–/fi:**p**/
Initial /d/–/t/	/**d**i:s/–/**t**i:s/
Final /d/–/t/	/fo:**d**/–/fo:**t**/

Dutch and English listeners' use of vowel duration as a cue to final obstruent voicing was further investigated in experiment 2. If Dutch listeners use vowel duration as a cue less than English listeners do, Dutch listeners may find it easier to ignore vowel duration when this cue is made unreliable than English listeners do. In experiment 2 it was investigated whether Dutch and English listeners relied on vowel duration as a cue to final obstruent voicing when this cue was uninformative and when it mismatched with other information in the signal.

II. EXPERIMENT 1

A. Method

1. Participants

Twenty native speakers of Dutch and 20 native speakers of British English took part in the experiment. The Dutch participants had a high level of proficiency in English as a second language. They had received on average 7 years of English instruction in primary and secondary education. The English participants did not know any Dutch. The Dutch participants were recruited from the Max Planck Institute participant pool, and the English participants from the participant pool of the Laboratory of Experimental Psychology of the University of Sussex. None reported any hearing loss. All were volunteers and received a small fee for participation.

2. Materials

The vowel contrast /æ/–/ɛ/ was to be tested in one position and the four consonant contrasts /z/–/s/, /v/–/f/, /b/–/p/, and /d/–/t/ in two positions. Therefore, nine pairs of monosyllabic CVC items were selected. Each pair differed in one phoneme pair, corresponding to the contrast to be tested. The nontarget consonants in the CVC items were obstruents, in order to minimize their influence on the target sounds. All

TABLE II. Experiment 1, acoustic measures of stimuli with /æ/ and /ɛ/: Mean $F1$ steady-state frequency (Hz), mean $F2$ steady-state frequency (Hz), and mean vowel duration (ms).

	/æ/	/ɛ/
$F1$	824	744
$F2$	1602	1823
Vowel duration	167.4	131.6

TABLE III. Experiment 1, acoustic measures of stimuli with initial and final fricatives: Mean vowel duration (ms), mean fricative duration (ms), and mean fricative power above 500 Hz (logarithm of the spectral power of the frication noise above 500 Hz in Pa^2).

	Initial				Final			
	/z/	/s/	/v/	/f/	/z/	/s/	/v/	/f/
Vowel duration	⋯	⋯	⋯	⋯	258.7	130.5	264.9	118.5
Fricative duration	112.8	178.3	111.2	160.9	160.5	257.8	144.8	232.7
Fricative power	−2.2	−2.2	−3.1	−3.1	−2.6	−2.2	−3.6	−3.3

items were nonwords in Dutch and English, according to the CELEX database (Baayen, Piepenbrock, and Gulikers, 1995). The items are presented in Table I.

The materials were recorded by a male native speaker of British English. The speaker read the items one by one, separated by a pause, in a clear citation style. The recording was made in a soundproof booth with a Sennheiser microphone and stored directly onto a computer at a sample rate of 16 kHz. For each target phoneme in each position, three tokens were extracted from the file with the speech editor PRAAT. For the items with a final stop, only tokens with a clearly audible release burst were selected. These tokens were kept unedited for the condition with release burst, and for the condition without release burst the signal was truncated at the last positive zero crossing before the release burst.

Acoustic measurements were made of several characteristics which may be relevant for the distinction of the contrasts. The results are presented in Table II for the target vowels, in Table III for the fricatives, and in Table IV for the stops.

3. Design

Each fricative contrast occurred in initial and final position. The stops occurred in three conditions: initial position, final position with release burst, and final position without release burst. The order of presentation of the initial and final positions was counterbalanced. As the items in the two final conditions were based on the same tokens, the final with release burst condition always occurred after the final without release burst condition. The target phonemes /s/ and /f/ also occurred as nontargets in stimuli for other contrasts. The blocks were ordered such that the subjects were not exposed to a phoneme before the contrast it was part of was being tested.

The items were presented in 11 blocks, each block representing one phoneme contrast in one condition. Each block consisted of four repetitions of six tokens, semirandomized such that the same phoneme occurred maximally five times in succession and the same token maximally once.

4. Procedure

Participants were tested one at a time in a quiet room. They were informed in their native language that they would hear a series of nonwords, which would be similar except for one sound. They were instructed to decide which one of two alternatives this sound was, and to indicate their response with a button press. Before each block, they received further information about the two response alternatives in that block, and about the position of the target phoneme. They were not instructed about the truncation in the condition without release burst. Before the /æ/–/ɛ/ block, participants heard some examples of nonwords containing these phonemes to make it clear, particularly to the Dutch participants, which sounds were intended. The other phonemes were not expected to cause uncertainty, and were not illustrated with examples. Each block started with six practice trials. The response buttons were labeled "A" and "E," "Z" and "S," "V" and "F," "B" and "P," or "D" and "T," respectively. The experiment was controlled with NESU (Nijmegen Experiment Set-Up) experimental software. Stimuli were presented binaurally over Sennheiser closed headphones at a comfortable listening level, one at a time. Participants responded by pressing one of two response buttons. No time limit was imposed for the responses. After each button press, presentation of the next item started.

TABLE IV. Experiment 1, acoustic measures of stimuli with initial and final stops: Proportion of initial stops with prevoicing, mean vowel duration (ms), mean *F*1 offset frequency (Hz), mean closure duration (ms), proportion of final stops with voicing during closure, mean closure voicing duration (as a percentage of total closure duration), and mean burst duration (ms).

