Effect of masker type on native and non-native consonant perception in noise

M. L. Garcia Lecumberri[a)]
Department of English Philology, University of the Basque Country, Paseo de la Universidad 5, 01006, Vitoria, Spain

Martin Cooke[b)]
Department of Computer Science, University of Sheffield, Regent Court, 211 Portobello Street, Sheffield, S1 4DP, United Kingdom

(Received 24 August 2005; revised 4 February 2006; accepted 6 February 2006)

Spoken communication in a non-native language is especially difficult in the presence of noise. This study compared English and Spanish listeners' perceptions of English intervocalic consonants as a function of masker type. Three maskers (stationary noise, multitalker babble, and competing speech) provided varying amounts of energetic and informational masking. Competing English and Spanish speech maskers were used to examine the effect of masker language. Non-native performance fell short of that of native listeners in quiet, but a larger performance differential was found for all masking conditions. Both groups performed better in competing speech than in stationary noise, and both suffered most in babble. Since babble is a less effective energetic masker than stationary noise, these results suggest that non-native listeners are more adversely affected by both energetic and informational masking. A strong correlation was found between non-native performance in quiet and degree of deterioration in noise, suggesting that non-native phonetic category learning can be fragile. A small effect of language background was evident: English listeners performed better when the competing speech was Spanish.
 [DOI: 10.1121/1.2180210]

PACS number(s): 43.71.Hw, 43.71.Es, 43.66.Dc [ARB] Pages: 1–XXXX

I. INTRODUCTION

Spoken communication in noise presents problems for all listeners, but is especially difficult in a foreign language (FL[1]). Several features distinguish native and non-native experience of a given language. Differences in the degree, type, quality, and time of exposure to the language inevitably result in less familiarity with linguistic patterning at all levels, from acoustic to pragmatic. Non-natives also have to deal with the possibility of interference from their first language (L1). In the case of phonological acquisition, L1 influences have been shown to be particularly strong and pervasive (Ioup, 1984; Leather and James, 1991; Polka, 1995). The relative contribution made by these and other factors to foreign language perception in noise is not well understood.

A number of studies have compared native and non-native speech perception performance in noise and reverberation (Florentine *et al.*, 1984; Takata and Nábělek, 1990; Mayo *et al.*, 1997; Hazan and Simpson, 2000; van Wijngaarden *et al.*, 2002; Bradlow and Bent, 2002; Cutler *et al.*, 2004; van Wijngaarden *et al.*, 2004). These studies differed in the speech and noise material employed, the languages tested and the range of language proficiencies of the participants, but three basic findings have emerged. First, native performance on speech in noise tasks exceeded that of non-natives, even for a bilingual-since-infancy group (Mayo *et al.*, 1997). Second, increasing FL experience correlated well with a reduced effectiveness of masking noise (Florentine *et al.*, 1984; Mayo *et al.*, 1997). Third, non-native listeners were less able to take advantage of linguistic context to decode speech presented in noise (Mayo *et al.*, 1997; van Wijngaarden *et al.*, 2004).

Many studies (Florentine *et al.*, 1984; Mayo *et al.*, 1997; Takata and Nábělek, 1990) have suggested that the effect of noise is greater for non-native than for native listeners. This additional native advantage has also been found in studies of easy versus hard word recognition (Bradlow and Pisoni, 1999; Imai *et al.*, 2005). However, two recent studies found that non-native listeners were not more adversely affected by noise (Bradlow and Bent, 2002; Cutler *et al.*, 2004). Bradlow and Bent (2002) compared native and non-native performance on a keyword identification task using sentences mixed with white noise at signal to noise ratios (SNRs) of −4 and −8 dB. The fall in performance between these SNRs was similar for natives and non-natives. However, Bradlow and Bent (2002) acknowledge that this might have been due to a floor effect for non-natives at the lower SNR. Cutler *et al.* (2004) compared the abilities of native and non-native listeners in identifying consonants and vowels in VC and CV syllables presented in babble noise at SNRs of 16, 8, and 0 dB. They found no interaction between language background and noise level: the native advantage in performance was approximately the same for each noise level.

[a)]Electronic mail: garcia.lecumberri@ehu.es; Tel: +34 945013967; Fax: +34 945013200.
[b)]Electronic mail: m.cooke@dcs.shef.ac.uk; Tel: +44 114 2221822; Fax: +44 114 2221810.

The primary purpose of the current study was to examine the effect of different kinds of masker on native and non-native speech perception. Speech perception in noise is governed by both energetic and informational masking. The former arises from the masking of stimulus components at the auditory periphery and produces uncertainty or complete loss of information about the level of the target signal in spectro-temporal regions where the masker is sufficiently intense. Sufficient redundancy exists in clean speech to allow robust identification in many masking conditions (Assmann and Summerfield, 2004) based on available "glimpses" of the target signal (Cooke, 2006). Informational masking (Carhart *et al.*, 1969) refers to the potentially distracting effect of the masker. Intelligibility may suffer if attentional resources are directed at processing the masker, or if the allocation of signal energy to the foreground or masker is unclear.

The role played by energetic and informational masking in non-native speech perception has received little attention. The reduction in acoustic information associated with energetic masking may expose deficiencies in non-native mental representations built from a more limited or less optimal exposure to speech signals. Attentional overload may have a disproportionate influence on non-native perceptual processing which is already taxed by the difficulties of listening to non-native sounds and processing higher level morphological, syntactic, semantic, and pragmatic structures in the FL (Bradlow and Bent, 2002; Cutler *et al.*, 2004). In addition, interference can arise from the non-native listeners' L1 sound system, which may be more activated when listening conditions render the task more difficult. Indeed, Cutler *et al.* (2004) suggest that L1 categories may exert greater influence when the FL stimuli are more difficult to perceive. Similarly, Mayo *et al.* (1997) proposed that for speech material with low predictability, competition between the L1 and FL sound systems may contribute to the added difficulty of non-native perception in noise.

Previous studies of non-native speech perception in noise have employed either steady-state or babble maskers. While N-talker babble can be an effective informational masker even for large N for tasks involving nonsense syllables (Simpson and Cooke, 2005), strong informational masking effects are typically observed when the masker consists of other speech material from one or more competing talkers (Brungart *et al.*, 2001). Non-native speech perception in the face of competing speech has not been investigated to date. The use of competing speech, babble, and stationary noise in the current study permits a comparison of native and non-native performance in differing degrees of energetic and informational masking. The presence of speech in the background in FL perception also raises the possibility of differential language-dependent informational masking: it may be easier for listeners to ignore competing foreign language speech.

