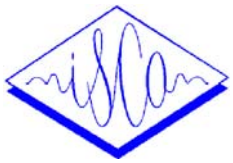
ISCA Archive
http://www.isca-speech.org/archive



INTERSPEECH 2004 - ICSLP
8th International Conference on Spoken Language Processing
ICC Jeju, Jeju Island, Korea
October 4-8, 2004

Perception of Non-native Phonemes in Noise

Nicole Cooper and Anne Cutler

Max Planck Institute for Psycholinguistics
Nijmegen, The Netherlands

anne.cutler@mpi.nl

Abstract

We report an investigation of the perception of American English phonemes by Dutch listeners proficient in English. Listeners identified either the consonant or the vowel in most possible English CV and VC syllables. The syllables were embedded in multispeaker babble at three signal-to-noise ratios (16 dB, 8 dB, and 0 dB). Effects of signal-to-noise ratio on vowel and consonant identification are discussed as a function of syllable position and of relationship to the native phoneme inventory. Comparison of the results with previously reported data from native listeners reveals that noise affected the responding of native and non-native listeners similarly.

1. Introduction

Listening to speech in noise is always difficult; but the increase in difficulty is greater for speech in a non-native language than for speech in the native language. This is a common subjective experience for non-native listeners, and it has been attested in listening experiments (e.g. [1,2,3]).

Most studies of listening in noise use simple sentences as stimuli, or words in a constant context. This does not enable a precise estimate of the extent to which non-native listeners' difficulty with speech in noise is due purely to phonetic discrimination problems. It is well known that discrimination of non-native phoneme contrasts can be highly prone to error, especially when the phoneme categories of the native and the non-native language mismatch (see, e.g. [4,5]). Failure to tell one word from another (as when English *right* and *light* sound similar to speakers of languages without the contrast [l] -[r]) is an obvious problem for non-native listening; if the ability to identify phonemes is affected to a greater extent by noise for non-native than for native listening, this factor alone could explain the greater overall effect of noise on non-native comprehension. However, it may also be the case that greater effects of noise are not due to phoneme recognition problems as such but to problems at higher levels of processing. Mayo et al. [3] reported that native listeners made more effective use of contextual plausibility when listening in noise than non-native listeners did; this difference too could account for the greater influence of noise for non-native listeners.

In the present study non-native listeners were presented with simple vowel-consonant or consonant-vowel sequences, and required to identify either the consonant or the vowel. In this procedure the identification of phonemes can be examined without any influence from lexical knowledge or contextual probabilities. The listeners were native speakers of Dutch whose English proficiency was high. Nevertheless there are mismatches between the phonemic inventories of Dutch and English which cause phoneme identification problems for Dutch listeners to English.

2. Method

2.1. Participants

Sixteen undergraduate students at the University of Nijmegen, all native speakers of Dutch with good knowledge of English and no hearing impairment, took part in the experiment and were paid a small sum for their participation.

2.2. Materials

Dutch has 35 phonemes: 19 consonants and 16 vowels [6]; American English has 24 consonants and 16 vowels [7]. Thus there are several English consonants without counterpart in Dutch. The vowels of the two languages also mismatch in many ways. Dutch has more high and mid vowels and fewer low vowels than English; especially the contrast in English *bat-bet* is difficult for Dutch listeners.

The American English phonemes were combined to form all possible CV and VC sequences, excepting those involving schwa. All phonemes occurred syllable-initially and -finally except for /h/, /w/, and /j/, which occurred only in initial position, and /ŋ/ and /ʒ/, which occurred only in final position.

The complete set of stimuli, comprising 345 syllables, was transcribed phonemically. A phonetically trained female native speaker of American English, seated in a quiet room, read the transcriptions into a high-quality microphone. The sampling rate during digitization was 16 kHz. Each syllable was then centrally embedded in one second of multispeaker babble, constructed from the recorded speech of three male and three female speakers, at three different signal-to-noise ratios (SNR; 16 dB, 8 dB, and 0 dB). These SNRs were chosen, on the basis of a pretest, to yield easy, intermediate, and difficult phoneme perception for the non-native listeners.