	Initial				Final			
	/b/	/p/	/d/	/t/	/b/	/p/	/d/	/t/
Prevoicing	2/3	0	3/3	0	⋯	⋯	⋯	⋯
Vowel duration	⋯	⋯	⋯	⋯	185.7	112.7	252.8	133.3
*F*1 offset frequency	⋯	⋯	⋯	⋯	246	261	272	310
Closure duration	⋯	⋯	⋯	⋯	82.1	117.8	59.1	113.6
Closure voicing	⋯	⋯	⋯	⋯	3/3	0	3/3	3/3
Closure voicing duration (%)	⋯	⋯	⋯	⋯	94.2	0	100	27.5
Burst duration	10.8	22.6	16.3	18.7	67.5	72.0	72.0	122.9

TABLE V. Experiment 1 results: Mean percentage of correct responses as a function of participants' native language and condition.

	Dutch				English			
	Medial	Initial	Final released	Final dereleased	Medial	Initial	Final released	Final dereleased
/æ/	96	⋯	⋯	⋯	97	⋯	⋯	⋯
/ɛ/	94	⋯	⋯	⋯	100	⋯	⋯	⋯
/z/	⋯	96	96	⋯	⋯	97	93	⋯
/s/	⋯	96	98	⋯	⋯	98	95	⋯
/v/	⋯	95	94	⋯	⋯	98	97	⋯
/f/	⋯	92	100	⋯	⋯	98	99	⋯
/b/	⋯	86	96	92	⋯	99	98	96
/p/	⋯	98	98	93	⋯	99	98	96
/d/	⋯	99	94	95	⋯	97	96	97
/t/	⋯	99	98	94	⋯	98	95	93

B. Results and discussion

One response with a reaction time (RT) longer than 10 000 ms due to a technical error was removed. One Dutch subject gave only "v" responses for the /v/–/f/ contrast in final position. All responses of this subject on both /v/–/f/ contrasts were removed from the analysis. Mean percentages of correct responses are presented in Table V. The sensitivity measure d' was calculated for each subject, for each contrast, and each condition separately, with a correction for near-perfect sensitivity (MacMillan and Creelman, 1991). Next, $\log \beta$ was calculated to investigate possible biases (McNicol, 1972). Mean values of d' and $\log \beta$ are presented in Table VI.

For the /æ/–/ɛ/ contrast, an analysis of variance (ANOVA) showed that the d's of the English listeners were significantly larger than those of the Dutch listeners [$F(1,39)=7.59, p<0.01$], indicating a higher sensitivity of the English listeners. However, a t-test showed that the Dutch listeners performed amply above chance ($d'=0$), with d's significantly larger than 3 [$t(19)=5.38, p<0.001$]. There was no effect of native language on bias [$F(1,39)=2.21, p>0.1$].

For the /z/–/s/ contrast, no interaction between condition and native language [$F(1,38)=1.93, p>0.1$], and no main effects of condition [$F(1,38)<1$] and native language [$F(1,38)<1$] were found. No difference in bias between the language groups was found for initial position [$F(1,39)<1$] or for final position [$F(1,39)<1$].

TABLE VI. Experiment 1 results: Mean d' and $\log \beta$ as a function of participants' native language and condition. (Higher values of d' indicate higher sensitivity. Negative values of $\log \beta$ indicate a bias towards the first, and positive values towards the second phoneme of a contrast.)

	Dutch		English	
	d'	$\log \beta$	d'	$\log \beta$
/æ/–/ɛ/	4.21	−0.28	4.96	0.59
Initial /z/–/s/	4.48	−0.25	4.57	0.00
Final /z/–/s/	4.55	0.33	4.17	−0.07
Initial /v/–/f/	4.20	−0.64	4.83	0.00
Final /v/–/f/	4.61	1.23	4.79	0.43
Initial /b/–/p/	3.93	1.76	5.14	0.00
Final /b/–/p/ released	4.68	0.34	4.72	0.02
Final /b/–/p/ dereleased	3.85	0.42	4.31	0.11
Initial /d/–/t/	5.11	−0.02	4.73	0.42
Final /d/–/t/ released	4.45	0.39	4.21	−0.31
Final /d/–/t/ dereleased	3.90	0.09	4.11	−1.00

For the /v/–/f/ contrast, no interaction between condition and native language [$F(1,37)=1.22, p>0.1$], and no main effects of condition [$F(1,37)<1$] and native language [$F(1,37)=2.40, p>0.1$] were found. No difference in bias between the language groups was found for initial position [$F(1,39)=1.07, p>0.1$] or for final position [$F(1,38)=2.27, p>0.1$].

For the /b/–/p/ contrast, a significant interaction between condition and native language was found [$F(2,76)=4.45, p<0.05$]. A planned comparison of initial position and final position with release burst yielded a significant interaction between condition and native language [$F(1,38)=8.65, p<0.01$]. Therefore, separate analyses were performed for both conditions and both language groups. In initial position, the English listeners' sensitivity was significantly higher than the Dutch listeners' sensitivity [$F(1,39)=19.75, p<0.001$]. In final position with release burst, there was no effect of native language [$F(1,39)<1$]. Comparing initial position and final position with release burst for the Dutch listeners only, a significantly lower sensitivity was found for initial position [$F(1,19)=5.68, p<0.05$]. For the English listeners, there was no difference between initial position and final position with release burst [$F(1,19)=2.98, p>0.1$].

In a planned comparison of final position with release burst and final position without release burst, no interaction was found between condition and native language [$F(1,38)=1.21, p>0.1$]. The effect of condition was significant [$F(1,38)=10.69, p<0.01$], with d' being larger for final position with release burst than for final position without release burst. There was no significant effect of native language [$F(1,38)<1$].

For initial position, the effect of native language on bias was significant [$F(1,39)=13.94, p<0.001$], with a bias towards "p" responses for the Dutch listeners, and no bias ($\log \beta=0$) for the English listeners. Neither the analysis of $\log \beta$ for final position with release burst [$F(1,39)<1$] nor that without release burst [$F(1,39)<1$] yielded a significant effect.