Since noisy conditions present added difficulties for all listeners, they constitute a good test for the investigation of native/non-native confusion patterns. It has been suggested that differences between native and non-native phonological performance are a reflection of non-native listeners' less well-developed phonetic categories caused by such factors as L1 interference and differences in the quantity, quality, and onset of FL exposure. At various stages of acquisition, non-native listeners' FL categories will be based on the influence of categories and cues from the FL and L1 to differing degrees. Two models, the Perceptual Assimilation Model (PAM) (Best, 1995) and the Speech Learning Model (Flege, 1995) propose that in FL sound perception, the influence of L1 phonetic categories can largely account for non-native listeners' identifications. PAM takes into account the degree of perceived similarity (exemplar rating) of FL sounds to NL categories. Thus, FL sounds may be considered "uncategorizable" within the listeners' L1 space or even regarded as "non-speech sounds." However, if the FL sounds are assimilated to some L1 sound, they may be seen as a "good" exemplar of the L1 category, or as either an "acceptable" or "deviant" exemplar. The present study examined native versus non-native confusion patterns in different listening conditions to investigate whether masking produces the same confusion patterns independent of a listener's L1 background.

Native speaker competence in tasks such as consonant identification may be expected to be fairly uniform. However, the diversity of FL competences is a problem that permeates all FL research and especially that concerning sound acquisition. It is well known that pronunciation may develop quite separately from other language skills (the *Joseph Conrad* phenomenon; Scovel, 1969). Bradlow and Bent (2002) mentioned that their non-native listeners were highly proficient in written English but were self-reported to have problems with their aural and oral skills. As learners' phonological competence improves, the distance between their developing FL system and that of native speakers' decreases, so they may perform increasingly like native speakers. A recent study by Imai *et al.* (2005) in a task involving natively spoken English word recognition found that the performance of Spanish speaking learners of English with a high phonological proficiency was more similar to native listeners than to low proficiency non-native listeners. In the present study, we were interested in within-group differences in non-native listeners concerning their performance in quiet and their degradation in identification rates in the presence of noise maskers. If non-native listeners' emergent FL sound system is robust, more competent listeners in quiet might be expected to show a smaller intelligibility reduction in masking conditions than listeners at a lower stage of FL phonological acquisition.

This study used English and Spanish listener groups to compare native and non-native perception of intervocalic consonants in quiet and in four noise maskers. Section II describes the listener groups, speech and masker corpus and experimental procedure. It also summarizes the main differences between the English and Spanish phonetic systems. Section III presents the results of native and non-native consonant identification in the four masking conditions.

II. METHODS

A. Participants

Twenty-one native speakers of British English and 61 native speakers of (European) Spanish participated in the

study. The English group was composed of monolingual students at the University of Sheffield whose age ranged from 18 to 24 years (mean: 21.4 years). English students were paid for their participation. The Spanish group consisted of students at the University of the Basque Country studying English as a foreign language (age range:20–25, mean: 21.2 years). They were enrolled in a one semester course in English Phonetics in the second year of a four year B.A. degree in English Language and Literature. As is typical in such courses, competence in English within the non-native group was not uniform, but all had attained the level of the Cambridge Advanced Exam. Spanish students received course credit for participating in the listening tests.

English listeners were screened for hearing loss (better than 20 dB hearing level in the range 250–8000 Hz) while Spanish listeners were asked to report hearing loss. One Spanish participant was excluded after reporting some hearing loss. Another Spanish listener left the course before completing all the tests and was also excluded. Results are presented for the remaining 59 non-native listeners.

Many of the listeners were native or second language speakers of Basque. However, the Spanish and Basque sound systems are so similar that this was not considered to be a relevant variable: for the English consonants in intervocalic context employed in the present study, the only difference between Spanish and Basque is that the latter has a voiceless palato-alveolar phoneme /ʃ/. However, non-Basque speakers are familiar with this sound due to language contact, since it appears in some widely-known words such as "kaixo" /kaiʃo/ ("hello") and "xabi" /ʃabi/ (a common forename).

B. Differences between the English and Spanish phonetic systems

The English and Spanish consonant systems mainly differ in the following respects. Both languages have six plosives arranged in three voiced/voiceless pairs. However, Spanish intervocalic voiced plosives are lenited to approximants and have voicing lead, whereas English voiced plosives have voicing lag or are devoiced (Navarro Tomás, 1918; Harris, 1969). English voiceless plosives are often aspirated whereas in Spanish they are unaspirated in most accents. Unlike English alveolar plosives, Spanish /t d/ have a lamino-dental articulation.

The main difference for fricatives and affricates is the absence of voiced fricative phonemes in most varieties of Spanish. Whereas English has /f v θ ð s z ʃ ʒ tʃ dʒ h/, standard peninsular Spanish has /f θ s tʃ x/, although in many southern Spanish and Latin American accents /θ/ is absent and /s/ is used in its place. The sound /j/ is variably classified as a weak fricative, an approximant or a realization of /i/ (Hualde, 2005). The other main difference between the two languages is that the [+back] fricative is velar in Spanish and glottal in English. The place of articulation of Spanish /s/ varies substantially across regional varieties. In our listeners' accent (Northern-Central Spain) /s/ is realized as an apico-alveolar whereas English /s/ can be characterized as lamino-alveolar. Some of the English voiced fricatives can be found as contextual variants or in other accents of Spanish. For instance, [v] is a variant of /b/ in some areas of Andalucia; [z] is a contextual allophone of /s/ before a voiced consonant; [ʒ] is the realization of /ʝ/ and /ʎ/ in Argentinian Spanish; [h] is found as a realization of /x/ in some areas such as the Canary Islands, Caribbean or as a realization of coda /s/ followed by consonant in central Spain; /dʒ/ is a variant of /ʝ/.

As for the nasals, Spanish has a palatal nasal phoneme /ɲ/ whereas the velar nasal is always a contextual allophone of /n/. Spanish has a larger inventory of liquids. As opposed to English postalveolar or retroflex approximant /ɹ/ Spanish contrasts an alveolar trill /r/ and a tap /ɾ/. The alveolar lateral which is often but differently velarized in American and British English is only velarized in Spanish as a result of anticipatory assimilation before velar plosives. The Spanish palatal lateral phoneme /ʎ/ is receding in most accents, converging with /ʝ/. However, because /ʎ/ is also a phoneme in Basque, it is still considerably present in the Spanish spoken in the Basque Country, though gradually giving way to /ʝ/ too. Finally, the English approximants /w j/ are present in Spanish either as allophones of the corresponding high vowels in rising diphthongs or, in the case of [j], as a variant of /ʝ/ (Navarro Tomás, 1918; Harris, 1969; Hualde, 2005).

A further issue is the differing relationship between phonemes and graphemes in the two languages. English is well known for the opacity of its spelling, whereas Spanish has often been called a "phonemic language." These statements are quite near the truth but need qualification. It is true that some English phonemes, particularly vowels and consonants such as /z θ ð/, are difficult to interpret from orthography; "g" may represent /g/ and /dʒ/, while "r" is only pronounced prevocalically in some varieties of English. However, other consonants such as /p t k b d v m n l/ have a much clearer spelling correspondence. On the other hand, Spanish sound-letter correspondence breaks down in a few cases. For instance, /b/ is represented by both "b" and "v" and "h" is not pronounced.