2.3. Procedure

Over eight sessions, each listener heard all CV and VC syllables in the three SNRs twice, once identifying the vowel and once identifying the consonant, for a total of 3870 trials (645 syllables x 3 SNR x 2 presentations). The presentation of items was self-paced. If the listener did not respond within 15 seconds after stimulus offset, the trial was recorded as a miss. Each listener was presented with the items in a different pseudo-random order. In each session, listeners had to identify blocks of initial or final consonants and blocks of vowels. They responded by clicking the word that contained the appropriate sound on a computer screen (see Tables 1 and 2 for examples of these words). Different words were used for vowels, initial consonants, and final consonants. Participants were familiarized with the words before the experiment.

10.21437/Interspeech.2004-197

3. Results

3.1. Overall

The response rate was very high, with a miss being recorded on less than 0.1% of trials.

In Figure 1 the mean percentages of correct responses are shown for vowels and consonants respectively as a function of signal-to-noise ratio. It can be seen that with only mild noise (16 dB SNR), consonants are more accurately identified than vowels, but while an increase in the level of noise has little effect on the accuracy of vowel identification, consonant identification is seriously impaired, and drops to below 40% at 0 dB SNR. ANOVAs confirmed that the interaction between noise level and phoneme type was significant ($F[2,30] = 811.5$, $p < .001$). Post-hoc analyses revealed that there was a significant difference between 16 dB and 8 dB SNR for consonants only, but for both vowels and consonants there was a significant difference between 8 and 0 dB SNR.

Figure 2 shows the effects of syllable position on the accuracy of responses (averaged across SNR). It can be seen that the effects were different for vowels and for consonants; the position effect interacted with phoneme type ($F[1,15] = 26.91$, $p < .001$). For vowels, identification accuracy was higher in final position than in initial position, while for consonants the reverse pattern was observed. With the kind of stimuli we used, isolated syllables in babble noise, it is not possible to predict the precise onset of a stimulus item; therefore the advantage of final position which we see here for vowels only may be considered the expected result.

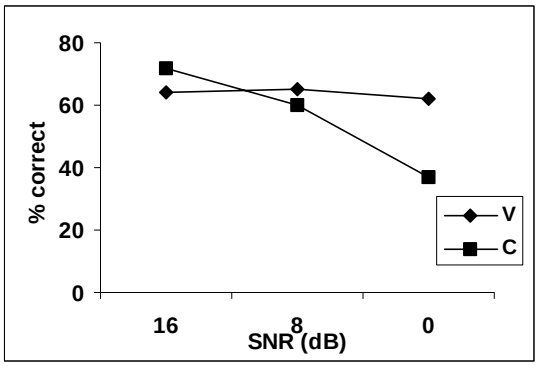


Figure 1: Mean percentages of correct responses for vowels (V) and consonants (C) as a function of increasing noise.

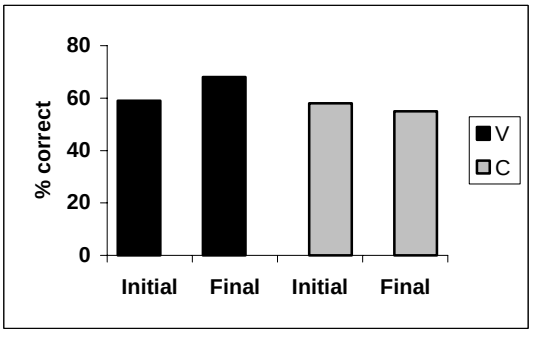


Figure 2: Mean percentages of correct responses for vowels and consonants in syllable-initial and syllable-final position.

The results for consonants thus demand explanation. Although, as we pointed out above, the consonant inventory of English is some 25% larger than that of Dutch, differences in inventory size alone cannot explain a selective impairment in one position. The obvious source for such a finding lies in differences in the phonological systems of the two languages. In English, obstruent voicing can contrast in syllable-final position: *bat* can contrast with *bad*, *lace* with *laze*. In Dutch, all syllable-final obstruents are voiceless. Thus it is likely that the greater part of the errors for final consonants concern obstruents with a minimal pair differing in voicing. Indeed, this was the case: the overall mean of 55.25% correct for final consonants breaks down to 69.79% for non-obstruents but only 50.38% for obstruents. We therefore ascribe this pattern to influence from the phonology of the native language.