The Dutch listeners' bias towards "p" responses in ini-

tial position suggested that their low sensitivity for the initial contrast resulted from a high number of errors on /b/ items rather than on /p/ items. As Table V shows, the Dutch listeners' percentage of correct responses for /p/ items was similar to that of the English listeners (98% vs 99%), whereas the Dutch listeners' percentage of correct responses for /b/ items was only 86%, compared to 99% for the English listeners. Acoustical examination of the /b/ items showed that two of the tokens were produced with prevoicing, and the third without prevoicing (Table IV). The tokens with prevoicing received 99% and 96% correct responses from the Dutch listeners, whereas the token without prevoicing received only 63% correct responses from the Dutch listeners. This score is identical to the percentage of correct responses for Dutch initial voiced stops produced without prevoicing found by Van Alphen and Smits (2004). As expected, the English listeners categorized all tokens of initial voiced stops accurately, regardless of the presence or absence of prevoicing.

Finally, for the /d/–/t/ contrast, no interaction was found between condition and native language [$F(2,76)=1.18, p>0.1$]. There was a significant effect of condition [$F(2,76)=10.54, p<0.001$]. A *posthoc* Bonferroni test showed that the sensitivity scores were higher in initial position (note that all initial /d/'s were prevoiced) than in final position with release burst ($p<0.05$) and in final position without release burst ($p<0.001$), and that the two conditions in final position did not differ significantly from one another ($p>0.1$). No effect of native language was found [$F(1,38)<1$]. There was no effect of native language on bias for initial position [$F(1,39)=1.08, p>0.1$], for final position with release burst [$F(1,39)=1.63, p>0.1$], or for final position without release burst [$F(1,39)=3.75, p=0.060$].

For both Dutch and English listeners, the removal of the release burst affected the categorization of the /b/–/p/ contrast but not of the /d/–/t/ contrast. The signal remaining after removal of the release burst may have contained clearer cues for the alveolar stops than for the bilabial stops. For example, the difference in vowel duration and in $F1$ offset frequency of /d/ and /t/ was larger than that of /b/ and /p/ (Table IV).

In general, the results for the four consonant contrasts show a consistent pattern. For the sensitivity measure, no interactions between position and native language were found for the contrasts /v/–/f/, /z/–/s/, and /d/–/t/. The Dutch listeners' categorization of /v/–/f/ and /z/–/s/ was as accurate in final position as in initial position, and as accurate as that of the English listeners. For the /d/–/t/ contrast, both groups performed better on initial position than on final position with release burst. For the /b/–/p/ contrast, there was an interaction between condition and native language. The English listeners outperformed the Dutch listeners on the initial position. The absence of prevoicing of initial stops hindered the Dutch listeners' categorization in English as much as it does in Dutch. As voiced stops are prevoiced less often in English than in Dutch, this may cause Dutch listeners to misperceive the voicing of initial stops in English more frequently than in Dutch. Dutch and English listeners performed equally well on the /b/–/p/ contrast in final position with release burst. Although Dutch does not allow for voiced obstruents in final, prepausal position, the Dutch listeners did not have a bias towards voiceless responses in final position.

The results are in line with the predictions of the PAM. Whereas the Dutch listeners categorized the /æ/–/ɛ/ contrast less accurately than the English listeners did, they categorized the English final voicing contrasts as accurately as the initial contrasts, and as accurately as the English listeners did.

Experiment 1 also tested categorization accuracy for final stops without a release burst. A difference between the Dutch and English listeners' categorization accuracy could have indicated a differential use of the duration of the preceding vowel as a voicing cue. However, the removal of the release bursts of final stops did not influence the Dutch and the English listeners differentially. For the /b/–/p/ contrast, Dutch and English listeners performed better on items with release burst than on the same tokens without release burst. The removal of the release burst affected the performance of the two language groups to the same extent. For the /d/–/t/ contrast there was no difference in sensitivity to items with or without release burst, neither for the Dutch nor for the English listeners. Thus, experiment 1 did not provide any evidence that the Dutch listeners used vowel duration less than the English listeners did.

However, experiment 1 was not a direct test of the use of vowel duration as a cue. Apart from vowel duration, several other cues remained available after removal of the release burst (e.g., $F1$ offset frequency, closure voicing; see Table IV). Thus, the Dutch listeners may have achieved a native-like level of accuracy without using vowel duration as a cue. On the other hand, the absence of a release burst may have stimulated the Dutch listeners to use vowel duration, while they may not do so when more perceptual cues are available. The results from this experiment are not decisive about these possibilities. In fact, any evidence of Dutch listeners using vowel duration as a cue for final voicing would leave open the possibility that their use of the cue was a reaction to the task at hand.

Therefore, the use of vowel duration as a cue to final obstruent voicing was investigated from a different angle in experiment 2. In this experiment, stimulus materials were constructed such that they did not stimulate but rather discouraged the use of vowel duration as a voicing cue. The question was addressed whether Dutch listeners use vowel duration as a voicing cue as persistently as English listeners do. If the Dutch listeners did not use vowel duration in experiment 2, this would not imply that they never do so. However, it could show that Dutch listeners do not use this cue as persistently as English listeners.

It was argued that among the languages that do not have voiced and voiceless obstruents in final position, Dutch prepares its listeners well for the use of vowel duration as a cue to English word-final obstruent voicing. As Dutch has long and short vowels, Dutch listeners are familiar with the use of phonetic vowel duration. They also have native-language experience with the use of vowel duration as a cue to word-medial obstruent voicing (Van den Berg, 1989). Although the role of vowel duration as a cue to voicing seems to be smaller in Dutch than in English, Dutch listeners' familiarity

with the cue in word-medial position may facilitate its use in word-final position in English. Especially advanced learners of English may have learned to use this word-final voicing cue, combining their native and non-native language experience. Therefore, Dutch listeners with a high level of proficiency in English might be expected to process the English obstruent voicing contrast in a native-like manner, with a native-like use of the vowel duration cue. If Dutch listeners do not use vowel duration in a native-like manner, however, this would raise the question whether non-native listeners can ever be expected to process non-native but familiar phoneme contrasts in unfamiliar positions in a native-like manner.