C. Speech and noise materials

Speech tokens were drawn from the vowel-consonant-vowel (VCV) corpus collected by Shannon *et al.* (1999). Although 23 consonants were available, a subset of 16 consonants /p b t d k g tʃ f v s z ʃ m n l r/ was employed. Members of the chosen subset have a clear consonantal character and are unambiguous with respect to their orthographic representations. This latter factor was important since the English listener group was not familiar with phonetic symbols. Consonants such as /θ ð ʒ ʒ/ with a problematic orthographic-phoneme correspondence were not used, while /ŋ/ was excluded because it is not phonotactically possible in the vowel context chosen. The vowel context used was /ɑ_ɑ/ and remained constant to avoid coarticulatory differences between stimuli. This vowel context has been found to be better than high back and front vowels for consonant identification because of its well-defined F1 and F2 transitions which favor consonant place and voicing identification (Hazan and Simpson, 2000).

Two tokens of each VCV from each of five male talkers

made up a test set of 160 items. Data for the English listeners in the quiet condition was taken from a pretest condition of a larger study (which involved the same group of native listeners as the current study) and contained 80 items rather than 160. Tokens were normalised to have equal root-mean-square (rms) energy and sampled at 25 kHz. An additional two examples of each token were used to create a set of 32 practice items which were appended to the front of the test set but not scored. Listeners were not informed about the presence of the practice items.

Four masker types—8-talker babble, speech-shaped noise, and competing English and Spanish speech—were employed. All maskers apart from the Spanish speech were derived from male talkers within dialect regions d1, d2, and d3 of the TIMIT corpus (Garofolo *et al.*, 1992). Babble was produced by summing utterances whose rms energy had been equalized. Speech-shaped noise was generated by passing white noise through a filter whose magnitude response was the long-term average spectrum of speech material from the aforementioned dialect regions of TIMIT. Competing English sentences were chosen at random from TIMIT. Spanish speech material was provided by a native male speaker of the language at the University of the Basque Country.

VCVs were combined with each of the four masker noise types such that the masker signal started 1 s before the onset of the VCV and continued up to the VCV offset. Leading maskers were chosen to increase the likelihood of informational masking effects in the single competing talker conditions. Speech and noise signals were combined such that the SNR in the overlapped region was 0 dB for each token. The duration of the overlapped region varied from 0.57 to 1.21 s (mean=0.88, s.d.=0.14).

D. Procedure

The native group was tested at the University of Sheffield while the non-native group was tested at the University of the Basque Country. At both sites, stimulus presentation and response collection was under computer control. Participants used a mouse to select their response category from a 4×4 grid representing an orthographic version of the 16 consonants presented on a computer screen. English listeners were not familiar with phonetic symbols but their native knowledge and literacy made it possible to use orthographic symbols. Although Spanish listeners had been introduced to phonetic symbols in other English language courses, the degree to which they were familiar with English sounds was not known since they had just started a 15-week English Phonetics course. Both groups of listeners were informed that the test involved the identification of English consonants. Participants were instructed to listen out for the consonant that they would hear between vowels in nonsense words and click on the corresponding symbol box. After a short delay, the next stimulus was presented. In this way, participants governed the presentation rate of the stimuli.

Each of the five conditions required 8–10 min to complete. At the University of Sheffield, participants were tested individually in an IAC single-walled acoustically-isolated booth using Sennheiser HD250 headphones. At the University of the Basque Country, participants were tested in groups of 15–20 in a quiet laboratory using Plantronics Audio-90 headphones. Stimuli were presented diotically and listeners were able to adjust the level to a comfortable setting.

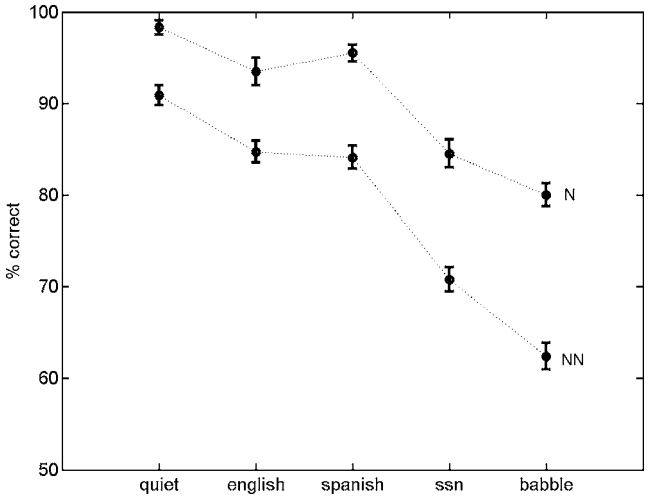

FIG. 1. Consonant identification scores of native (N) and non-native (NN) listeners in quiet and four masking conditions, each at a SNR of 0 dB: English=competing talker in English, Spanish=competing talker in Spanish, ssn=speech-shaped noise, babble=8-talker babble noise. Dotted lines indicate which points belong to the same listener group, and error bars define 95% confidence intervals.

To determine whether the difference in stimulus presentation equipment in the two countries could influence the results, a separate group of seven native English listeners were tested at the University of Sheffield in quiet and in the speech-shaped noise condition using a setup (quiet room, PC soundcard, Plantronics Audio-90 headphones) similar to that used in the University of the Basque Country. Identification results were compared with those obtained in quiet and speech-shaped noise by the main English experimental group. Means for both quiet (99.7 vs 98.3; $t(26)=1.44$, $p=0.162$) and speech-shaped noise (83.4 vs 84.6; $t(26)=-0.649$, $p=0.404$) were not significantly different.

Listener performance in quiet was measured first. Listeners were then presented with the four masking conditions in a random order. All five conditions (quiet plus the four maskers) were tested in a single session lasting 45–50 min.

