Patterns of English phoneme confusions by native and non-native listeners

Anne Cutler[a)]
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Andrea Weber
University of the Saarland, Saarbrücken, Germany

Roel Smits and Nicole Cooper
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

(Received 15 May 2004; revised 3 September 2004; accepted 7 September 2004)

Native American English and non-native (Dutch) listeners identified either the consonant or the vowel in all possible American English CV and VC syllables. The syllables were embedded in multispeaker babble at three signal-to-noise ratios (0, 8, and 16 dB). The phoneme identification performance of the non-native listeners was less accurate than that of the native listeners. All listeners were adversely affected by noise. With these isolated syllables, initial segments were harder to identify than final segments. Crucially, the effects of language background and noise did not interact; the performance asymmetry between the native and non-native groups was not significantly different across signal-to-noise ratios. It is concluded that the frequently reported disproportionate difficulty of non-native listening under disadvantageous conditions is not due to a disproportionate increase in phoneme misidentifications. © *2004 Acoustical Society of America.*
[DOI: 10.1121/1.1810292]

PACS numbers: 43.71.Es [RLD] Pages: 3668–3678

I. INTRODUCTION

All four authors of this paper fluently speak and understand both English and Dutch; for each of us, at least one of these languages is not the native language. As non-native listeners, we are all too familiar with the phenomenon that listening to non-native language seems disproportionately difficult under disadvantageous listening conditions, such as against a noisy background.

Despite the very large literature on phoneme perception in non-native languages [see, e.g., Strange (1995) for overview papers], the evidence concerning the effects of noise and other distortions on non-native versus native perception remains relatively scant. A series of studies by Nábělek and colleagues (Nábělek and Donahue, 1984; Takata and Nábělek, 1990) demonstrated that speech stimuli which native and non-native listeners reported equally accurately in the clear were reported significantly less accurately by the non-native listeners against a noisy or reverberant background. The stimuli in question were the sentences of the Modified Rhyme Test (MRT; Kreul *et al.*, 1968), English words in the context *Say the word—again*. Gat and Keith (1978) had found the same result with similar materials. The MRT stimuli presented as synthetic speech to non-native listeners by Greene *et al.* (1985) produced a greater performance decrease compared to natural productions than the decrease observed with native listeners (see Pisoni, 1987). Van Wijngaarden *et al.* (2002) found that German and English sentences were perceived significantly better under noise by native than by Dutch listeners. Florentine (1985a, b) and Mayo *et al.* (1997), using the Speech Perception in Noise test (Kalikow *et al.*, 1977), found greater relative effects of noise on non-native than on native reports of high-predictability sentences (e.g., *The boat sailed across the bay*). Conrad (1989) reported that the greater the rate of compression applied to simple sentences (e.g., *The traveler saw a lighthouse in the distance*), the larger were the differences in listening accuracy between native and non-native listeners.

These results confirm what non-native listeners so frequently report: disadvantageous conditions affect listening to a greater degree in the non-native than in the native language. However, they do not uniquely indicate the source of this disproportionate effect. One obvious possibility is, of course, gross disruption of phonetic processing. Where the phoneme categories of the non-native language fail to match those of the native language, phonetic decisions can be influenced by the native repertoire (Best, 1995; Strange, 1995); it may be that this influence becomes stronger when stimuli are harder to perceive. Interestingly, though, a number of results suggest that the difficulty may not be (exclusively) located at the phonetic processing level. When semantically anomalous sentences (e.g., *A jaunty fork raised a vacant cow*) were presented as natural or synthetic speech to native and non-native listeners by Mack (1988), it was the native listeners who showed the proportionally greater increase in errors from the natural to the synthetic condition. Hazan and Simpson (2000) studied the effects of cue enhancement (selective amplification of the acoustic cues critical for particular contrasts) on phoneme perception in noise; their investigation revealed that all listeners' identification performance

[a)]Address for correspondence: Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. Electronic mail: anne.cutler@mpi.nl

benefited from such enhancement, but the benefit for non-native listeners was never greater than, and sometimes less than, that for native listeners. In the studies of Florentine and her colleagues (Florentine, 1985a; Mayo *et al.*, 1997), it was mainly in the effective use of contextual predictability that native listeners outstripped non-native listeners when perceiving speech in noise. Together, these studies suggest that non-native listening difficulty may have a more complex etiology than disruption of phonetic processing.

The previous literature does not, however, motivate a definitive conclusion. The aim of the present study was, therefore, to provide a new data set of phonetic identification in noise by native and non-native listeners, using materials for which higher-level factors such as lexical frequency or contextual plausibility were irrelevant, and covering almost the entire phoneme inventory of a language. The usefulness to speech perception research of very large data sets, ideally covering a complete phoneme inventory, needs no special advocacy—for American English, the classic studies of Peterson and Barney (1952), Miller and Nicely (1955), and Wang and Bilger (1973) remain valuable, now supplemented by the more recent studies of Hillenbrand *et al.* (1995) and Benkí (2003a). Smits *et al.* (2003) similarly reported perceptual data on the complete diphone set of Dutch. On the basis of such sets, it is possible to estimate the contribution of phoneme perceptibility to recognition of any spoken word of the language; our present aim was to provide such a necessary basis for understanding Dutch listeners' recognition of American English, under differing listening conditions.

The noise masking which we used was, as in the studies of Takata and Nábělek (1990), Florentine (1985a, b), and Mayo *et al.* (1997), multi-speaker babble, which best mimics difficult listening conditions in the natural experience of non-native listeners. The stimuli were CV and VC syllables covering almost all such possible sequences in American English. The native listeners were speakers of American English; the non-native listeners were Dutch. These non-native listeners were fluent users of English, but dominant in their native language. Where the phoneme categories of Dutch [16 vowels, 19 consonants (Booij, 1995; Gussenhoven, 1999)] fail to match those of American English, misidentifications are expected in the non-native responses; English contains a number of consonants with no Dutch counterpart (the final consonants of *path*, *smooth*, *edge*, and *egg*) and several vowel contrasts which collapse to a single near category in Dutch (e.g., the contrast in *bat–bet* in any variety of English, and the contrast *cot–cut* in American English). The question particularly at issue here, however, is whether under noisier listening conditions relatively more such misidentifications are observed.

We might further expect our non-native listener population to experience difficulty with syllable-final consonants, since in Dutch all consonants in syllable-final position are voiceless; English final voicing contrasts such as *at* versus *ad* should therefore prove difficult. Several recent studies of native listening in English have in fact reported better recognition of consonants presented in noise in syllable-initial than in syllable-final position (Redford and Diehl, 1999; Benkí, 2003a). Again, however, the question of principal interest here is not only whether non-native listeners experience special problems with perceiving voice in final position, but whether these problems become disproportionately more marked under increasing levels of noise.

In summary: If the repeated demonstrations of greater difficulty of non-native than of native listening under noise reflect disproportionate effects of noise on phoneme identification, then we will observe a phoneme identification difference between native and non-native listeners which increases with increasing noise, as the sentence score differences collected by Mayo *et al.* (1997) did. However, if the extra difficulties of non-native listening under noise are not exclusively, or not at all, due to problems at the level of phoneme identification, then we may observe some other pattern of results: insignificant increase in the difference between native and non-native scores, a constant difference between native and non-native scores across noise conditions, or even a decrease in the native versus non-native difference with increasing noise.

II. METHOD

A. Participants

Sixteen native listeners of American English, mostly students at the University of South Florida, participated in the experiment; they received either course credit or a small monetary compensation. Sixteen Dutch-native listeners, students at the University of Nijmegen, also participated; they received a small monetary compensation. In all cases listeners were rewarded per session and additionally with a bonus upon completion of the eight-session experiment. The Dutch native listeners had all completed their school education in the Netherlands, involving 7 to 8 years of English instruction beginning on average at age 11. All were fluent in English but none had lived for longer periods in an English-speaking country.

B. Materials

Twenty-four consonants and 15 vowels were combined to form all possible standard American English CV and VC sequences, excluding those with schwa. All vowels (12 monophthongs and three diphthongs) occurred either in initial or in final position; thus lax vowels were allowed in syllable-final position although stand-alone syllables ending in lax vowels do not occur in the language. Twenty-two consonants (not /ŋ/ or /ʒ/) occurred in initial position, and 21 consonants (not /h/, /w/, /j/) in final position. The full phoneme set can be found in Appendix A. The complete set of stimuli comprised 645 (330 CV and 315 VC) syllables.