In experiment 2, Dutch and English listeners' categorization of English final voiced and voiceless obstruents was investigated again. For reasons of comparison, categorization of initial voicing contrasts was tested as well. For practical reasons regarding the construction of phoneme continua, only fricatives were tested.

III. EXPERIMENT 2

A. Method

1. Participants

Twenty-eight native speakers of Dutch and 28 native speakers of British English, none of whom had participated in experiment 1, took part in the experiment. The Dutch participants had a high level of proficiency in English as a second language (as in experiment 1), whereas the English participants did not know any Dutch. The Dutch participants were recruited from the Max Planck Institute participant pool, and the English participants were recruited from the participant pool of the Laboratory of Experimental Psychology of the University of Sussex or at the University of Birmingham. None reported any hearing loss. All were volunteers and received a small fee for participation.

2. Materials

The same nonwords for initial and final fricative contrasts were used as in experiment 1. The materials were recorded by the same native speaker of British English who recorded the materials for experiment 1. The speaker read the items one by one, separated by a pause, in a clear citation style. The materials were recorded with a Sennheiser microphone in a soundproof booth onto digital audiotape and downsampled to 16 kHz during transfer to a computer. For each contrast, two target sounds and one or two carriers were extracted from the sound file, using the speech editor XWAVES. The target sounds were used to create voicing continua which were spliced onto the appropriate carriers, as described below.

From the nonword /zi:f/, /i:f/ was extracted, removing the initial /z/, with the cut being made at the first positive zero crossing after the offset of frication noise. From /fu:s/ and /fu:z/, /fu:/ was extracted, truncating the signal at the last positive zero crossing before the onset of frication noise. In a similar way, /u:k/ was extracted from /vu:k/, removing the initial /v/, and /ku:/ was extracted from /ku:f/ and /ku:v/, removing the final /f/ and /v/. These elements served as carriers.

An initial /z/ was extracted from another token of /zi:f/, truncating the signal at the first positive zero crossing after the offset of the frication noise. A final /z/ was extracted from another token of /fu:z/, with the cut being made at the last positive zero crossing before the start of the frication noise. Similarly, an initial and a final /s/, /v/, and /f/ were extracted from other tokens of /si:f/, /fu:s/, /vu:k/, /ku:v/, /fu:k/, and /ku:f/, respectively. For the initial and final /s/ and /f/, a portion in the center of the fricative was removed, such that the duration of each voiceless fricative matched the duration of its voiced counterpart. The initial /s/ and /f/ were shortened by 2 and 8 ms, respectively, and the resulting durations were 115 ms for the /s/ and 148 ms for the /f/. The final /s/ was shortened by 80 to 187 ms, and the final /f/ was shortened by 56 to 127 ms. The final /s/ and /f/ were shortened by 30% and 31%, respectively. The four pairs of phonemes that were thus obtained served as the endpoints of the four continua. For each continuum, nine intermediate steps were generated, following the procedure of Stevenson (1979) and Repp (1981). In this procedure, the amplitudes of two waveforms are added in varying proportions. The proportions have a ratio of 0–1 and 1–0 in the two endpoints, and are equally spaced in the intermediate steps, always adding up to 1.

For each continuum, the two endpoints and the nine intermediate steps were spliced onto the appropriate carriers. Thus, the resulting stimuli ranged from /zi:f/ to /si:f/ and from /fu:z/ to /fu:s/ for the alveolar fricatives, and from /vu:k/ to /fu:k/ and from /ku:v/ to /ku:f/ for the labiodental fricatives. For the two initial continua, there was one carrier each. For the two final continua, there were two carriers each. One was originally pronounced with a voiceless final fricative and contained a phonetically short vowel (of 118 ms for the /z/–/s/ contrast and 98 ms for the /v/–/f/ contrast); the other was originally pronounced with a voiced final fricative and contained a phonetically long vowel (of 233 ms for the /z/–/s/ contrast and 257 ms for the /v/–/f/ contrast).

3. Design

Stimuli were blocked by contrast, position, and carrier. Each block was presented to half of the participants. As there were two carriers for final contrasts and one for initial contrasts, half of the participants only heard the two final contrasts, and the other half heard the two final contrasts and the two initial contrasts. The order of the blocks with initial and final contrasts was counterbalanced where applicable. As explained above, the nontarget consonants in the CVC items were obstruents. As this restriction yields a limited number of items that are nonwords in both languages, /f/ was part of the carrier in the items where /z/ and /s/ were the target sounds. Therefore, the /v/–/f/ contrast was always tested before the /z/–/s/ contrast. Each block consisted of 20 repetitions of the 11 steps of the continuum, semirandomized such that the same step could not occur twice in succession.

Crucially, each participant was presented with only one carrier for each final contrast. For each participant, the dura-

TABLE VII. Experiment 2 results: Mean percentage of "z" or "v" responses to initial fricatives as a function of the place on an 11-step stimulus continuum ranging from /z/ to /s/ or from /v/ to /f/ and participants' native language.

	Dutch		English	
	/z/–/s/	/v/–/f/	/z/–/s/	/v/–/f/
1 (Voiced)	85	93	95	98
2	83	94	91	96
3	83	94	94	97
4	83	93	92	93
5	83	87	91	88
6	79	72	83	76
7	63	53	78	55
8	56	37	65	27
9	41	23	50	12
10	28	15	26	6
11 (Voiceless)	10	9	8	3

tion of the vowel for each final contrast was unvarying. Thus, vowel duration was not informative for the voicing contrast. For all participants, there was a mismatch between vowel duration and other information in the signal for a subset of the stimuli. When voiced fricatives were preceded by a short vowel, or voiceless fricatives by a long vowel, vowel duration and information in the frication noise pointed in opposite directions.