III. RESULTS

A. Effect of masker type

Figure 1 compares native and non-native identification of consonants in quiet and in the four masking conditions. As expected, performance deteriorated for both groups in all masking conditions. A repeated measures ANOVA with one within-subjects factor (masker type) and one between-subjects factor (nativeness) confirmed the effect of masker ($F(4,75)=375$, $p<0.0005$, $\eta^2=0.952$) and group ($F(1,78)=212$, $p<0.0005$, $\eta^2=0.731$). Listeners in both groups performed best in the presence of competing speech and worst in a babble masker. Spanish listeners suffered more from the presence of noise. Post hoc comparisons (with Bonferroni adjustment for multiple comparisons) showed statistically significant ($p<0.0005$) differences between natives and non-

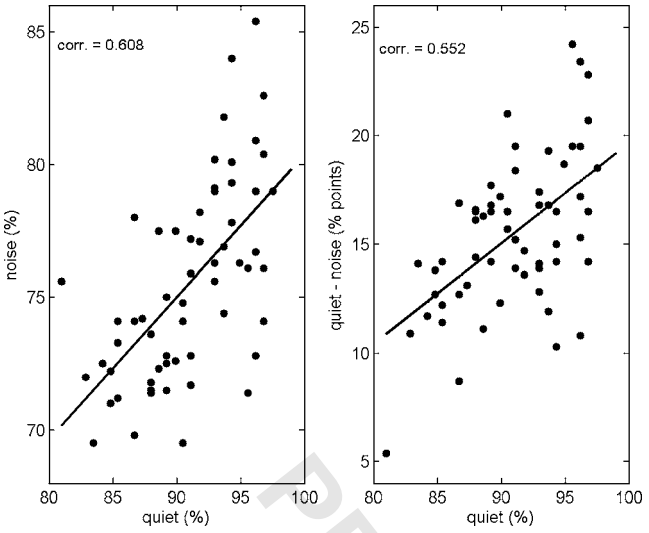

FIG. 2. Scatter plots of non-native performance in quiet versus noise (left panel) and quiet vs degradation in noise, measured as the percentage points difference between quiet and noise (right panel). Each point represents an individual listener. Noise scores are averaged over the four masking conditions. The best (least squares) linear fit and Pearson correlation coefficient are shown.

natives in each noise condition. Similarly, post-hoc comparisons of noise conditions demonstrated significant ($p < 0.0005$) differences in masking effectiveness apart from the two competing talker conditions ($p > 0.05$).

The interaction between masker type and nativeness was highly significant ($F(4,75) = 15.4$, $p < 0.0005$, $\eta^2 = 0.452$). The native advantage of 7 percentage points in quiet increased to an average of 10 for the two competing speech conditions, 14 for speech-shaped noise and 18 for babble. All pairwise comparisons of noise conditions (with the two competing speech conditions combined into a single mean measure) showed a significant interaction between masker type and nativeness. The largest interaction with nativeness was between quiet and babble ($F(1,78) = 60.24$, $p < 0.0005$, $\eta^2 = 0.436$) while the smallest was between the mean of the competing talker conditions and speech-shaped noise ($F(1,78) = 8.37$, $p < 0.005$, $\eta^2 = 0.097$).

B. Within-group differences for non-native listeners

As expected, the native group exhibited far less variability in quiet conditions than the non-native group (s.d. of 0.9 for natives vs 4.2 for non-natives). Non-native groups are far less homogenous than native listeners due to differences in the onset, amount and quality of exposure to the FL and the stage of language learning. Given the spread of identification rates in quiet for non-native listeners, it was possible to explore the extent to which performance in noise is predictable on the basis of performance in quiet conditions. Such an analysis is not meaningful for the native listeners since very few errors were made in quiet by this group. Figure 2 (left panel) plots performance in quiet against mean performance in the 4 masking conditions for each non-native listener. A clear relationship is visible: the better the ability to recognize VCVs in quiet, the better the performance in noise (Pearson correlation $= 0.608$, $p < 0.001$). In fact, non-native performance in quiet was significantly correlated with identification rates in each of the four masking conditions individually.

If the spread of non-native performance in quiet represents different levels of phonetic competence in the foreign language, the interaction between noise condition and nativeness described above could be explained by disproportionately worse performance in noise by the least competent non-natives. If so, the deterioration in performance between quiet and noisy conditions should be negatively-correlated with performance in quiet. Figure 2 (right panel) plots the difference in percentage points between identification rates in quiet and the mean of the four noise conditions as a function of performance in quiet for each non-native listener. Contrary to the prediction that the least competent listeners in quiet conditions would show greater deterioration in noise, the reverse was found (Pearson correlation $= 0.552$, $p < 0.001$). Those listeners who performed well in quiet displayed the greatest absolute drop in identification rates in noise relative to quiet. Significant correlations were found for each of the four noise conditions independently.

C. Language of competing speech

A planned comparison of the effect of language for the competing talker maskers showed no significant difference overall. However, there was a significant interaction of language and nativeness ($F(1,78) = 6.82$, $p = 0.011$, $\eta^2 = 0.08$). Within-group comparisons revealed that this was due to a small but significant difference for the native group ($F(1,78) = 5.23$, $p = 0.025$, $\eta^2 = 0.063$). The language of the competing talker had no effect on the non-native group. This result suggests that English listeners were better able to tune out an unknown language, while both maskers were equally disturbing for Spanish learners of English.

D. Consonant identification

Figure 3 compares native and non-native identification scores for each of the 16 consonants in quiet (upper panel) and averaged over the four masking conditions (lower panel). Figure 4 depicts the native advantage, measured in percentage points, in quiet (upper panel) and in noise (middle panel). The lower panel of Fig. 4 shows the additional native advantage in noise over that in quiet, calculated as the difference between the measures in the middle and upper panels. Tables I to IV show consonants confusions for the two groups in quiet and in noise. Consonants /b v f s z/ are identified much less accurately by non-natives in quiet (Fig. 4, upper panel). However, it is striking that for most of these sounds, non-natives are not further disadvantaged in noise (Fig. 4, lower panel). In fact, most of the increased native advantage in noise comes from poorer identification of the consonants /p d g k l r tʃ f/. The remaining consonants /m n t ʃ/ are identified both in quiet and in noise at similar rates by both groups.

Among the sounds for which non-native listeners' perceptions are considerably worse in quiet conditions, some clear L1 influences can be observed. Sounds such as /v/ and /z/ which are not part of the Spanish phonological inventory,