3.2. Comparison with Native Listening

The same stimulus materials as used in the present study had been presented to native American English listeners, with results reported by Weber and Smits [8]. A comparison of the present results with theirs reveals that the non-native listeners performed consistently worse than the native listeners across all SNRs. However, there was little evidence that noise had a significantly greater effect on non-native than on native listening. For vowels, the non-native score was 80.6% of the native score at 16 dB SNR, and actually improved slightly with noise, to 82% of the native score at 8 dB and 80.8% of the native score at 0 dB SNR. For consonants, the trend was in the opposite (and predicted) direction: 84.9% at 16 dB worsened to 82.4% at 8 dB and 76.2% at 0dB SNR. However, joint ANOVAs across the two data sets showed no significant interaction of phoneme type with language, or of SNR with language, or of all three factors together.

The order of difficulty of individual phonemes at each SNR was very similar for the two language groups; even at 0dB SNR the performance of the two groups was highly correlated ($r = .91$), again suggesting similar performance.

3.3. Effects of Native Phoneme Inventory

Complete confusion matrices for the responses at 0 dB SNR are shown in Tables 1 and 2, for vowels and consonants respectively. The responses are pooled across syllable positions and are presented as percentage of each phoneme response (columns) given each stimulus (rows). Example words which were used for listeners to click on to signal their responses head the columns; for consonants (except [w,j,h]) the syllable-final alternatives are given. Where the rows do not sum to 100%, the remainder was missing data.

The rank ordering of the phonemes was similar at better SNRs; we here present in full the 0 dB results because these show the most errors and thus allow the best opportunity for examining effects on identification performance of mismatch between the native and the non-native inventory. The vowel inventories of Dutch and American English are similarly crowded [6,7]; as described in section 3.2, the non-native listeners found the same vowels easy or difficult as the native listeners. Both the hardest third of the vowels (back and central vowels) and the easiest third (diphthongs and high front vowels) were identified more than 97% as well at 0dB as at 16dB. Confusions of *bat* with *bet* were almost the same at the three SNRs. The vowel results thus show no particular interaction of the effects of noise with the native repertoire.

Table 1: Dutch listeners' confusion matrix for English vowels at 0 dB SNR.

		response															
		beat	bit	wait	bet	bat	hot	cut	caught	boat	cook	boot	buy	boy	shout	bird
		i	ɪ	eɪ	ɛ	æ	ɑ	ʌ	ɔ	oʊ	ʊ	u	aɪ	ɔɪ	aʊ	ɚ
stimulus	i	**86.8**	8.3	1.0	1.7		.1	.3			.7		.1	.1	.1	.4
	ɪ	.9	**90.8**	.3	3.8	.1	.1	.1		.1	.1	.7	.6		.1	1.9
	eɪ	15.3	9.3	**65.8**	4.8	2.6		.1	.4	.1		.1	.4	.3		.6
	ɛ	.3	14.1	.6	**59.4**	23.5		.1	.1						.3	1.3
	æ		1.3	3.1	36.5	**53.9**	.4	.9		.1			.7		2.0	1.0
	ɑ	.3	.3		.6	15.1	**26.3**	27.0	22.1	1.9	.3	.3	2.8	.3	1.6	.9
	ʌ		.1	.4	1.3	7.6	26.2	**42.9**	12.4	2.6	.3	.4	1.5	.1	2.2	2.0
	ɔ					2.8	48.7	5.4	**36.3**	2.9	.3	.6	.3	1.6	.3	.7
	oʊ		.3	.3		.1	9.9	.7	3.5	**61.5**	8.7	10.3	.4	2.2	1.5	.3
	ʊ	.3	.4			.1	8.4	4.2	3.5	2.8	**62.4**	11.9	.6	2.5	1.5	1.2
	u	13.4	.9	.1	.3	.1	1.6	1.2	1.5	2.8	38.8	**37.4**	.3		1.2	.6
	aɪ		3.2	19.6	.4	1.9	.1	.4	.6	.7			**69.3**	1.2	.1	2.3
	ɔɪ			.9			2.9		.4	.9	.4	.1	1.3	**92.2**	.7	.1
	aʊ		.1	.3	2.3	.6	3.3	.3	3.9	16.7	.6	.9	2.8	.6	**67.3**	.3
	ɚ	.6	.3	.1	.9	.1	.4	12.9	.6	.1			.6		.4	**82.8**