4. Procedure

The procedure was as described for experiment 1. Each block was preceded by a practice part containing two presentations of each of the 11 steps of the continuum in semirandomized order. The response buttons were labeled "Z" and "S," or "V" and "F," respectively.

B. Results and discussion

Eight responses with RTs longer than 10 000 ms due to a technical error were removed from the analysis. The categorization curves of each contrast in each position and for each subject separately were fitted with logistic regression. From the regression models, 50-percent crossover points were retrieved, reflecting the location of the category boundary. From the models, a measure of the steepness of the categorization curve at the 50-percent crossover point was computed, indicating how categorical perception was. In five response sets, the percentage of correct responses at step 1 or 11 did not exceed 50% (one response set representing one contrast in one position for one subject). No logistic regressions were performed on those response sets.

The categorization results for the initial contrasts are presented in Table VII. There were no differences between the Dutch and English listeners in the steepness of the slopes, either for the /z/–/s/ contrast $[F(1,27)=1.55, p>0.1]$ or for the /v/–/f/ contrast $[F(1,26)<1]$.

The categorization results for the final /z/–/s/ contrast are presented in Fig. 1. If vowel duration was used for the categorization of ambiguous fricatives from the middle region of the voicing continuum, this should have resulted in a shift between the curves corresponding to the long and short vowel conditions. However, an ANOVA on the 50-percent

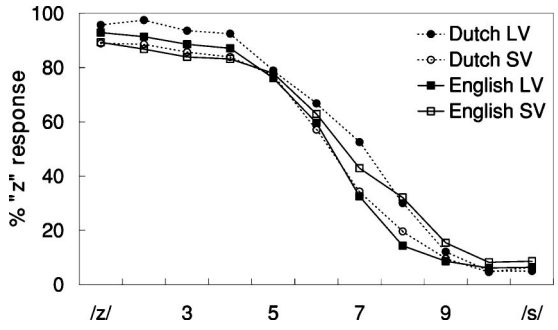

FIG. 1. Experiment 2: Mean percentage of "z" responses to final fricatives as a function of the place on an 11-step stimulus continuum ranging from /z/ to /s/, preceding vowel duration (*LV*: long vowel; *SV*: short vowel), and participants' native language.

crossover points showed no interaction between vowel duration and native language $[F(1,54)=2.53, p>0.1]$, and no main effects of vowel duration $[F(1,54)=1.17, p>0.1]$ or native language $[F(1,54)<1]$.

The categorization results for the final /v/–/f/ contrast are presented in Fig. 2. The graphs show that vowel duration had a differential effect on the Dutch and the English participants' categorization results. A significant interaction was found between the effects of vowel duration and native language on 50-percent crossover point $[F(1,52)=4.32, p<.05]$.

For the Dutch listeners, the categorization curves were similar in the conditions with the short and with the long preceding vowel. Although the curve for items with a long vowel was located slightly further towards the voiceless side of the continuum than the curve for the short vowel, the 50-percent crossover points were not statistically different $[F(1,27)<1]$.

For the English listeners, the categorization curve for the items with a long vowel was strongly shifted relative to the curve for items with a short vowel. The 50-percent crossover points were significantly different in the two conditions $[F(1,24)=18.63, p<0.001]$, with a larger 50-percent crossover point for the condition with longer vowel duration, showing a preference for "v" responses which persisted further towards the voiceless side of the continuum.

Moreover, the English listeners categorized even the /f/ endpoint as "v" 31% of the time when it was preceded by a long vowel. The difference between the curves for the short

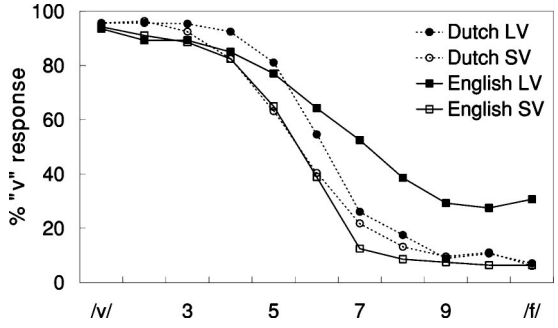

FIG. 2. Experiment 2: Mean percentage of "v" responses to final fricatives as a function of the place on an 11-step stimulus continuum ranging from /v/ to /f/, preceding vowel duration (*LV*: long vowel; *SV*: short vowel), and participants' native language.

and long vowel conditions was located on the voiceless side of the continuum. ANOVAs on arcsine-transformed proportions showed that from steps 6 to 11 the proportion of "v" responses was significantly higher for items with a long vowel than for items with a short vowel ($p<0.01$).

There was no effect of native language on steepness of the slope in the short vowel condition [$F(1,27)<1$]. In the long vowel condition, the Dutch listeners' categorization curve was steeper than the English listeners' curve [$F(1,24)=4.76, p<0.05$], indicating that the Dutch listeners' categorization was more categorical than the native English listeners' categorization.

Vowel duration thus affected the categorization of the /v/–/f/ contrast and the /z/–/s/ contrast differentially. For the final /v/–/f/ contrast, a change in vowel duration led to a shift in the categorization curve for the English but not for the Dutch listeners. For the final /z/–/s/ contrast, there was no shift for either language group. A significant three-way interaction [$F(1,48)=10.12, p<0.001$] among the effects of vowel duration, native language, and place of articulation on 50-percent crossover point confirmed that the contrasts differed in this respect.

The finding that there was no effect of vowel duration for the final /z/–/s/ contrasts is not surprising in itself, as vowel duration was not informative in this experiment. As vowel duration was kept constant for each participant throughout the whole block, it did not have any cue value for the voicing contrast. Nevertheless, the English listeners but not the Dutch listeners showed an effect of vowel duration in their categorization of the final /v/–/f/ contrast. As the English listeners based their categorization decisions on the uninformative vowel duration, their categorization for the /v/–/f/ contrast preceded by a long vowel was less categorical than the Dutch listeners' categorization.