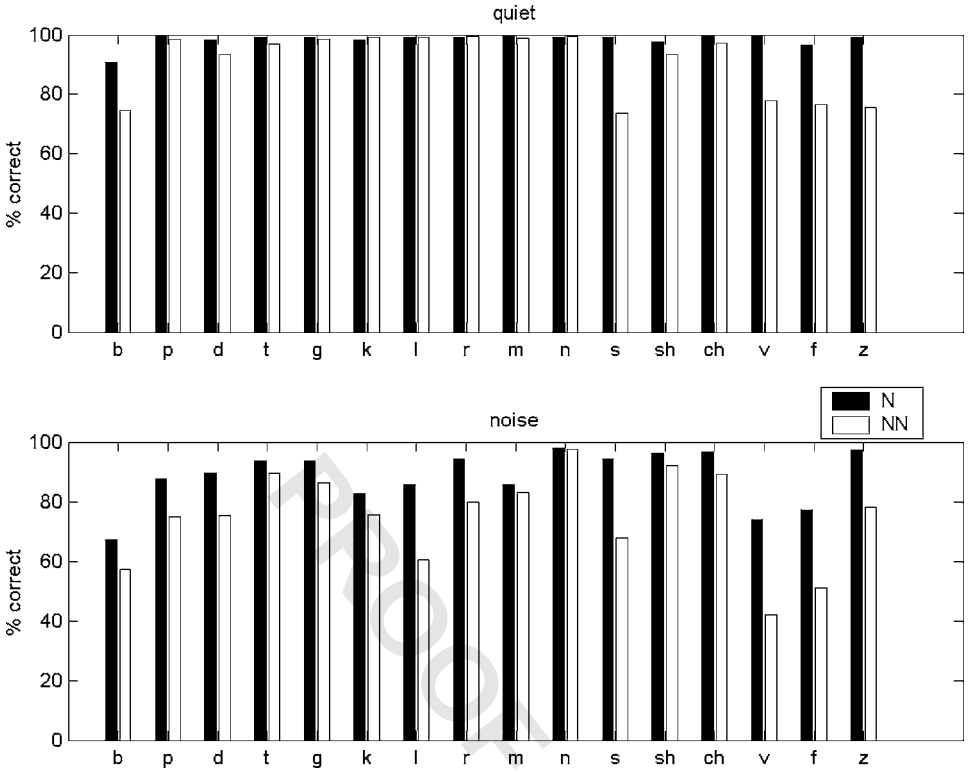

FIG. 3. Native and non-native identification scores for individual consonants in quiet (upper panel) and averaged over the noise conditions (lower panel). Here, and in Fig. 4, labels "sh" and "ch" correspond to the consonants ʃ and tʃ.

are not, nevertheless, "uncategorizable" in L1 terms (Best, 1995). Rather, they represent what PAM would call "deviant" to "acceptable" exemplars of the L1 phonetic categories /b/ and /s/, respectively. The sound [v] is a realization of /b/ in some varieties and, further, orthographically "b" and "v" are realized as /b/. Thus, English /v/ may be perceived by Spanish listeners as a deviant realization of /b/. Indeed, as the confusion matrix of Table II shows, non-native listeners' mainly misidentify /v/ in quiet as /b/. Contrary to appearances, English intervocalic /b/ may be classified as a "poor" exemplar of Spanish /b/ since the phonetic cues that code intervocalic voiced plosives in Spanish (full voicing and lenition) are quite different to those used in English and can even result in an acceptable exemplar of /p/. Table II shows that non-native listeners hear /b/ as /p/ 21% of the time. Regarding /s/ and /z/, the former is a phonetic category in both English and Spanish with realizational differences while [z] is a contextual allophone in Spanish. Consequently, both

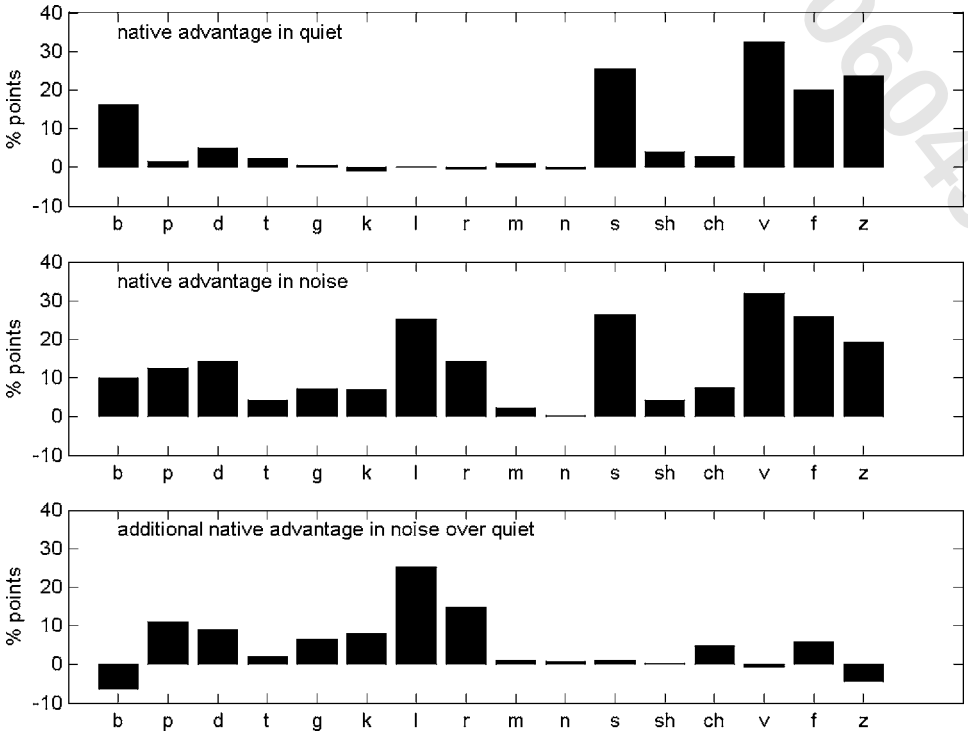

FIG. 4. Differences in native and non-native consonant identification rates, measured in percentage points. The upper panel shows the native advantage in quiet. The middle panel displays the native advantage in noise, where the noise scores are averages over the four masking conditions. In the lower panel, the additional native advantage in noise over quiet is shown. This is simply the difference between middle and upper panels.

PROOF COPY 060604JAS

TABLE I. Consonant confusions in quiet for the native group, expressed as percentages. Rows represent stimuli presented while columns denote participant responses. Rows may not sum to 100 since percentages are rounded.

	b	p	d	t	g	k	l	r	m	n	s	ʃ	tʃ	v	f	z
b	91	3	5		1											
p		100														
d			99	1												
t				99			1									
g		1			99											
k						98							2			
l							99							1		
r		1						99								
m									100							
n					1					99						
s			1								99					
ʃ												98	2			
tʃ													100			
v								1			1			90		9
f			1	1							2				96	
z											1					99

TABLE II. Consonant confusions in quiet for the non-native group.

	b	p	d	t	g	k	l	r	m	n	s	ʃ	tʃ	v	f	z
b	75	21	4													
p		98		1												
d			93	6												
t				97									2			
g					98	1										
k						99										
l			1				99									
r								99								
m									99	1						
n										100						
s											74	5				21
ʃ											5	93	1			1
tʃ				2									97			
v	15													78	3	3
f											5			3	77	15
z											18	7				75

FL sounds fall under the one L1 category with different goodness ratings, which accounts for their nonoptimal identification levels (Best, 1995). This is borne out by the confusions in Table II since there are mutual confusions between the two categories: /z/ is misperceived as /s/ on 18% of presentations and /s/ is classified as /z/ 21% of the time. The remaining consonant showing greatest native advantage in the quiet condition is /f/ which is very similar to a NL prototype. Here, the high level of confusion is not predicted by models such as those of Best (1995) and Flege (1995), unless there is another FL sound competing in the same prototype area. Indeed, /v/ could be a candidate. However, an examination of the confusion matrix suggests that /f/ is most often confused with /z/, which points towards an acoustic influence, with friction as the salient perceptual cue.