The 645 syllables were transcribed phonemically. A phonetically trained female native speaker of American English (born in the Mid West, who had lived as child and teenager in four different states) read these transcriptions in a quiet room via a Sennheiser ME64 microphone to Digital Audio Tape. The sampling rate at recording was 44.1 kHz, later downsampled to 16 kHz. Stop consonants in final position were released.

Each syllable was centrally embedded in 1 s of multi-speaker babble noise. The babble was constructed from a

recording of three male and three female speakers having a conversation in English in a quiet room. The recording was made directly onto DAT tape using a Sennheiser microphone placed in the middle of a table around which the speakers were seated. For each speaker, a 1-s stretch was selected during which no background noises were present and he or she was speaking alone at a normal (i.e., not too loud or soft or excited) tone. These six stretches of speech were then equalized for rms amplitude and added together. The test syllables were normalized for rms amplitude of the vowel and were then combined with the babble noise at three signal-to-noise ratios (SNRs): 0, 8, and 16 dB (normalized vowel amplitude/babble amplitude). These SNRs were chosen on the basis of a pretest to yield difficult, intermediate, and easy English phoneme perception for the Dutch non-native listeners. The whole stimulus set thus comprised 1935 tokens (645 syllables×3 SNRs).

C. Procedure

Each listener participated in eight testing sessions, made up of 3870 trials in total. Each of the 1935 tokens was presented twice (always in separate sessions), once with the listener's task being to identify the consonant and once to identify the vowel. In each session, listeners received two stimulus blocks, one for consonant and one for vowel identification; the blocks consisted solely of CV or solely of VC stimuli. Every listener received the items of a block in a different pseudo-random order. SNRs were mixed within blocks.

The presentation of items was self-paced. If the listener did not respond by 15 s after stimulus offset, the trial was recorded as a miss. Listeners signaled responses by clicking on a word exemplifying the appropriate sound on a computer screen. Prior to the experiment they were familiarized with these example words: they saw the display screen and heard the same speaker as in the experiment pronounce each alternative, e.g., *v as in very*. Different words were used for vowels, initial consonants, and final consonants; the words are listed in Appendix A.

III. RESULTS

No response (''miss'') was registered on in total 0.64% of trials (less than 0.1% for the non-native listeners; just over 1% for the native listeners); these trials were discarded from the data set.

The principal findings of our study can be derived at a glance from Fig. 1, which shows the overall proportions of correct responses for the two listener groups for vowels and for consonants at the three signal-to-noise ratios. First, it can be seen that the identification performance of the non-native listeners is significantly and consistently worse than that of the native listeners, but that this performance disadvantage is unaffected by signal-to-noise ratio. Second, it can be seen that an increase in signal-to-noise ratio results in a clear improvement in performance (for both listener groups) in the identification of consonants, but has very little effect on the identification of vowels, which (again for both listener groups) is at a relatively high level even at 0 dB SNR. The

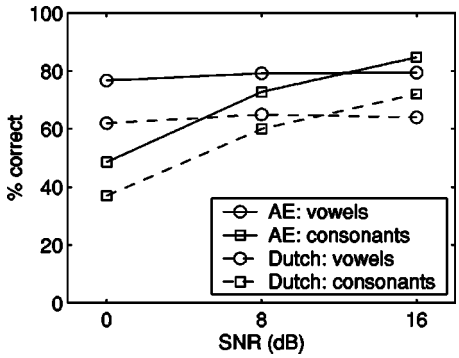

FIG. 1. Percentages of correctly recognized vowels and consonants as a function of SNR, separately by language group (''AE''=American English). Data have been pooled across initial and final positions, phonetic contexts, and subjects.

pattern of results thus strongly suggests that the greater difficulty of non-native than of native listening under noise is not due to disproportionate effects of noise on phoneme identification.

Analyses of variance across subjects confirmed that overall performance was better for native (grand mean 73.6% correct) than for non-native listeners (grand mean 60%; $F[1,30]=21.1$, $p<0.001$) and for vowels (grand mean 71.1%) than for consonants (62.5%; $F[1,30]=44.66$, $p<0.001$); further, performance was strongly affected by SNR (grand mean of 75% at 16 dB SNR, 69.3% at 8 dB SNR, and 56.1% at 0 dB SNR; $F[2,60]=2191$, $p<0.001$). The latter two effects interacted significantly—consonant performance showed significantly more effect of SNR than did vowel performance ($F[2,60]=2066$, $p<0.001$)—but the native versus non-native comparison interacted neither with SNR nor with the vowel/consonant factor. *Posthoc* analyses revealed that for both listener groups there was a significant difference in performance between 0 and 8 dB SNR for both vowels and consonants (vowel difference 2% for each group, consonant difference 24% for native and 23% for non-native), but a significant difference between 8 and 16 dB SNR for consonants only (vowel difference 0% for native and 1% for non-native, consonant difference 12% for each group).

Figure 2, which presents overall identification in initial versus final position in the carrier syllable, shows a further clear effect: for native listeners, identification of both consonants and vowels is better in final than in initial position. Non-native listeners show the same final-position advantage

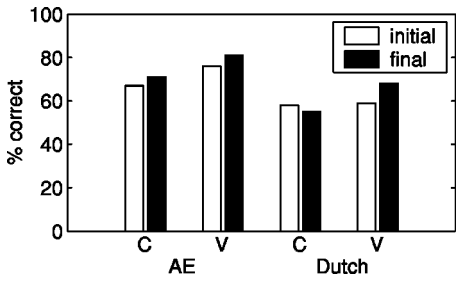

FIG. 2. Percentages of correctly recognized consonants (C) and vowels (V) in initial and final positions, separately by language group (''AE''=American English). Data have been pooled across SNRs, phonetic contexts, and subjects.

TABLE I. Confusion matrix for initial consonants at 0 dB SNR categorized by the American English listeners. Percentages of correct responses have been pooled over participants and vowel contexts.

	Response																				
Stimulus	pie p	tie t	car k	far f	thin θ	see s	she ʃ	chin tʃ	hi h	be b	do d	go g	very v	there ð	zoo z	joke dʒ	yell j	my m	no n	lie l	row r	win w
p	**15.4**	2.9	4.6	10.4	3.3				39.2	6.7	1.3	2.1	2.5	1.3		0.8	1.3	1.3	0.4	0.4	0.8	0.8
t	10.4	**19.6**	9.6	5.4	6.3	1.3	0.8	0.4	27.9	2.9	1.3	3.3	0.4	3.3	1.3	0.4	0.8	0.4	0.8	0.4		0.8
k	11.7	14.6	**25.8**	1.7	2.1		0.4	0.8	27.9	2.1	0.8	2.1	0.4	0.8			0.4	0.4	2.1			1.7
f	22.9	2.1	3.8	**19.2**	7.5	0.4			14.2	8.8	1.3	0.8	3.8	5.8		0.4	1.7					3.8
θ	12.5	5.4	3.8	13.3	**18.3**			0.4	10.4	7.5	2.1	1.3	3.8	14.6	0.8		0.4	0.8	0.4	0.4	0.4	1.7
s	0.4	2.1	0.4	9.2	10.0	**51.7**	2.1	0.4	2.1	2.5				9.6	8.8							
ʃ	0.4		0.4			0.8	**76.7**	19.6				0.4	0.4	0.4		0.4						
tʃ		5.0	0.8	1.7	0.8		1.3	**83.8**	0.4			0.4		1.7		2.9	0.4		0.4			
h	14.6	5.0	4.6	9.6	4.6	0.4		0.4	**36.7**	7.1	0.4	1.7	2.9	2.1		0.4		1.7	0.4	0.4	0.4	1.7
b	2.1		1.3	5.8	5.8				15.0	**19.6**	1.3	1.3	5.0	8.3	0.4	0.8	3.8	10.8	0.4	4.2	1.7	3.8
d		2.9		1.3	7.9	0.4		0.8	4.6	7.9	**14.6**	2.9	0.8	14.6	0.4	0.8	7.1	3.3	19.6	6.7		0.4
g	1.7	0.8	1.3	2.9	2.5	0.4			10.4	3.8	2.5	**29.6**	4.2	2.1	0.4	0.8	19.2	1.3	7.9	2.9	1.3	1.3
v	2.9	1.7	0.4	5.8	4.2	0.8		0.4	8.8	18.3	0.4	3.3	**17.5**	14.6		0.8	2.1	4.2	1.7	0.8	1.7	5.0
ð		1.3		1.3	14.6	0.8		0.4	1.7	9.6	4.2	3.3	5.8	**30.4**	4.2	2.1	1.3	1.3	4.6	10.0		0.8
z		1.7			9.2	2.5			0.8	1.7	1.7	2.5	8.3	21.3	**31.3**	2.1	0.4	0.8	3.3	1.3	1.7	7.5
dʒ	0.4	0.4	0.4	0.4	2.5	0.4		4.6	1.3	1.7	4.2	2.9	0.4	8.3		**68.8**	0.8	0.4	0.4	1.3		
j		0.8			0.4				2.9	3.3	5.4	3.3	1.3	1.3	1.7	2.9	**65.8**	2.5	2.5	1.7		2.1
m	0.4		0.4	1.7	0.4				3.3	3.8	0.8	1.3	6.3	0.4	0.4		0.8	**63.8**	5.8	5.8	1.3	1.7
n					0.8				0.4	0.4	0.4		0.4	0.4	0.4		0.4	12.5	**77.9**	4.2	0.8	0.4
l	0.4				2.5			0.4	0.8	5.0	1.7	2.1	3.3	4.2	0.4		2.1	12.5	5.0	**54.2**	2.1	0.8
r	0.8	0.4	1.3	1.3					7.1	5.4	0.4	2.1	5.0	0.8				2.5	0.4		**68.8**	2.9
w	0.8	0.4							1.7	4.2		0.8	2.9	0.4			4.2	5.8		2.9	0.4	**73.3**