The different results for the alveolar and labiodental fricatives may be caused by their acoustic characteristics. Word-initial alveolar fricatives have a higher amplitude than labiodental fricatives (Jongman, Wayland, and Wong, 2000). A similar difference may exist in final position. Indeed, the spectral power of the final /z/ and /s/ was higher than that of the final /v/ and /f/ in experiment 1 (Table III) and in experiment 2 (where the mean logarithms of the spectral power of the frication noise above 500 Hz in Pa^2 were -3.1 and -4.3, respectively). Further, in final position, alveolar fricatives have a longer noise duration than labiodental fricatives (Crystal and House, 1988). This was also the case in experiment 1 (Table III) and in the original final fricatives in experiment 2. Therefore, the information in the alveolar frication signal may generally be more easily perceptible for the listener than the information in a labiodental. A less informative frication signal may stimulate listeners to exploit other sources of information. This may explain why the English participants took vowel duration into consideration in their decisions for the final /v/–/f/ contrast, but not for the /z/–/s/ contrast.

English listeners categorized the same tokens significantly more often as "v" when they were preceded by a phonetically long vowel than when they were preceded by a short vowel. Even the endpoint /f/ was categorized as "v" 31% of the time in the long vowel condition. Apparently, the long duration of the vowel pointed towards a voiced fricative so strongly that it overruled the other information in the signal in many cases. Note that the reverse did not happen on the other side of the continuum: tokens at the voiced end of the continuum received a high percentage of voiced responses, even when preceded by a short vowel. Several factors may have contributed to this asymmetry. In the first place, the final /f/ was shortened to match the duration of the final /v/. As frication duration is a cue to voicing (e.g., Watson, 1983), the shortening made the final /f/ more /v/-like. Note that this shortening did not lead to a high percentage of "v" responses for the Dutch listeners, or for the English listeners in the short vowel condition. Neither was there a high percentage of "z" responses to the endpoint /s/, which was shortened to a similar extent. Thus, the shortening cannot explain the high percentage of "v" responses for the English listeners in the long vowel condition, but it may have made the /f/ endpoint more acceptable as a "v" than vice versa. Second, listeners have experience with vowel shortening in fast speech. In fast speech, vowels are reduced relatively more than consonants (Gay, 1978), so that not only the vowel duration itself, but also the ratio of vowel and fricative duration changes. Indeed, vowel lengthening before voiced obstruents decreases at faster speaking rates (Smith, 2002). As a result of this experience with absolute and relative vowel shortening, listeners may find it easier to ignore short vowel duration as a cue to voicing when it mismatches with other cues than to ignore long vowel duration. In the third place, phonologically voiced obstruents are phonetically often unvoiced (Stevens *et al.*, 1992). Therefore, listeners may show asymmetric weighting of the presence or absence of phonetic voicing. The presence of phonetic voicing may signal a voiced obstruent relatively strongly, while its absence may not point as strongly towards a voiceless interpretation. This may have contributed to the finding that tokens on the voiced side of the continuum were predominantly perceived as voiced, whereas in the long vowel condition, tokens on the voiceless side of the continuum were less often perceived as voiceless.

IV. GENERAL DISCUSSION

The results of the two experiments presented in this paper show that a native-like level of accuracy may be reached for the categorization of non-native phonemes, even though the phonemes are not necessarily processed in a native-like manner.

Experiment 1 investigated the accuracy with which Dutch listeners categorized English contrasts with different degrees of correspondence in Dutch phonology. The /æ/–/ɛ/ contrast, which the PAM predicts to be among the most difficult English contrasts for Dutch listeners, was indeed found to be the most difficult contrast in this experiment. Although the Dutch listeners performed amply above chance, the English listeners showed a significantly higher sensitivity than the Dutch listeners did. The obstruent voicing contrasts are matched by similar contrasts in Dutch, and the PAM predicts them to be easy to distinguish. Although Dutch voicing contrasts do not occur in final position, the Dutch listeners cat-

egorized the English final voicing contrasts as accurately as (or even more accurately than) the initial contrasts, and as accurately as the English listeners did. Dutch listeners were not biased towards voiceless responses in final position.

The PAM does not make any predictions about the perception of familiar contrasts in unfamiliar positions. The present study suggests that an unfamiliar position does not necessarily complicate the perception of familiar but non-native contrasts. For example, in experiment 1 the Dutch listeners categorized the /b/–/p/ contrast more accurately in the unfamiliar final position than in the familiar initial position. In order to make predictions about the perception of familiar contrasts in familiar and unfamiliar positions, it is important to take into account that speech sounds have different acoustic characteristics in different positions. The extent to which these characteristics overlap with those of the native speech sounds seems an important predictor of the ease with which non-native listeners can distinguish between the sounds.

In experiment 1, no evidence was found that the Dutch listeners used vowel duration as a cue to final voicing less than the English listeners did. The removal of the release burst from final stops did not affect the Dutch listeners' categorization more than it affected the English listeners' categorization. However, the Dutch listeners may have achieved a native-like level of accuracy without using vowel duration as a cue. Flege (1989) found a native-like level of accuracy for Chinese listeners' categorization of unedited tokens of English final /d/–/t/. Nonetheless, the Chinese listeners were found to rely on cues in the release burst more than the English listeners did. They had achieved a native-like accuracy through a non-native-like manner of processing. When the release burst was removed, they were no longer able to maintain a native-like level of accuracy. The Dutch listeners may have been better able to adapt to the removal of the release burst. Their knowledge about English voicing cues may have been sufficient to use those cues that remained available when the burst had been removed (e.g., $F1$ offset frequency, closure voicing). They may have achieved a native-like level of accuracy for the categorization of stops without release burst without using vowel duration. On the other hand, it is also possible that the absence of a release burst stimulated the Dutch listeners to use vowel duration, while they may not do so when more perceptual cues are available. Thus, the results from experiment 1 left the possibility open that non-native listeners do not need to process non-native phonemes in a native-like manner in order to achieve a native-like level of accuracy.