Most of the sounds which show the greatest native advantage in quiet are not additionally disadvantaged in noise (Fig. 4). In fact, two categories, /z/ and /b/, show *reduced* native advantage in noise. In the case of /z/, as shown by the perception scores in Fig. 4 (upper panel), this could indicate that the voicing feature used to identify it is resistant to noise maskers (Miller and Nicely, 1955; Hazan and Simpson, 2000). For /b/, the decrease in native advantage is due to a large drop in performance for native listeners in noise (Fig. 3, lower panel, and Table III). Confusion patterns in noise (Tables III and IV) show that native listeners most often confuse /b/ with /v/ whereas non-native listeners confuse it predominantly with /p/. This disparity indicates the use of different cues in noise for the two groups. For native listeners, voicing appears to be the most salient cue while non-native listeners seem to employ the same L1 cues as those used in the quiet condition, namely voice onset time and lack of lenition. Therefore, L1 influences appear to determine confusions but this influence is not stronger in noise.

For the sounds /p d g k l r/ the native advantage only becomes apparent in noise. An examination of their confusion patterns shows that both voiceless plosives are mainly confused with other voiceless plosives, a pattern which is also visible for native listeners, suggesting that place of articula-

TABLE III. Consonant confusions in noise (averaged across the four masking conditions) for the native group.

	b	p	d	t	g	k	l	r	m	n	s	ʃ	tʃ	v	f	z
b	67	3	2	1			6	2	2				1	14	1	
p	2	88		1		5								1	2	
d			89		7		1			2						
t		3	1	94		2										
g	1		1		94	1				1				1		
k	1	11		1	1	83									2	
l	1						86	3	6	1				2		
r	1	1					1	94						3		
m	1						4		86	7					1	
n									1	98						
s											94				1	2
ʃ			1									96	2			
tʃ						1						1	97			
v	13	2					2	3	3					74	2	
f	5	5	1			3	1							8	77	
z											1			1		97

TABLE IV. Consonant confusions in noise for the non-native group.

	b	p	d	t	g	k	l	r	m	n	s	ʃ	tʃ	v	f	z
b	57	19	4	2	3	1	2	1	1				1	7	2	
p	3	75	1	4	2	10			1					1	2	
d	1		75	7	11	1	1	1		2						
t		1	2	89		3					1		1			1
g	1		3	1	86	6		1		1				1		
k	1	11	1	5	2	76									2	
l	7	3	3		2		60	4	12	3	1			4		1
r	5				5		3	80	3					3		
m	1	1			1		3	1	83	7	1			1	1	
n									1	98						
s				1							68	4			1	25
ʃ											4	92	2			1
tʃ			1	2	1	1					1	5	89			
v	29	9	1	1	5	1	1	2	3					42	5	1
f	6	21	1	1	1	3			1		3		1	6	51	4
z											16	4				78

PROOF COPY 060604JAS

tion information is less resistant to noise than voicing (Miller and Nicely, 1955). Non-native listeners confuse /g/ with its voiceless counterpart, following the same influences observed above for /b/ confusions, i.e., L1 cues for differentiating voiced/voiceless plosives. Finally, the two liquids show the biggest native advantage in noise and the greatest degree of confusion dispersions. In both cases there are prototypes in the learner's L1 system to which the English category could be assimilated as a good or acceptable exemplar, predicting good identifications (Best, 1995): English /ɹ/ could be related to both Spanish /r/ and /ɾ/; English nonvelarized /l/ is similar to Spanish /l/. Indeed, this is the pattern observed in the quiet condition. However, the considerable realizational differences (/l/ velarization and /r/ retroflexion in American English) between the FL and the L1 phonetic categories show up under the strain of unfavourable listening conditions. Some of the non-native confusions for /l/ are also visible to a lesser degree in the native group.

For the sounds /f/ and /v/ which are poorly identified by non-natives in quiet and in noise, native listeners' perceptions also deteriorate considerably in noise. These sounds are confused with similar categories by both groups. Non-natives mainly confuse /f/ with /p/ in noise rather than /s/ in quiet, but in both cases inherent susceptibility to masking rather than L1 influences is the likely cause. The case of /v/ is particularly interesting. In noisy conditions, native and non-native confusions are dominated by /b/, suggesting that friction is masked by noise. This is a natural, perhaps universal, confusion with historical instantiations such as the merging of /v/ and /b/ in Spanish and the change of Latin intervocalic /b/ to /v/ in some Romance languages.

IV. DISCUSSION

A. Effect of different masker types on native versus non-native consonant identification in noise

On a consonant identification task, English outperformed Spanish listeners in quiet conditions by 7 percentage points. The performance differential in all masking conditions was larger than in quiet, ranging from 9 to 18 percentage points. Masking conditions which caused most errors for natives were also those which resulted in larger differences in native and non-native identification rates.

The ranking of masking effectiveness of the three noise types employed in this study was identical for natives and non-natives. Both groups identified consonants in a competing talker background at a higher rate than in a steady noise masker, and both performed worst when the masker was 8-talker babble. This ranking of masking effectiveness has been found in other studies using native listeners (e.g., Simpson and Cooke, 2005) and probably reflects the relative amount of energetic and informational masking produced by the various masker types when added at a fixed SNR.

A single competing talker produces less energetic masking than 8-talker babble, which in turn is a less effective energetic masker than speech-shaped noise. However, speech-shaped noise has no additional informational masking effect, while both babble and competing speech can produce significant amounts of informational masking (Carhart *et al.*, 1969; Brungart *et al.*, 2001; Freyman *et al.*, 2004). The degree of informational masking obtained depends on several factors such as the similarity of target and masker and the number of talkers in the background. For example, Freyman *et al.* (2004) demonstrated a maximal effect of informational masking with two talkers in the background.

The study by Simpson and Cooke (2005) is particularly relevant, since it used the same VCV corpus in a large number of masking conditions, including the three employed here, albeit at a more adverse fixed SNR of −6 dB. Simpson and Cooke demonstrated that a competing talker produces negligible amounts of informational masking for this VCV corpus. Instead, maximal informational masking arises from 8-talker babble. Consequently, the masking effect of babble in the current study may be interpreted as due to the combined effects of energetic and informational masking. The 14 percentage point difference in performance between the two groups in stationary noise is less than the 18 percentage points deficit in babble, which has a smaller energetic masking effect than stationary noise. Consequently, it is reasonable to conclude that non-native listeners are more adversely affected then native listeners by informational masking. Further studies with more confusable competing speech tasks are required before the native informational masking advantage can be quantified.