for vowels; for consonants, however, their performance is worse in VC than in CV. Analyses of variance confirmed that the overall advantage for final position was significant ($F[1,30]=27.34$, $p<0.001$), but this effect interacted with the vowel/consonant factor ($F[1,30]=20.4$, $p<0.001$), the final advantage being larger for vowels than for consonants. Moreover, the three-way interaction of these two factors with listener language was also significant ($F[1,30]=11.44$, $p<0.002$); *posthoc* tests revealed the source of this interaction to be a significant interaction (reversal of the position effect) of initial/final and vowel/consonant for non-native listeners ($F[1,15]=26.91$, $p<0.001$), but no significant interaction for native listeners ($F<1$).

Since the overall advantage for final position which we observed contrasts with previous findings of Redford and Diehl (1999) and Benkí (2003a), we conducted further analyses in direct comparison with these earlier studies; these analyses are described in Appendix B.

Detailed results are presented as confusion matrices (separately for native and non-native listeners and for consonants and vowels in initial versus final position) in Tables I–VIII. These tables show the identification results at 0 dB (the more accurate results at better SNRs are available at http://www.mpi.nl/world/persons/private/anne/materials.html). Where rows do not sum to 100%, the remainder was missing data.

It can readily be seen from the confusion matrices that the phonemes which were most difficult for non-native listeners were also difficult for native listeners. Thus although the Dutch listeners had difficulty identifying the English consonants without counterpart in Dutch, these consonants were also difficult for the native listeners; and although the Dutch listeners confused the vowel sounds which share one near Dutch category, native listeners made such confusions, too. Characteristics of the masking babble noise presumably influence these patterns. In fact, the percent correct identification rate of the two listener groups across phonemes was very highly correlated: at 16 dB SNR $r=0.83$, at 8 dB 0.87 and at 0 dB 0.91, in all cases $p<0.001$. It can also be seen from the matrices that there were no strong effects of phonotactic legality of syllable-final lax vowels; errors on these vowels, for both listener groups, tended to be other lax vowels. The one clear effect of native phonology on non-native listening appeared in the Dutch listeners responses to syllable-final consonants; as described above, Dutch phonotactics prohibit voicing contrasts in final position, and the Dutch listeners made many more voicing errors on final consonants such as /b, d, g/ than native listeners did.

Since Miller and Nicely (1955), it has been customary to view perceptual data of the present type in terms of percentage of information transmitted for broad feature classes. In contrast to raw percent correct, transmitted information (TI) takes account of response biases, and, regardless of the number of response alternatives, gives a result of zero when subjects guess randomly. The number of response alternatives varies across features, so only TI measures allow direct comparisons of the accuracy with which different features can be recognized. Thus only TI allows us to compare native and non-native featural sensitivity. Smits (2000) further explains these advantages, and equations for TI calculation are presented by Miller and Nicely (1955).

Figure 3 presents TI analyses of our data set, and Table IX shows the phonemes associated with the featural values we used. We considered the broad features of consonants to be place and manner of articulation, and voicing, rather than the more detailed feature systems (coronal, anterior, conso-

TABLE II. Confusion matrix for final consonants at 0 dB SNR categorized by the American English listeners. Percentages of correct responses have been pooled over participants and vowel contexts.

	Response																				
Stimulus	lip p	hot t	sick k	off f	path θ	pass s	fish ʃ	such tʃ	grab b	odd d	egg g	love v	smooth ð	buzz z	beige ʒ	edge dʒ	am m	on n	ring ŋ	ill l	far r
p	**50.0**	16.3	14.2	5.8	5.8				0.8	0.4	0.4		2.1				0.4				0.8
t	5.0	**77.1**	5.8	0.8	4.6			2.1		0.4		0.4	2.5	0.4							
k	11.3	12.5	**63.3**	0.8	5.0			2.1	0.4	0.4	0.4	0.8	2.1								
f	10.0	10.0	6.7	**45.0**	12.9		0.8		0.8			1.3	5.4		0.8		0.4			0.4	1.7
θ	9.2	17.9	4.2	30.8	**19.2**	0.8	0.4	0.8	0.4	0.4	0.4	2.5	7.5		0.4			0.4	0.8		0.4
s	0.8	2.9	0.8	12.9	8.8	**65.4**	2.9	0.4			0.4		1.7	0.8	0.4						
ʃ					0.4	1.3	**80.8**	14.2					0.4		2.5	0.4					
tʃ	0.4	3.8	0.4					**89.6**							1.7	4.2					
b	1.3	1.3	4.2	3.8	2.5	0.4		0.4	**35.0**	10.4	9.2	15.4	4.6	0.4	2.5	1.3	2.1	1.7		0.4	1.3
d		3.3	0.4	3.8	2.5	1.7			3.8	**42.9**	4.6	6.7	5.8	1.7	5.4	5.8	0.4	5.8	2.9		0.4
g	0.4	3.3	1.3	2.9	5.4	0.4		0.8	5.4	9.2	**35.4**	14.2	5.4	0.8	2.1	1.7	1.3	2.1	2.5	0.4	0.8
v	0.4	0.8	1.3	9.2	2.5			0.4	2.9	4.6	7.9	**47.5**	5.8	0.4	3.8	1.7	2.9	1.3	0.8	0.8	1.7
ð		2.1		1.7	4.2	2.5	0.4	0.4	2.9	22.5	5.0	17.5	**16.7**	5.8	5.4	7.5		1.3	0.8	0.4	0.4
z	0.4				2.5	7.5	0.4	0.4	1.3	10.0	1.3	12.5	9.6	**37.1**	5.4	4.6	0.4	2.5		0.4	0.8
ʒ				0.4		0.8	2.5	2.1		2.5	1.3	4.2	4.6	3.8	**51.7**	23.3	0.4	1.7		0.4	0.4
dʒ	0.4	0.8			0.4		0.4	3.3	0.4	5.8	2.5	0.8	0.4	0.4	17.9	**64.6**	0.4				0.8
m		0.4		1.3	0.4				1.7	1.3	2.1	7.1	0.4			0.4	**56.3**	12.1	14.2	0.4	0.8
n					0.4				0.4	5.4	1.3	3.3	1.7	1.3	1.3	0.8	12.5	**59.6**	10.4		0.4
ŋ		0.4	0.4	0.8				0.4	0.4	1.3	9.2	5.8	1.3		0.4		15.4	25.4	**35.0**	0.8	1.7
l	0.4	0.8	0.8	7.9	0.4					0.8	0.8	6.7	3.3		0.4	0.4	1.3	0.4		**70.8**	0.8
r		0.4	0.4	1.3	0.8				0.4	0.8	0.4	3.8	1.3			1.3	1.3	0.4		0.8	**84.2**