Therefore, experiment 2 tested the use of the duration of the preceding vowel as a cue to final fricative voicing with items which were constructed such that they did not stimulate the use of vowel duration as a voicing cue. In experiment 2, categorization of initial fricative voicing contrasts was tested as well. In line with the results from experiment 1, no differences were found between Dutch and English listeners' categorization of initial fricative voicing contrasts.

In order not to stimulate the Dutch listeners to use vowel duration more than they would normally do, but rather discourage its use, the vowel duration cue was kept uninformative. Vowel duration even mismatched with other information in the signal for some of the tokens. Nevertheless, the English listeners tried to use vowel duration for the categorization of /v/ and /f/. Especially, the categorization of tokens with long vowels was often consistent with vowel duration. Apparently, for the English listeners, vowel duration was such an important cue for final voicing that it often overruled other information in the signal. The Dutch listeners, on the other hand, did not use vowel duration at all. Even in the middle range of the continuum, for tokens with an ambiguous identity, there was no effect of vowel duration on the Dutch listeners' categorization of final /z/–/s/ or /v/–/f/. Thus, in experiment 2, Dutch listeners did not categorize final voiced and voiceless obstruents in a native-like manner. The Dutch listeners were able to ignore vowel duration when it was uninformative and misleading. They differed in this respect from the English listeners, who ignored vowel duration for the categorization of the final /z/–/s/ contrast, but relied heavily on it for the categorization of the final /v/–/f/ contrast. As a result, for the /v/–/f/ contrast preceded by a long vowel, Dutch listeners' categorization curve was steeper than that of the native English listeners. As vowel duration was not informative and sometimes mismatched with other voicing cues, English listeners' use of vowel duration for the /v/–/f/ contrast resulted in less categorical perception.

The Dutch listeners, who did not use vowel duration as a cue in experiment 2, may do so in other circumstances where the cue is informative. Indeed, they may have used vowel duration as a cue in experiment 1. However, the results from experiment 2 showed that the Dutch listeners did not use vowel duration as persistently as the English listeners did. This may be a result from their native-language experience, where vowel duration is a less important cue to (word-medial) obstruent voicing than in English. From their native-language experience, the Dutch listeners may have inferred that vowel duration is only a minor cue to English final obstruent voicing as well. Another explanation could be that Dutch listeners are regularly exposed to English spoken by native speakers of Dutch. Elsendoorn (1985) has shown that the difference in vowel duration before voiced and voiceless final obstruents in the English spoken by Dutch learners is smaller than that in the English of native speakers. From exposure to English spoken by Dutch learners, Dutch listeners may have learned to ignore vowel duration as a voicing cue when it is uninformative.

It was argued that Dutch prepares its listeners well for the distinction of English word-final obstruent voicing. Dutch has obstruent voicing contrasts which are perceptually similar to English contrasts, which makes the English contrasts easy to distinguish according to the PAM (Best *et al.*, 1988). Dutch allows for voiceless obstruents in word-final position, which has been found to facilitate perception of the distinction (Flege and Wang, 1989). Dutch provides experience with phonetic vowel duration for the distinction of phonemically long and short vowels, which has been found to facilitate the use of this cue for the final consonant voicing contrast (Crowther and Mann, 1992), and Dutch provides experience with the use of vowel duration as a cue to word-medial obstruent voicing (Van den Berg, 1989). Indeed,

Dutch listeners were found to categorize English final obstruent voicing with a native-like level of accuracy. Nevertheless, they were found to use vowel duration as a cue less persistently than the English listeners did.

As even native listeners of Dutch, who had the necessary experience for a native-like use of vowel duration, and who had a high level of proficiency in English, did not use perceptual cues in a native-like manner, this raises the question whether non-native listeners can ever be expected to process non-native but familiar phoneme contrasts in unfamiliar positions in a native-like manner.

Although the Dutch listeners in this study had a high level of proficiency in English, they had not reached a level of ultimate attainment. Their English perception skills were still open to improvement. However, such improvement may not involve the use of perceptual cues for contrasts which they could already accurately distinguish. As the results from the experiments in this paper show, a native-like level of accuracy can be achieved, even when the perceptual cues are not processed in a native-like manner. Possibly, Dutch listeners can learn to use vowel duration as a cue to English word-final obstruent voicing in a fully native-like manner with a native-like persistence, for example through laboratory-based training, but there may be no need to learn this for normal language use. Presumably, the second language learner's goal is not to process language in a native-like manner, but rather to be able to understand (and produce) the language well enough to meet the learner's communicative needs. If a native-like use of perceptual cues is not necessary for accurate perception, it is possible that listeners may never learn to perceive non-native phonemes in a native-like manner. The benefits of native-like processing may be too small, or even nonexistent.

It should be relatively easy for Dutch listeners to learn to use vowel duration in an English native-like manner, and nonetheless, in experiment 2 the Dutch listeners did not use vowel duration as a cue when the English listeners did. As even these Dutch listeners did not use vowel duration in a native-like way, it seems likely that native listeners of languages for whom the cue is harder to learn would in many cases not use the cue in a native-like manner either. The reasoning can be extended to other perceptual cues as well. The vowel duration cue provides a great amount of information about the English voicing contrast, and it is a very important cue for English listeners. It therefore seems a good candidate for native-like use by non-native listeners. As even this cue was not used in a native-like way, it seems likely that other, less informative cues would in many cases not be used in a native-like manner either. Thus, for the distinction of non-native but familiar contrasts in unfamiliar positions, a native-like manner of phonetic processing may hardly ever be attained.