Regarding the effect of the two competing speech conditions, there was a small effect of masker language for the native group but not for non-native listeners. Native listeners showed better performance when the language of the masker was unknown to them (Spanish) than when it was their L1 (English). In contrast, non-native listeners performed at a similar level irrespective of whether the masker language was English or Spanish, presumably because they spoke both languages. This suggests that the strong attentional component of competing speech had a larger effect when the language was known to the listeners. Although FL listeners commonly report that they are able to tune out FL speech better than their L1, in this case the fact that the task involved FL consonant identification may have made it harder to tune out the FL masker (Cutler, personal communication). Since the effect found was relatively small, further studies with other languages and speech material are necessary. If a language unknown to both groups were to be found to constitute a less effective masker for both, the asymmetry found in the present study could be interpreted as the result of reduced attentional demands. Sentence length speech material in the foreground might be expected to produce stronger language activation for the FL group.

The current study found a significant interaction between masker type and nativeness. The closest work to that reported here is Cutler *et al.* (2004), who found no interaction between noise level and nativeness. There are a number of differences between Cutler *et al.* (2004) and this study. First, Cutler *et al.* (2004) varied masking effectiveness by presenting tokens at a number of SNRs, while the current study employed different masker types known to differ in their degree of energetic masking. There is no direct way to compare the effectiveness of the maskers used in the two studies. However, the most difficult condition in both studies

occurred with babble at 0 dB SNR. Further, it is known that a competing speaker provides 6–8 dB less masking than stationary noise at the same level (Miller, 1947; Festen and Plomp, 1990). Since Cutler *et al.* (2004) used babble noise at 16, 8, and 0 dB and our stationary noise masker produced less masking than the babble condition, there are grounds for arguing that the range of masker effectiveness was broadly similar in the two studies.

A second difference between the two studies concerns the non-native samples (Dutch and Spanish). Dutch learners of English have been shown to display outstanding phonetic performance in English (Bongaerts, 1999; Broersma, 2005). In part, this is due to the quality and quantity of English exposure in Holland compared to that in Spain. In addition, interlanguage phonological distance, which has been considered to be a crucial factor in FL sound acquisition (Bongaerts *et al.*, 2000; Hazan and Simpson, 2000), is considerably smaller between English and Dutch than between English and Spanish.

There are other differences between the two studies. Cutler *et al.* (2004) used CV and VC syllables rather than VCV tokens and asked listeners to identify all phonemes rather than just the consonants. VCV tokens used in the current study were produced by American speakers while the native listeners were British. However, Cutler *et al.* (2005) demonstrated that American and Australian English listeners produced statistically indistinguishable performance on a task involving a subset of the stimuli employed by Cutler *et al.* (2004). Further, the placement of tokens relative to maskers differed in the two studies in such a way that listeners may have been able to better predict the onset of the VCV tokens used in the current study.

B. Fragility of non-native categories

Non-natives were less able to identify consonants in quiet conditions, which agrees with most of the literature on FL sound perception (Pisoni *et al.*, 1994; Best, 1995; Flege, 1995). Clear L1 influences were observed among consonant confusion patterns. The additional native advantage seen in the presence of noise (Fig. 4, lower panel) for sound identification tasks such as that employed here which involved low-level phonetic processing, might be explained by interaction of the L1 and FL phonetic systems (Mayo *et al.*, 1997) compounded by noise degradation (Hazan and Simpson, 2000). Noise may expose the lack of robustness of a non-native listener's FL categories and reveal the use of different phonetic cues for certain FL categories due to L1 interference and incomplete FL sound acquisition. The L1 experience of a native listener will include exposure to adverse conditions, so native identification is more likely to involve the use of multiple, redundant cues and appropriate cue weighting strategies to overcome the effects of energetic masking. In contrast, non-natives may have developed fewer cues and less sophisticated weighting strategies due to limited or faulty exposure to the FL categories and less experience with the FL in noisy conditions. Additionally, non-native listeners may use cues influenced by their L1, and such cues may be affected by noise in a different manner from those used by native listeners. For example, Spanish listeners may make less use of aspiration and listen for lenition to differentiate voiceless/voiced plosives.

Noise constitutes a good testing ground in which to compare native and non-native consonant identification since native performance departs from the near ceiling levels observed in a quiet background. Thus noise can be used to examine whether confusions pattern similarly for both listener groups and hence distinguish between "universal" misidentifications due to acoustic factors such as inherent maskability, and non-native specific confusions ascribable to L1 influences.

The fact that some consonants only show native advantage in noise indicates that the phonetic categories non-native listeners were using in less demanding situations (and which in some cases allowed them to reach the level of native listeners' performance) are too fragile to withstand adverse listening conditions. Therefore, although perception in quiet reveals important information about a non-native listener's level of sound acquisition, it may still be seen as a "performance" measure that only taps indirectly into "competence." As has been suggested (Mayo *et al.*, 1997; Pallier *et al.*, 1997; Bosch *et al.*, 2000), even bilinguals can be shown to differ in competence from monolingual native speakers, although these differences only become apparent under careful testing. Sound perception in noise, in which some of the usual phonetic cues are not reliable due to energetic and informational masking, appears to be a good test of the robustness of FL categories and phonetic competence.

V. CONCLUSIONS

English listeners outperformed Spanish listeners in a task involving the identification of medial consonants in English VCV tokens. The native performance advantage increased when tokens were presented at a fixed SNR of 0 dB in speech-shaped noise, 8-talker babble and in two competing speech maskers. Since stationary noise has no informational masking effect, this result suggests that non-native listeners are more adversely affected by pure energetic masking than are native listeners. The ranking of masking effectiveness across masking conditions was identical for the native and non-native groups, with competing speech being the least effective masker and babble the most challenging. Given that, at a fixed SNR, babble is a less effective energetic masker than stationary noise, these findings also suggest that non-native listeners were more adversely affected by informational masking.

Non-native listener performance in quiet correlated well with performance in noise. However, performance degradation in noise was also positively correlated with identification rate in quiet, suggesting a lack of robustness in non-native FL phonetic categories which is only apparent under adverse conditions such as speech perception in noise.

Spanish listeners performed at the same level in the presence of both English and Spanish competing speech maskers. However, English listeners were slightly better

when the language of the competing speech was Spanish as opposed to English, perhaps due to a reduced attentional engagement for an unknown language.

At the level of individual consonants, those most poorly identified by non-natives in the quiet condition did not suffer any further disadvantage in noise. Most of the additional native advantage was due to a subset of consonants which were identified well in quiet conditions by both native and non-native listeners. A more comprehensive study of consonant identification in noise might reveal important information about the structure of phonetic categories used by non-native learners. Finally, further studies with words and sentences are necessary to discern the effect of competing talkers, masker language, and attentional factors on native and non-native perception in everyday speech conditions.

ACKNOWLEDGMENTS

This work was supported by grants from the Spanish Ministry of Science and Technology and the Basque Government (BFF 2003-04009; 9/UPV 00103.130). The authors thank Sarah Simpson for help in collecting data from the English listeners and Ann Bradlow, Anne Cutler, and two anonymous reviewers for their valuable suggestions.