nantal, sonorant, continuant, etc.) used in formal phonology (e.g., Kenstowicz, 1994). The values of place of articulation were held to be labial, dental, alveolar, palatal, velar, and glottal; this classification strikes a balance between a very detailed phonetic inventory of places within the English consonant inventory, which would have very few consonants at many places, and a gross classification into only labial, coronal, and dorsal. As values of manner of articulation we used stop, affricate, fricative, liquid, glide, and nasal. Voicing had two values, voiced and voiceless. The features used for vowels were height (three values: high, mid, and low), backness (three values: front, central, and back) and tenseness (two values: tense and lax). Because the three diphthongs always change value on height and tenseness, and two of the three also change value on backness, we excluded them from the vowel calculations in Fig. 3; for the TI calculations it was therefore also necessary to discard diphthong responses to monophthongal stimuli, a total of 915 cases (1.85% of the total monophthongal vowel dataset).

Statistical analyses of the comparisons in Fig. 3 showed a significant improvement in percentage of transmitted featural information with increasing SNR, for five of the six broad feature classes (all comparisons $p<0.001$; vowel tenseness insignificant). For all three vowel features, and for consonant manner, information was transmitted more efficiently in final position in the syllable than in initial position

TABLE III. Confusion matrix for initial vowels at 0 dB SNR categorized by the American English listeners. Percentages of correct responses have been pooled over participants and consonant contexts.

	Response														
Stimulus	beat i	bit ɪ	wait eɪ	bet ɛ	bat æ	hot ɑ	cut ʌ	caught ɔ	boat oʊ	cook ʊ	boot u	buy aɪ	boy ɔɪ	shout aʊ	bird ɝ
i	**78.9**	8.3	0.3	2.7				0.3		0.3	1.8	1.2			3.9
ɪ	1.5	**81.8**	0.9	8.0	0.9		1.2	0.3		0.3	1.5	0.9			1.8
eɪ	5.7	5.4	**74.4**	4.5	5.7				0.3	0.3		0.3		0.3	1.2
ɛ	0.6	4.2	2.4	**84.2**	2.7		1.2					0.3			3.0
æ		1.2	6.5	3.9	**78.3**	0.6		1.2				0.3	0.3	4.8	2.1
ɑ		0.6	1.2	0.3	9.8	**42.3**	12.5	26.8	0.9			0.6		0.9	1.8
ʌ			0.3		1.2	12.5	**64.9**	8.3	1.8	1.2		1.2		4.2	3.0
ɔ		0.3	0.3		0.6	36.3	4.5	**47.3**	3.9	1.2		0.6	2.1	0.9	
oʊ		0.3				4.8	1.2	0.9	**69.6**	8.9	6.5		3.3	2.1	0.3
ʊ						2.1	14.0	2.1	0.9	**63.7**	6.8	0.3	3.0	2.1	0.9
u	3.6	1.5	0.3			0.6	3.0	1.5	1.8	19.3	**62.5**	0.3	1.2	1.8	1.2
aɪ		8.3	2.1				0.3					**87.2**	0.6		0.6
ɔɪ	0.3		0.6	0.3			0.3	0.6	1.2	0.6	0.3		**92.9**	3.0	
aʊ	0.3		0.3	2.1	0.6	3.3		7.1	2.4	0.3	0.3		0.9	**81.5**	0.3
ɝ	0.6	0.3		1.5			1.5	0.3							**95.5**

TABLE IV. Confusion matrix for final vowels at 0 dB SNR categorized by the American English listeners. Percentages of correct responses have been pooled over participants and consonant contexts.

	Response														
Stimulus	beat i	bit ɪ	wait eɪ	bet ɛ	bat æ	hot ɑ	cut ʌ	caught ɔ	boat oʊ	cook ʊ	boot u	buy aɪ	boy ɔɪ	shout aʊ	bird ɝ
i	**93.5**	0.3		3.7				0.3			1.4			0.3	
ɪ	0.9	**84.4**	0.3	10.5	0.3		2.0		0.3			0.9		0.3	
eɪ	0.6	2.0	**91.5**	2.0	2.8										
ɛ	0.6	6.3	2.3	**73.6**	8.5		2.3	2.3		0.3		0.3		0.3	0.3
æ		0.6	1.1	12.2	**82.7**			1.1			0.3	0.3			0.3
ɑ			1.1	0.9	8.2	**33.5**	24.4	27.0	0.6	0.3		0.3		1.1	
ʌ			0.9	2.3	6.0	11.4	**65.3**	11.1	0.3	0.9	0.3		0.3	0.3	
ɔ			0.9		2.6	23.9	3.7	**65.3**	0.9	0.6				0.9	
oʊ	0.3	0.3		0.3		1.7		0.6	**90.6**	0.3	0.9		1.4	2.3	
ʊ	0.3			0.6		2.0	21.6	0.6	0.6	**68.2**	2.6	0.3	0.3		
u	3.7			0.9		0.3	0.9	0.6	0.3	6.8	**81.8**		0.3	2.6	
aɪ		6.0	0.9	0.3		0.3		0.3	0.3			**91.5**		0.3	
ɔɪ						0.6	0.3	0.9	1.4	0.9	0.3	0.3	**92.0**	2.3	
aʊ	0.3			0.9	0.6	2.0		1.7	8.2	0.6			1.7	**82.4**	0.6
ɝ	0.3	0.3		1.1			0.3								**97.7**

(three comparisons $p<0.001$, vowel backness $p<0.05$). Place and voicing showed no significant main effect of position in the syllable. There was an interaction between SNR and position within the syllable for all feature classes, reflecting in each case a greater improvement with increasing SNR for phonemes in syllable-initial position than for phonemes in syllable-final position (five comparisons $p<0.001$, consonant manner $p<0.05$). For all types of phonetic information, the masking effects of noise (especially at 0 dB SNR) are thus greatest in syllable-initial position. Listener group language did not interact with SNR for any feature comparison, but interacted with syllable position for consonant voicing ($p<0.001$) and vowel height ($p<0.05$). For the native listeners, voicing information was perceived better in final position, but for the Dutch listeners, as expected, voice was much less well perceived in final position [Fig. 3(e)]. The vowel height interaction was due to the advantage of final position over initial position being larger for non-native than for native listeners [Fig. 3(b)].

For each listener group separately, we compared the relative informativeness of types of featural information. There were significant differences in informativeness among the consonant features for the native ($F[2,30]=16.94$, $p<0.001$) and non-native listeners ($F[2,30]=57.69$, p

TABLE V. Confusion matrix for initial consonants at 0 dB SNR categorized by the Dutch listeners. Percentages of correct responses have been pooled over participants and vowel contexts.