Table 2: Dutch listeners' confusion matrix for English consonants at 0 dB SNR.

		response																						
		lip	hot	sick	off	path	pass	fish	such	hi	grab	odd	egg	love	smooth	buzz	beige	edge	yell	am	on	ring	ill	far	win
		p	t	k	f	θ	s	ʃ	tʃ	h	b	d	g	v	ð	z	ʒ	dʒ	j	m	n	ŋ	l	r	w
stimulus	p	**27.5**	8.5	10.4	7.7	5.8			.4	9.6	16.5	3.5	1.7	1.5	2.5	.2	.4		.4	.6	.8		.8	.4	.6
	t	14.6	**29.6**	9.0	4.6	8.5	1.3		2.1	5.6	4.4	10.4	1.3	.8	3.5	.2	.2	.4		1.0	1.7		.4	.4	
	k	16.7	10.2	**35.2**	4.0	4.6	.4	.6	.6	6.9	3.3	2.5	8.1	.8	1.3	.2		.8	.6	.2	1.0		.6	.8	.4
	f	16.0	11.9	7.9	**18.5**	8.3	.8	.8	.4	4.6	10.6	6.0	1.7	4.4	4.6	.2	.4	.4	.2	.6	.2		.4	.4	.4
	θ	13.3	15.4	3.8	15.6	**14.8**	.6	.4	1.0	3.5	8.5	5.2	1.0	2.7	8.3		.4	.6	.2	.6	1.5		1.5	.2	.4
	s	.2	3.1	.2	14.8	19.6	**34.0**	2.9	.6		.6	.8		2.9	9.0	9.6	.8	.4						.4	
	ʃ		.2			.8	6.7	**69.6**	14.4				.2		.6	.8	5.0	1.0	.4					.2	
	tʃ	1.9	2.5	.6	1.3	1.9	.8	5.6	**56.7**	.6	1.0	1.9	.2	.2	2.1		2.7	19.8	.2						
	h	26.3	4.6	12.1	11.3	5.0	.4	.4	.8	**17.9**	8.3	1.3	.4	4.6	1.7	.4			.8		.8		1.7	.8	.4
	b	6.3	4.0	6.3	6.0	3.5	.2		.4	6.3	**29.4**	8.8	2.1	6.3	3.5	.6	.2	1.5	1.5	4.8	1.9		2.9	1.3	2.5
	d	1.9	9.2	.8	1.9	5.6	.4	.2		4.4	7.1	**25.2**	2.3	2.9	10.4	1.0	1.5	3.1	3.1	2.3	7.1	1.3	6.3	1.0	.8
	g	2.1	7.1	6.9	1.9	5.0		.2	1.0	4.6	6.0	12.7	**21.5**	2.9	4.2	.4	.4	3.5	12.1	1.7	2.1	.4	1.7	.8	.8
	v	4.4	7.9	3.1	10.4	5.2	.2	.8	.2	3.3	17.3	8.5	2.5	**12.7**	6.7	.6	.6	.6	1.0	2.9	2.5	.6	2.3	2.7	2.7
	ð	1.5	6.3	1.7	2.9	11.5	2.7	1.0	1.0	1.0	9.8	19.6	1.0	5.6	**13.5**	2.1	.6	4.8	.6	.8	2.3		6.5	1.5	1.5
	z		2.3	.6	1.5	9.8	7.7	1.3	1.9		4.2	6.9	1.0	3.8	16.9	**26.5**	2.5	1.7	1.3	2.1	3.3	.6	.8	1.7	1.9
	ʒ		3.3	.4		2.1	2.5	14.2	4.2		.4	3.8	.4	1.3	4.2	6.7	**45.0**	9.2		.4	.4		.8	.8	
	dʒ	2.1	1.7	1.3	1.0	2.5	.4	1.9	15.8	1.0	1.3	7.7	1.0		5.8	.2	4.8	**46.0**	2.9	.2	.6		1.5	.2	
	j	1.3		.8	.8	.8		.8	.4	2.1	4.2	2.5	1.3	.4	1.3	.4		4.6	**69.6**	2.5	4.2		1.3		.8
	m	1.9	5.0	1.3	1.7	2.5	.4			1.0	4.8	3.3	.6	2.3	1.0	.4				**45.6**	15.0	4.2	3.5	3.3	2.1
	n	.2	4.8		.6	.8	.2	.2	.2	1.0	1.3	4.8		.2	1.0	1.0	.2	1.5	.2	10.6	**61.0**	4.6	3.3	1.3	.8
	ŋ		6.7	.4	2.1	2.1	.4				2.5	5.8	6.3		1.7	1.7				13.8	22.5	**30.4**	2.1	1.7	
	l	3.1	4.8	1.7	3.5	2.3		.2	.2	1.0	4.2	5.4	1.0	2.3	2.5	.2			.6	5.4	3.3	.2	**52.1**	3.1	2.7
	r	1.5	4.6	1.0	1.5	1.9		.2	.2	2.9	8.1	3.8	1.0	1.5	1.9	.2	.2	.4		.4	1.3		1.0	**63.1**	3.3
	w	1.7		.4	.4	.4	.4			1.7	5.8	.8		2.1	.4					5.8	.8		2.5	1.7	**75.0**