[1]Second language (L2) learning usually refers to learning a language in the area where it is spoken, after another language (L1) has been acquired. Foreign language (FL) indicates that the language is not a language in the community and thus it has a foreign status (notwithstanding the fact that a FL learner can go the FL country to pursue its learning). In this paper, the term FL is used to refer to learning a language after the L1 has been established.

Assmann, P. F., and Summerfield, Q. (**2004**). "The perception of speech under adverse acoustic conditions," in *Speech Processing in the Auditory System, Springer Handbook of Auditory Research*, edited by S. Greenberg, W. A. Ainsworth, A. N. Popper, and R. R. Fay (Springer, Berlin), Vol. **18**.

Best, C. T. (**1995**). "A Direct Realist View of Cross-Language Speech, Perception," in *Speech Perception and Linguistic Experience*, edited by W. Strange (Timonium, MD), pp. 171–204.

Bongaerts, T. (**1999**). "Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners," in *Second Language Acquisition and the Critical Period Hypothesis*, edited by D. Birdsong (Erlbaum, Mahwah), pp. 133–159.

Bongaerts, T., Mennen, S., and van der Slik, F. (**2000**). "Authenticity of pronunciation in naturalistic second language acquisition: The case of very advanced late learners of Dutch as a second language," Studia Linguistica **54**, 298–308.

Bosch, L., Costa, A., and Sebastián-Gallés, N. (**2000**). "First and second language vowel perception in early bilinguals," European J. of Cognitive Psychology **12**, 189–221.

Bradlow, A. R., and Bent, T. (**2002**). "The clear speech effect for non-native listeners," J. Acoust. Soc. Am. **112**, 272–284.

Bradlow, A. R., and Pisoni, D. B. (**1999**). "Recognition of spoken words by native and non-native listeners: Talker-, listener-, and item-related factors," J. Acoust. Soc. Am. **106**, 2074–2085.

Broersma, M. (**2005**). "Perception of familiar contrasts in unfamiliar positions," J. Acoust. Soc. Am. **117**, 3890–3901.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic màsking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **100**, 2527–2538.

Carhart, R., Tillman, T. W., and Greetis, E. S. (**1969**). "Perceptual masking in multiple sound backgrounds," J. Acoust. Soc. Am. **45**, 694–703.

Cooke, M. P. (**2006**). "A glimpsing model of speech perception in noise," J. Acoust. Soc. Am. **119**, 1562–1573.

Cutler, A., Weber, A., Smits, R., and Cooper, N. (**2004**). "Patterns of English phoneme confusions by native and non-native listeners," J. Acoust. Soc. Am. **116**, 3668–3678.

Cutler, A., Smits, R., and Cooper, N. (**2005**). "Vowel perception: Effects of non-native language vs non-native dialect," Speech Commun. **47**, 32–42.

Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.

Flege, J. E. (**1995**). "Second language speech learning: Theory, findings and problems," in *Speech Perception and Linguistic Experience*, edited by W. Strange (York, Timonium), pp. 233–277.

Florentine, M., Buus, S., Scharf, B., and Canevet, G. (**1984**). "Speech reception thresholds in noise for native and non-native listeners," J. Acoust. Soc. Am. **75**, s84.

Freyman, R. L., Balakrishnan, U., and Heifer, K. S. (**2004**). "Effect of number of masking talkers and auditory priming on informational masking in speech recognition," J. Acoust. Soc. Am. **115**, 2246–2256.

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., Pallett, D. S., and Dahlgren, N. L. (**1992**). "DARPA TIMIT Acoustic Phonetic Continuous Speech Corpus CDROM," NIST.

Harris, J. W. (**1969**). *Spanish Phonology* (MIT Press, Cambridge).

Hazan, V., and Simpson, A. (**2000**). "The effect of cue-enhancement on consonant intelligibility in noise: Speaker and listener effects," Lang Speech **43**, 273–294.

Hualde, J. I. (**2005**). *The Sounds of Spanish* (Cambridge University Press, Cambridge).

Imai, S., Walley, A. C., and Flege, J. E. (**2005**). "Lexical frequency and neighborhood density effects on the recognition of native and Spanish-accented words by native English and Spanish listeners," J. Acoust. Soc. Am. **117**, 896–907.

Ioup, G. (**1984**). "Is there a structural foreign accent? A comparison of syntactic and phonological errors in second language acquisition," Lang. Learn. **34**, 1–17.

Leather, J., and James, A. (**1991**). "The acquisition of second language speech," Stud. Second Lang. Acquis. **13**, 305–341.

Mayo, L. H., Florentine, M., and Buus, S. (**1997**). "Age of second-language acquisition and perception of speech in noise," J. Speech Lang. Hear. Res. **40**, 686–693.

Miller, G. A. (**1947**). "The masking of speech," Psychol. Bull. **44**, 105–129.

Miller, G. A., and Nicely, P. (**1955**). "An analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. **27**, 338–352.

Navarro Tomás, T. (**1918**). *Manual de Pronunciatión Española*, 21st ed. (CSIC, Madrid, **1982**).

Pallier, C., Bosch, L., and Sebastián-Gallés, N. (**1997**). "A limit on behavioural plasticity in speech production," Cognition **84**, B9–B17.

Pisoni, D. B., Lively, S. E., and Logan, J. S. (**1994**). "Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception," in *The Development of Speech Perception: The Transition From Speech Sounds to Spoken Words*, edited by J. C. Goodman and H. C. Nusbaum (MIT Press, Cambridge) pp. 121–166.

Polka, L. (**1995**). "Linguistic influences in adult perception of non-native vowel contrasts," J. Acoust. Soc. Am. **97**, 1286–1296.

Scovel, T. (**1969**). "Foreign accents, language acquisition, and cerebral dominance," Lang. Learn. **19**, 245–253.

Shannon, R. V., Jensvold, A., Padilla, M., Robert, M. E., and Wang, X. (**1999**). "Consonant recordings for speech testing," J. Acoust. Soc. Am. **106**, L71–L74.

Simpson, S., and Cooke, M. P. (**2005**). "Consonant identification in N-talker babble is a nonmonotonic function of N," J. Acoust. Soc. Am. **118**, 2775–2778.

Takata, Y., and Nábělek, A. K. (**1990**). "English consonant recognition in noise and in reverberation by Japanese and American listeners," J. Acoust. Soc. Am. **88**, 663–666.

van Wijngaarden, S. J., Bronkhorst, A. W., Houtgast, T., and Steeneken, H. J. M. (**2004**). "Using the Speech Transmission Index for predicting non-native speech intelligibility," J. Acoust. Soc. Am. **115**, 128l–1291.

van Wijngaarden, S. J., Steeneken, H. J. M., and Houtgast, T. (**2002**). "Quantifying the intelligibility of speech in noise for non-native listeners," J. Acoust. Soc. Am. **111**, 1906–1916.

PROOF COPY 060604JAS