	Response																					
Stimulus	pie p	tie t	car k	far f	thin θ	see s	she ʃ	chin tʃ	hi h	be b	do d	go g	very v	there ð	zoo z	joke dʒ	yell j	my m	no n	lie l	row r	win w
p	**30.8**	3.3	9.2	9.6	2.9			0.4	19.2	11.7	1.3	1.3	2.9	1.7	0.4		0.8	0.8	1.3	1.3		1.3
t	24.6	**14.2**	12.5	7.5	7.9	0.8		2.9	11.3	7.1	0.4	2.1	1.3	1.7				2.1	3.3	0.4		
k	25.0	7.9	**25.8**	3.8	4.2	0.4	0.8	0.4	13.8	4.2	1.3	3.8	1.3	0.8	0.4	0.4	1.3	0.4	1.7	1.3	0.4	0.8
f	24.6	2.1	9.2	**15.0**	7.1		0.4	0.4	9.2	15.0	1.7	2.9	5.4	4.2		0.4	0.4	0.4	0.4		0.4	0.8
θ	18.8	6.3	3.8	13.3	**12.1**	0.4	0.4	0.4	7.1	14.2	2.5	1.7	2.9	7.5			0.4	1.3	2.9	2.9		0.8
s	0.4	2.5	0.4	12.5	24.6	**30.4**	0.8	0.4		0.8	1.3		3.3	7.9	14.6							
ʃ		0.4			1.3	6.7	**72.5**	18.3									0.8					
tʃ	3.3	4.2	1.3	2.1	2.5	1.3	4.6	**70.8**	1.3	1.3	0.4		0.4	0.8		5.4	0.4					
h	26.3	4.6	12.1	11.3	5.0	0.4	0.4	0.8	**17.9**	8.3	1.3	0.4	4.6	1.7	0.4		0.8		0.8	1.7	0.8	0.4
b	7.5	0.4	5.8	9.2	1.7	0.4		0.4	12.5	**28.3**	2.5	0.4	4.6	2.1			2.9	7.1	2.9	5.0	1.3	5.0
d	2.5	2.1	1.3	1.3	5.4				8.8	12.1	**10.8**	2.5	2.1	12.9	0.4	0.4	6.3	4.2	12.5	12.5		1.7
g	3.3	1.3	9.2	2.9	2.5		0.4	1.3	9.2	10.0	5.0	**17.1**	1.7	3.3		0.8	24.2	0.8	2.5	2.5	0.4	1.7
v	7.5	2.9	2.5	8.8	5.0	0.4			6.7	30.0	1.3	1.7	**9.6**	7.9			2.1	3.8	1.7	0.4	2.5	5.4
ð	2.5	1.3	2.5	1.7	14.6	2.1	0.4	0.8	2.1	17.1	10.0		1.3	**18.8**	1.3	1.7	1.3	1.3	3.3	12.1	0.8	2.9
z		0.8	1.3	1.3	9.6	3.3	0.4	1.3		7.9	5.0		2.5	23.8	**27.1**	1.3	2.5	2.1	5.0	0.4	0.8	3.8
dʒ	4.2	0.4	2.5	0.4	2.1		0.8	18.3	2.1	2.1	6.7	2.1		7.5		**40.4**	5.8	0.4	1.3	2.9		
j	1.3		0.8	0.8	0.8		0.8	0.4	2.1	4.2	2.5	1.3	0.4	1.3	0.4	4.6	**69.6**	2.5	4.2	1.3		0.8
m	3.8	0.8	2.5	2.9	0.4				2.1	9.6	0.8		2.5					**50.0**	10.0	5.0	5.4	4.2
n					0.4				2.1	1.7	1.3					0.8	0.4	12.9	**73.8**	4.6	0.4	1.7
l	5.8	1.3	1.7	1.3	1.3			0.4	2.1	8.3	1.7	0.8	2.5	3.3			1.3	10.0	4.2	**46.7**	2.1	5.4
r	2.5	1.3	1.7	2.1				0.4	5.8	14.6	0.8	2.1	1.3	0.8				0.8	0.4	0.4	**58.3**	6.7
w	1.7		0.4	0.4	0.4	0.4			1.7	5.8	0.8		2.1	0.4				5.8	0.8	2.5	1.7	**75.0**

TABLE VI. Confusion matrix for final consonants at 0 dB SNR categorized by the Dutch listeners. Percentages of correct responses have been pooled over participants and vowel contexts.

	Response																				
Stimulus	lip p	hot t	sick k	off f	path θ	pass s	fish ʃ	such tʃ	grab b	odd d	egg g	love v	smooth ð	buzz z	beige ʒ	Edge dʒ	am m	on n	ring ŋ	ill l	far r
p	**24.2**	13.8	11.7	5.8	8.8			0.4	21.3	5.8	2.1		3.3		0.8		0.4	0.4		0.4	0.8
t	4.6	**45.0**	5.4	1.7	9.2	1.7		1.3	1.7	20.4	0.4	0.4	5.4	0.4	0.4	0.8				0.4	0.8
k	8.3	12.5	**44.6**	4.2	5.0	0.4	0.4	0.8	2.5	3.8	12.5	0.4	1.7			1.3		0.4			1.3
f	7.5	21.7	6.7	**22.1**	9.6	1.7	1.3	0.4	6.3	10.4	0.4	3.3	5.0	0.4	0.8	0.4	0.8			0.8	0.4
θ	7.9	24.6	3.8	17.9	**17.5**	0.8	0.4	1.7	2.9	7.9	0.4	2.5	9.2		0.8	1.3					0.4
s		3.8		17.1	14.6	**37.5**	5.0	0.8	0.4	0.4		2.5	10.0	4.6	1.7	0.8					0.8
ʃ					0.4	6.7	**66.7**	10.4			0.4		1.3	1.7	10.0	2.1					0.4
tʃ	0.4	0.8		0.4	1.3	0.4	6.7	**42.5**	0.8	3.3	0.4		3.3		5.4	34.2					
b	5.0	7.5	6.7	2.9	5.4			0.4	**30.4**	15.0	3.8	7.9	5.0	1.3	0.4	2.9	2.5	0.8		0.8	1.3
d	1.3	16.3	0.4	2.5	5.8	0.8	0.4		2.1	**39.6**	2.1	3.8	7.9	1.7	2.9	5.8	0.4	1.7	2.5		2.1
g	0.8	12.9	4.6	0.8	7.5			0.8	2.1	20.4	**25.8**	4.2	5.0	0.8	0.8	6.3	2.5	1.7	0.8	0.8	1.3
v	1.3	12.9	3.8	12.1	5.4		1.7	0.4	4.6	15.8	3.3	**15.8**	5.4	1.3	1.3	1.3	2.1	3.3	1.3	4.2	2.9
ð	0.4	11.3	0.8	4.2	8.3	3.3	1.7	1.3	2.5	29.2	2.1	10.0	**8.3**	2.9	1.3	7.9	0.4	1.3		0.8	2.1
z		3.8		1.7	10.0	12.1	2.1	2.5	0.4	8.8	2.1	5.0	10.0	**25.8**	5.0	2.1	2.1	1.7	1.3	1.3	2.5
ʒ		3.3	0.4		2.1	2.5	14.2	4.2	0.4	3.8	0.4	1.3	4.2	6.7	**45.0**	9.2	0.4	0.4		0.8	0.8
dʒ		2.9		1.7	2.9	0.8	2.9	13.3	0.4	8.8			4.2	0.4	9.6	**51.7**					0.4
m		9.2		0.4	4.6	0.8				5.8	1.3	2.1	2.1	0.8			**41.3**	20.0	8.3	2.1	1.3
n	0.4	9.6		1.3	1.3	0.4	0.4	0.4	0.8	8.3		0.4	2.1	2.1	0.4	2.1	8.3	**48.3**	9.2	2.1	2.1
ŋ		6.7	0.4	2.1	2.1	0.4			2.5	5.8	6.3		1.7	1.7			13.8	22.5	**30.4**	2.1	1.7
l	0.4	8.3	1.7	5.8	3.3		0.4			9.2	1.3	2.1	1.7	0.4			0.8	2.5	0.4	**57.5**	4.2
r	0.4	7.9	0.4	0.8	3.8		0.4		1.7	6.7		1.7	2.9	0.4	0.4	0.8		2.1		1.7	**67.9**

<0.001), and among the vowel features also for both the native ($F[2,30]=120.23$, $p<0.001$) and non-native groups ($F[2,30]=186.61$, $p<0.001$). Interfeatural comparisons showed that for native listeners consonantal manner information was transmitted most accurately and place information least accurately; manner and voicing did not differ significantly but each was significantly more accurately perceived than place ($p<0.001$ for manner, $p<0.01$ for voicing). For non-native listeners manner was also transmitted most accurately (significantly more so than place, $p<0.001$), but voicing least accurately (significantly less so than place, $p<0.001$). For both groups the vowel features ordered similarly: backness was more accurately transmitted than height, and height was more accurately transmitted than tenseness (all comparisons $p<0.001$).

These analyses thus further confirm the parallel effects of the noise masking on the performance of the native and non-native listener groups. Although the groups differed overall in one respect, namely in sensitivity to final voicing contrasts, importantly, on no type of information at all did listener group language interact with SNR.