ACKNOWLEDGMENTS

This work was supported by a grant from the Max Planck Society for the Advancement of Science and by a Marie Curie fellowship. I thank Antje Meyer of the Behavioral Brain Science Center, University of Birmingham, and Alan Garnham of the Laboratory of Experimental Psychology, University of Sussex, for giving me the opportunity to work in their labs. Many thanks to Anne Cutler, James McQueen, and three anonymous reviewers for their helpful comments on earlier versions of this manuscript.

Baayen, H., Piepenbrock, R., and Gulikers, L. (**1995**). *The* CELEX *Lexical Database (CD-Rom)* (Linguistic Data Consortium, University of Pennsylvania, Philadephia, PA).

Best, C. T. (**1994**). ''The emergence of native-language phonological influences in infants: A perceptual assimilation model,'' in *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*, edited by J. C. Goodman and H. C. Nusbaum (MIT, Cambridge, MA), pp. 167–224.

Best, C. T., McRoberts, G. W., and Sithole, N. M. (**1988**). ''Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants,'' J. Exp. Psychol. Hum. Percept. Perform. **14**, 345–360.

Best, C. T., and Strange, W. (**1992**). ''Effects of phonological and phonetic factors on cross-language perception of approximants,'' J. Phonetics **20**, 305–330.

Booij, G. (**1995**). *The Phonology of Dutch* (Oxford University Press, Oxford).

Byrd, D. (**1993**). ''54 000 American stops,'' UCLA Working Papers **83**, 97–115.

Crowther, C. S., and Mann, V. (**1992**). ''Native language factors affecting use of vocalic cues to final consonant voicing in English,'' J. Acoust. Soc. Am. **92**, 711–722.

Crystal, T. H., and House, A. S. (**1988**). ''A note on the durations of fricatives in American English,'' J. Acoust. Soc. Am. **84**, 1932–1935.

Elsendoorn, B. A. G. (**1985**). ''Production and perception of Dutch foreign vowel duration in English monosyllabic words,'' Lang Speech **28**, 132–254.

Flege, J. E. (**1989**). ''Chinese subjects' perception of the word-final English /t/–/d/ contrast: Performance before and after training,'' J. Acoust. Soc. Am. **86**, 1684–1697.

Flege, J. E., and Hillenbrand, J. (**1987**). ''A differential effect of release bursts on stop voicing judgments of native French and English listeners,'' J. Phonetics **15**, 203–208.

Flege, J. E., and Wang, C. (**1989**). ''Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/–/d/ contrast,'' J. Phonetics **17**, 299–315.

Gay, T. (**1978**). ''Effect of speaking rate on vowel formant movements,'' J. Acoust. Soc. Am. **63**, 223–230.

Jongman, A., Sereno, J. A., Raaijmakers, M., and Lahiri, A. (**1992**). ''The phonological representation of [voice] in speech perception,'' Lang. Speech **35**, 137–152.

Jongman, A., Wayland, R., and Wong, S. (**2000**). ''Acoustic characteristics of English fricatives,'' J. Acoust. Soc. Am. **108**, 1252–1263.

Kuijpers, C. T. L. (**1996**). ''Perception of the voicing contrast by Dutch children and adults,'' J. Phonetics **24**, 367–382.

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

MacMillan, N. A., and Creelman, C. D. (**1991**). *Detection Theory: A User's Guide* (Cambridge University Press, Cambridge).

McNicol, D. (**1972**). *A Primer of Signal Detection Theory* (Australasian, Sydney, Australia).

Peterson, G. E., and Lehiste, I. (**1960**). ''Duration of syllable nuclei in English,'' J. Acoust. Soc. Am. **32**, 693–703.

Raphael, L. J. (**1972**). ''Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English,'' J. Acoust. Soc. Am. **51**, 1296–1303.

Repp, B. H. (**1981**). ''Perceptual equivalence of two kinds of ambiguous speech stimuli,'' Bull. Psychon. Soc. **18**, 12–14.

Schouten, M. E. H. (**1975**). ''Native-Language Interference in the Perception of Second-Language Vowels: An Investigation of Certain Aspects of the Acquisition of a Second Language,'' Doctoral dissertation, Utrecht University.

Slis, I. H., and Cohen, A. (**1969a**). ''On the complex regulating the voiced-voiceless distinction. I,'' Lang. Speech **12**, 80–102.

Slis, I. H., and Cohen, A. (**1969b**). ''On the complex regulating the voiced-voiceless distinction. II,'' Lang. Speech **12**, 137–155.

Slis, I. H., and Van Heugten, M. (**1989**). ''Voiced–voiceless distinction in Dutch fricatives,'' in *Linguistics in The Netherlands 1989*, edited by H.

Bennis and A. v. Kemenade (Foris, Dordrecht, Netherlands), pp. 123–132.
Smith, B. L. (**1978**). ''Effects of place of articulation and vowel environment on 'voiced' stop consonant production,'' Glossa **12**, 163–175.
Smith, B. L. (**2002**). ''Effects of speaking rate on temporal patterns of English,'' Phonetica **59**, 232–244.
Stevens, K. N., Blumstein, S. E., Glicksman, L., Burton, M., and Kurowski, K. (**1992**). ''Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters,'' J. Acoust. Soc. Am. **91**, 2979–3000.
Stevenson, D. C. (**1979**). ''Categorical Perception and Selective Adaptation Phenomena in Speech,'' Doctoral dissertation, University of Alberta, 1979.
Trubetzkoy, N. S. (**1977**). *Grundzuege der Phonologie*, 6th ed. (Van den Hoeck and Ruprecht., Goettingen [original work published 1939]).
Van Alphen, P. M., and Smits, R. (**2004**). ''Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: The role of prevoicing,'' J. Phonetics **32**, 455–491.
Van den Berg, R. J. H. (**1989**). ''Perception of voicing in Dutch two-obstruent sequences: Covariation of voicing cues,'' Speech Commun. **8**, 17–25.
Watson, I. (**1983**). ''Cues to the voicing contrast: A survey,'' Cambridge Papers in Phonetics and Experimental Linguistics 2.