IV. GENERAL DISCUSSION

The identification performance of non-native listeners in our study fell clearly short of the native listeners' perfor-

TABLE VII. Confusion matrix for initial vowels at 0 dB SNR categorized by the Dutch listeners. Percentages of correct responses have been pooled over participants and consonant contexts.

	Response														
Stimulus	beat i	bit ɪ	wait eɪ	bet ɛ	bat æ	hot ɑ	cut ʌ	caught ɔ	boat oʊ	cook ʊ	boot u	buy aɪ	boy ɔɪ	shout aʊ	bird ɝ
i	**75.6**	16.7	1.8	1.8		0.3	0.6			1.5			0.3	0.3	0.9
ɪ	1.5	**86.0**	0.6	5.4	0.3		0.3		0.3	0.3	1.5	1.2			2.4
eɪ	25.0	14.6	**46.7**	6.5	4.2		0.3	0.6	0.3				0.6		1.2
ɛ	0.3	12.5	0.9	**58.3**	25.0		0.3							0.3	2.1
æ		1.8	1.2	33.6	**56.3**	0.3	0.6					0.9		4.2	1.2
ɑ	0.3			0.6	13.4	**29.2**	31.5	15.5	1.2	0.3	0.3	3.9	0.3	1.2	1.8
ʌ			0.3	0.9	4.8	27.4	**44.3**	7.7	3.9	0.3	0.6	1.5	0.3	3.9	4.2
ɔ					0.6	63.4	5.4	**22.0**	2.4	0.6	0.9	0.6	3.0		1.2
oʊ			0.6			8.6	0.9	2.1	**53.0**	14.9	14.9	0.9	1.8	1.2	0.6
ʊ	0.6	0.3			0.3	11.6	4.2	2.4	3.9	**50.9**	17.3	0.6	3.6	2.4	1.5
u	8.6	1.5	0.3		0.3	2.7	0.9	1.8	4.8	46.7	**29.2**	0.3		1.8	1.2
aɪ		3.0	24.7	0.6	2.7		0.3	0.6	0.6			**64.3**	0.9		2.4
ɔɪ			1.5			3.6		0.3	0.9	0.6	0.3	0.9	**90.5**	1.2	0.3
aʊ			0.3	2.7	0.6	2.7		2.7	16.4	0.9	0.9	2.4	0.3	**70.2**	
ɝ	0.9	0.3	0.3	0.9		0.3	19.0	0.9	0.3			0.6		0.6	**75.9**

TABLE VIII. Confusion matrix for final vowels at 0 dB SNR categorized by the Dutch listeners. Percentages of correct responses have been pooled over participants and consonant contexts.

	Response														
Stimulus	beat i	bit ɪ	wait eɪ	bet ɛ	bat æ	hot ɑ	cut ʌ	caught ɔ	boat oʊ	cook ʊ	boot u	buy aɪ	boy ɔɪ	shout aʊ	bird ɚ
i	**97.4**	0.3	0.3	1.7								0.3			
ɪ	0.3	**95.5**		2.3		0.3								0.3	1.4
eɪ	6.0	4.3	**84.1**	3.1	1.1			0.3			0.3	0.9			
ɛ	0.3	15.6	0.3	**60.5**	22.2			0.3						0.3	0.6
æ		0.9	4.8	39.2	**51.7**	0.6	1.1		0.3			0.6			0.9
ɑ	0.3	0.6		0.6	16.8	**23.6**	22.7	28.4	2.6	0.3	0.3	1.7	0.3	2.0	
ʌ		0.3	0.6	1.7	10.2	25.0	**41.5**	16.8	1.4	0.3	0.3	1.4		0.6	
ɔ					4.8	34.7	5.4	**50.0**	3.4		0.3		0.3	0.6	0.3
oʊ		0.6			0.3	11.1	0.6	4.8	**69.6**	2.8	6.0		2.6	1.7	
ʊ		0.6				5.4	4.3	4.5	1.7	**73.3**	6.8	0.6	1.4	0.6	0.9
u	17.9	0.3		0.6		0.6	1.4	1.1	0.9	31.3	**45.2**	0.3		0.6	
aɪ		3.4	14.8	0.3	1.1	0.3	0.6	0.6	0.9			**74.1**	1.4	0.3	2.3
ɔɪ			0.3			2.3		0.6	0.9	0.3		1.7	**93.8**	0.3	
aʊ		0.3	0.3	2.0	0.6	4.0	0.6	5.1	17.0	0.3	0.9	3.1	0.9	**64.5**	0.6
ɚ	0.3	0.3		0.9	0.3	0.6	7.1	0.3				0.6		0.3	**89.5**

mance. Also, all listeners were adversely affected by noise—the higher the SNR, the better the performance, for both native and non-native listeners. Crucially, however, the effects of language background and of noise did not interact. Of the possible patterns of results we listed in the Introduction, the one we have observed is a constant disadvantage for non-native compared with native listeners, irrespective of the degree of noise-masking. Thus our study clearly suggests that the disproportionate effects of noise on listening to non-native, as opposed to native, language are not due to exacerbation of the difficulty of phoneme identification.

We did observe effects of the native inventory on the non-native listeners' identifications. Where the non-native listeners performed as well as the native listeners (for instance, in the vowel confusion matrices, for the vowels of *beat*, *bit*, and *boy*), it was for phonemes which occupy highly similar positions in the two inventories (Gussenhoven, 1999; Ladefoged, 1999). Where the inventories mismatched, non-native performance often fell behind. However, these native-inventory effects were not heightened under noise. If anything, the similarity of non-native and native performance became stronger under noise, as our correlation analyses showed.

Phonological constraints of the native language also affected performance; the absence of a voicing contrast in final position in Dutch was reflected in the non-native listeners' poor performance on voicing decisions in final position [Fig. 3(e)], which led to a reversal, for non-native consonant identifications only, of the otherwise constant advantage of final over initial position with our stimuli (Fig. 2). Again, however, this effect did not significantly interact with the effects of SNR; if anything, the advantage of initial over final voicing decisions for Dutch listeners was actually greater at 16 dB than under more severe noise.

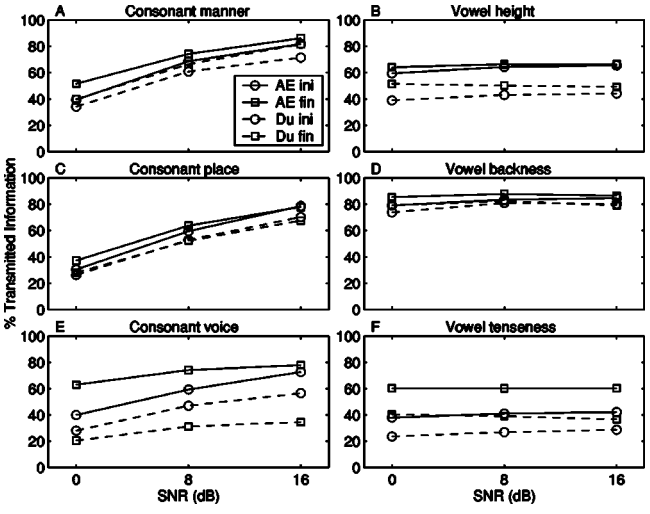

FIG. 3. Percentages of transmitted information for six phonological features as a function of SNR. Data are presented separately by position ("ini" =initial, "fin"=final) and language group ("AE"=American English, "Du"=Dutch), and have been pooled across phonetic contexts and subjects.

TABLE IX. Feature system used for the information-theoretical analyses.

Feature	Values	Phonemes
Consonant manner	stop	/p t k b d g/
	affricate	/tʃ dʒ/
	fricative	/f θ s ʃ h v ð z ʒ/
	liquid	/l r/
	glide	/j w/
	nasal	/m n ŋ/
Consonant place	labial	/p f b v m w/
	dental	/θ ð/
	alveolar	/t s d z n l/
	palatal	/ʃ tʃ ʒ dʒ r/
	velar	/k g ŋ/
	glottal	/h/
Consonant voice	voiced	/b d g v ð z ʒ dʒ j m n ŋ l r w/
	voiceless	/p t k f θ s ʃ tʃ h/
Vowel height	high	/i ɪ ʊ u/
	mid	/eɪ ɛ ʌ oʊ ɚ/
	low	/æ ɑ ɔ/
Vowel backness	front	/i ɪ eɪ ɛ æ/
	central	/ɚ/
	back	/ɑ ʌ ɔ oʊ ʊ u/
Vowel tenseness	tense	/i eɪ ɔ oʊ u ɚ/
	lax	/ɪ ɛ æ ɑ ʌ ʊ/

Thus although the non-native listeners in our study unquestionably performed below native phoneme-identification levels, they did so at more advantageous SNRs as well as under more severe noise, and the degree to which they suffered additional difficulty appeared to remain fairly constant across SNRs within the range tested here. In all our analyses, adverse effects of noise on non-native listening seemed to parallel adverse effects of noise for native listeners. We conclude from these results that it is not disproportionately increasing problems of phoneme identification that underlie the extra difficulty of listening to non-native language in noise.

Instead, we suggest that non-native listening is disproportionately affected by noise because non-native listening is, at all processing levels, slower and less accurate than native listening. Phoneme identification is, as we have seen, less accurate. Phoneme identification problems may be particularly important in that all later levels of processing will be affected by the decisions made at the phonemic level; but at all later levels, non-native listening is also less efficient. Segmentation of continuous speech into words is less efficient, because of interference from native prosodic expectations (Cutler *et al.*, 1986; Cutler and Otake, 1994) and from native phonotactic expectations (Weber, 2001). Lexical recognition is less efficient: phoneme identification problems can cause pseudo-homophony (Japanese listeners may have difficulty distinguishing *right* from *light*, Dutch listeners may confuse *bat* with *bet*), and this can lead to additional competition in the word recognition process (Broersma, 2002; Weber and Cutler, 2004). Spurious competition also arises from the native vocabulary, while native recognition is less likely to be affected by competition from nondominant non-native languages (Weber and Cutler, 2004). Syntactic processing is less efficient, even at high levels of proficiency in the non-native language (Sorace, 1993); prosodic distinctions between idiomatic and literal utterances are less efficiently processed (Vanlancker-Sidtis, 2003); and semantic processing, including the exploitation of prosody for information structure, is less efficient (Akker and Cutler, 2003). The effect of disadvantageous listening conditions, such as a babble of voices, is to slow down the process from the beginning, allowing the cumulative effects of lesser efficiency at all levels to become more noticeable and perhaps to exceed thresholds of auditory memory storage. Compensatory sources of information which all listeners will call upon under difficult listening conditions—knowledge of relative lexical frequencies of occurrence, of transitional probabilities, and of contextual plausibility—will also be less extensive, and less efficiently exploited, in non-native listening [as, indeed, Florentine and her colleagues observed (Florentine, 1985a; Mayo *et al.*, 1997)]. Interestingly, Van Wijngaarden *et al.* (2002) showed that a measure of linguistic entropy (letter-by-letter guessing of visually presented materials) significantly predicted the speech recognition performance of non-native listeners in noise; these authors therefore also concluded that less effective use of context, especially reduced exploitation of semantic redundancy, was a major factor in non-native listening difficulty in noise.

V. CONCLUSION

The identification of phonemes is adversely affected by increasing noise to a similar extent for native and for non-native listeners. Non-native identification scores were around 80% of native identification scores at each of the SNRs used in the present study. This pattern of results suggests that the robustly observed disproportionate difficulty which non-native listeners experience with speech in noisy conditions cannot be simply attributed to exacerbation of phoneme misidentification by noise interference; instead, it may reflect cumulative effects of lesser efficiency at all levels of processing, and lesser ability to exploit contextual redundancy.

ACKNOWLEDGMENTS

Participant testing in Florida was enabled by postdoctoral support to AW from NICHD Grant No. 00323 to Winifred Strange, whose support with this project is gratefully acknowledged. Participant testing in Nijmegen was supported by a research stipend from the Max Planck Society to NC. Further support was provided by a SPINOZA award from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek to AC. The native listening results were reported to the 15th International Congress of Phonetic Sciences, Barcelona (Weber and Smits, 2003), and the non-native listening results to the 8th International Conference on Spoken Language Processing, Jeju, Korea (Cooper and Cutler, 2004). We further thank Natasha Warner for recording the speech materials used in this study, and Randy Diehl, José Benkí, Sander van Wijngaarden, and an anonymous reviewer for comments which helped us improve this report.

APPENDIX A: PHONEMES USED IN THIS STUDY

Phonemes used in the study, with for each phoneme the illustrative word used to guide listeners' responses.

Final consonants		Initial consonants		Vowels	
/p/	liP	/p/	Pie	/i/	bEAt
/t/	hoT	/t/	Tie	/ɚ/	bIRd
/k/	siCK	/k/	Car	/u/	bOOt
/b/	graB	/b/	Be	/ɪ/	bIt
/d/	oDD	/d/	Do	/ʊ/	cOOk
/g/	eGG	/g/	Go	/eɪ/	wAIt
/f/	oFF	/f/	Far	/ɔ/	cAUght
/θ/	paTH	/θ/	THin	/ʌ/	cUt
/s/	paSS	/s/	See	/ɛ/	bEt
/v/	loVE	/v/	Very	/ɑ/	hOt
/ð/	smooTH	/ð/	THere	/aɪ/	bUY
/z/	buZZ	/z/	Zoo	/oʊ/	bOAt
/m/	aM	/m/	My	/æ/	bAt
/tʃ/	suCH	/tʃ/	CHin	/ɔɪ/	bOY
/dʒ/	eDGE	/dʒ/	Joke	/aʊ/	shOUt
/n/	oN	/n/	No		
/ʃ/	fiSH	/ʃ/	SHe		
/ʒ/	beiGE	/w/	Win		
/ŋ/	riNG	/j/	Yell		
/l/	iLL	/h/	Hi		
/r/	faR	/l/	Lie		
		/r/	Row		

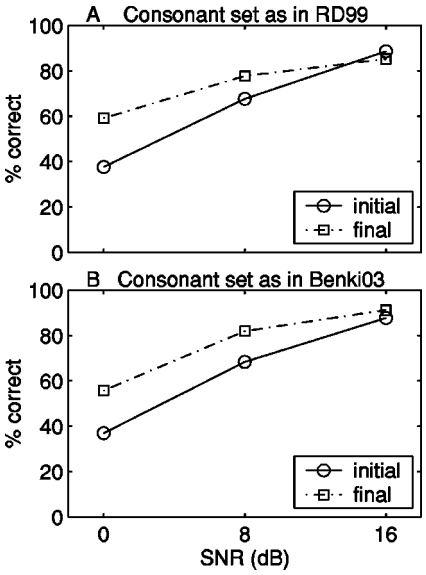

FIG. 4. Percentages of correctly recognized consonants in initial and final position, when consonant sets are restricted to those used by Redford and Diehl (1999), panel (a), and Benkaí (2003a,b), panel (b). Only the data for the American English listeners are given. In the calculations for panel (a), voicing errors have been disregarded, as in Redford and Diehl (1999).

APPENDIX B: ANALYSES OF OUR NATIVE DATA IN COMPARISON WITH REDFORD AND DIEHL (1999) AND BENKI (2003a)

Since we did not find the consistent advantage for initial over final consonants which the listening in noise studies of Redford and Diehl (1999) and Benkí (2003a) had led us to expect, we conducted analyses on our consonant data set in direct comparison with the previous work. These earlier studies differed from ours *inter alia* in that they tested a narrower range of phonemes. We thus extracted from our native listening data set the subset most closely matching the data presented in each of those studies; the relevant subsets are displayed in Fig. 4.

Figure 4(a) shows the American English listeners' identification performance for the voiceless stops and fricatives [p,t,k,f,θ,s,ʃ], with voicing errors ignored [as reported by Redford and Diehl (1999)]. It can be seen that the advantage for final over initial positions holds over 0 and 8 dB SNR in our data set, but disappears—to be replaced by a marginal advantage for initial over final position—at 16 dB SNR. Redford and Diehl used a SNR of 15 dB; thus, at the conditions most closely approximating theirs, we find a result more similar to theirs. At less favorable SNRs, however, the advantage for final position is also robust with this subset.

Figure 4(b) presents the American English listeners' identification performance for the ten initial consonants [b,p, ,d,t,k,s,h,m,l,r] and the ten final consonants [p,d,t,g,k, ,s,z,m,n,l] as used by Benkí (2003a). Again, the final position advantage is stronger at 0 and 8 dB SNR than at 16 dB SNR. However, Benkí used less favorable SNRs (from −14 to −5 dB) than any used in our study, so that our final position advantage with the same subset contrasts with his results.

There were further differences between the studies: we used just one speaker (Redford and Diehl had seven); final stops were released in Benkí's and our stimuli but not in Redford and Diehl's; neither earlier study used multi-speaker babble noise; in both earlier studies, stimuli were CVC syllables in a carrier phrase (respectively *Say—some more* or *Say—again*, and *You will write—please*), while in our study, listeners heard CV or VC syllables in isolation. We chose this format because we were interested in the implications of our findings for natural listening in noise, in which precise onsets are not predictable, and certainly are not accompanied by a constant preceding context. In our study, the syllables varied in length and were centrally embedded in the longer sample of noise, so that the moment of onset of the syllable to be identified was unpredictable and not cued by the preceding context. Under these conditions, the initial vowel or consonant was generally somewhat difficult to identify.

Redford and Diehl (1999) interpreted their positional finding as a result of greater articulatory distinctiveness of initial consonants, a result supported by acoustic evidence that their speakers' initial consonants were longer, louder, and different in fundamental frequency from the final consonants. Benkí (2003a) similarly cited articulatory differences as a likely source of his initial-position advantage. Note also that Benkí (2003b) found that the disadvantage of final consonants largely disappeared when the stimuli presented were words in sparse phonetic neighborhoods, making the final consonant relatively more probable. It seems clear that the relative perceptibility of phonemes as a function of position is not constant, but depends upon the particular characteristics of stimuli and procedure used in a phoneme identification experiment.

Akker, E., and Cutler, A. (**2003**). "Prosodic cues to semantic structure in native and nonnative listening," Bilingualism: Language and Cognition **6**, 81–96.

Benkí, J. R. (**2003a**). "Analysis of English nonsense syllable recognition in noise," Phonetica **60**, 129–157.

Benkí, J. R. (**2003b**). "Quantitative evaluation of lexical status, word frequency, and neighborhood density as context effects in spoken word recognition," J. Acoust. Soc. Am. **113**, 1689–1705.

Best, C. T. (**1995**). "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Booij, G. (**1995**). *The Phonology of Dutch* (Clarendon P, Oxford).

Broersma, M. (**2002**). "Comprehension of non-native speech: Inaccurate phoneme processing and activation of lexical competitors," in *Proceedings of the 7th International Conference on Spoken Language Processing* (Center for Spoken Language Research, University of Colorado Boulder, Denver) (CD-ROM), pp. 261–264.

Conrad, L. (**1989**). "The effects of time-compressed speech on native and EFL listening comprehension," Stud. Second Language Acquisition **11**, 1–16.

Cooper, N., and Cutler, A. (**2004**). "Perception of non-native phonemes in noise," in *Proceedings of the 8th International Conference on Spoken Language Processing* (Jeju, Korea).

Cutler, A., and Otake, T. (**1994**). "Mora or phoneme? Further evidence for language-specific listening," J. Memory Lang. **33**, 824–844.

Cutler, A., Mehler, J., Norris, D., and Seguí, J. (**1986**). "The syllable's differing role in the segmentation of French and English," J. Memory Lang. **25**, 385–400.

Florentine, M. (**1985a**). "Non-native listeners' perception of American-English in noise," in *Proceedings of Inter-Noise '85*, pp. 1021–1024.

Florentine, M. (**1985b**). "Speech perception in noise by fluent, non-native listeners," Proc. Acoust. Soc. Japan **26**, 1–8.

Gat, I. B., and Keith, R. W. (**1978**). "An effect of linguistic experience," Audiology **17**, 339–345.

Greene, B. G., Pisoni, D. B., and Gradman, H. L. (**1985**). ''Perception of synthetic speech by nonnative speakers of English,'' Speech Research Laboratory, Progress Report 11, Indiana Univ.
Gussenhoven, C. (**1999**). ''Dutch,'' in *Handbook of the International Phonetic Association* (Cambridge U.P., Cambridge, UK), pp. 74–77.
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.
Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (**1995**). ''Acoustic characteristics of American English vowels,'' J. Acoust. Soc. Am. **97**, 3099–3111.
Kalikow, D. N., Stevens, K. N., and Elliott, L. L. (**1977**). ''Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability,'' J. Acoust. Soc. Am. **61**, 1337–1351.
Kenstowicz, M. (**1994**). *Phonology in Generative Grammar* (Blackwell, Cambridge, MA).
Kreul, E. J., Nixon, N. C., Kryter, K. D., Bell, D. W., Lang, J. S., and Schubert, E. G. (**1968**). ''A proposed clinical test of speech discrimination,'' J. Speech Hear. Res. **11**, 536–552.
Ladefoged, P. (**1999**). ''American English,'' in *Handbook of the International Phonetic Association* (Cambridge U.P., Cambridge, UK), pp. 41–44.
Mack, M. (**1988**). ''Sentence processing by non-native speakers of English: Evidence from the perception of natural and computer-generated anomalous L2 sentences,'' J. Neurolinguist. **3**, 293–316.
Mayo, L. H., Florentine, M., and Buus, S. (**1997**). ''Age of second-language acquisition and perception of speech in noise,'' J. Speech Hear. Res. **40**, 686–693.
Miller, G. A., and Nicely, P. E. (**1955**). ''An analysis of perceptual confusions among some English consonants,'' J. Acoust. Soc. Am. **27**, 338–352.
Nábělek, A. K., and Donahue, A. M. (**1984**). ''Perception of consonants in reverberation by native and non-native listeners,'' J. Acoust. Soc. Am. **75**, 632–634.
Peterson, G. E., and Barney, H. L. (**1952**). ''Control methods used in a study of the vowels,'' J. Acoust. Soc. Am. **24**, 175–184.
Pisoni, D. B. (**1987**). ''Some measures of intelligibility and comprehension,'' in *From Text to Speech: The MITalk System*, edited by J. Allen, D. H. Klatt, and S. Hunnicutt (Cambridge U.P., Cambridge, UK), pp. 151–171.
Redford, M. A., and Diehl, R. L. (**1999**). ''The relative perceptual distinctiveness of initial and final consonants in CVC syllables,'' J. Acoust. Soc. Am. **106**, 1555–1565.
Smits, R. (**2000**). ''Temporal distribution of information for human consonant recognition in VCV utterances,'' J. Phonetics **27**, 111–135.
Smits, R., Warner, N., McQueen, J. M., and Cutler, A. (**2003**). ''Unfolding of phonetic information over time: A database of Dutch diphone perception,'' J. Acoust. Soc. Am. **113**, 563–574.
Sorace, A. (**1993**). ''Unaccusativity and auxiliary choice in non-native grammars of Italian and French: Asymmetries and predictable indeterminacy,'' J. French Lang. Studies **3**, 71–93.
Strange, W. (**1995**). *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research* (York P, Timonium, MD).
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., Steeneken, H., and Houtgast, T. (**2002**). ''Quantifying the intelligibility of speech in noise for non-native listeners,'' J. Acoust. Soc. Am. **111**, 1906–1916.
Vanlancker-Sidtis, D. (**2003**). ''Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one,'' Appl. Psycholinguist. **24**, 45–57.
Wang, D. M., and Bilger, R. C. (**1973**). ''Consonant confusions in noise: A study of perceptual features,'' J. Acoust. Soc. Am. **54**, 1248–1266.
Weber, A. (**2001**). *Language-specific Listening: The Case of Phonetic Sequences*, MPI Series in Psycholinguistics, Vol. 16, Ph.D. dissertation, University of Nijmegen, The Netherlands.
Weber, A., and Cutler, A. (**2004**). ''Lexical competition in non-native spoken-word recognition,'' J. Memory Lang. **50**, 1–25.
Weber, A., and Smits, R. (**2003**). ''Consonant and vowel confusion patterns by American English listeners,'' in *Proceedings of the 15th International Congress of Phonetic Sciences* (Palau de Congressos, Barcelona, Spain), pp. 1437–1440.