The results for the consonants pattern differently. Overall, performance on consonant identification at 0dB was only 51% of performance at 16dB. Performance on the best third (sonorants) was more robust (0dB 60% of 16dB) and on the bottom third was less robust (0db overall 50% of 16dB). Strikingly, however, the performance on the four English consonants with no Dutch counterpart (the final consonants of *path*, *smooth*, *edge* and *egg*) was very badly hit by noise: at 0dB identification was only 41% of the 16dB level. These four consonants were in fact largely responsible for the greater effect of noise for native than for non-native listeners (section 3.2); without them, non-native performance at 0dB is 81.5% of overall native performance at 0dB. Thus it appears that particularly the identification of unfamiliar consonantal categories is susceptible to disruption by noise.

4. Conclusions

This study has shown that noise seriously impairs the efficiency of non-native phoneme identification. It has also shown that the identification of non-native phonemes is subject to influence from the categorical structure of the native phoneme inventory as well as from the phonotactic constraints applying to the native language. Rather surprisingly, however, the influence of these native-language factors in general did not significantly increase under increasing levels of noise: the factors exercised an effect at more advantageous SNRs as well as under worse noise conditions. Only the identification of consonants without native counterpart appeared to be more affected by noise than other phonemic decisions. Even more surprising was the comparison with native listening data on the same materials, which revealed highly similar effects of noise on identification performance of the two listener groups. The performance of the non-native listeners was consistently worse than that of the native listeners, but the disadvantage which they displayed did not increase (indeed, in the case of vowel identification it even to a small extent decreased) as a function of increasing levels of concurrent noise.

These results lead us to suggest that the exceptional difficulty of listening to non-native language in noise [1,2,3] is not due, or at least not solely due, to problems of phonetic identification. Of course phonetic identification problems exist, but they exist even with listening in the clear, and noise does not disproportionately exacerbate them. It is known that these problems directly lead to an increase in the set of potential competitors for word recognition in non-native listening [9,10], and we suggest that it is rather at the word recognition level, or at still higher levels of processing [3], that the peculiar difficulty of noisy conditions for non-native listening arises.

5. Acknowledgements

This research was supported by a research stipend from the Max Planck Society to the first author and by a SPINOZA grant from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek to the second author. We thank Natasha Warner, Roel Smits and Andrea Weber for substantial help.

6. References

[1] Nábělek, A.K. and Donahue, A.M., "Perception of consonants in reverberation by native and non-native listeners", *J. Acoust. Soc. Am.*, 75: 632-634, 1984.

[2] Takata, Y. and Nábělek, A.K., "English consonant recognition in noise and in reverberation by Japanese and American listeners", *J. Acoust. Soc. Am.*, 88: 663-666, 1990.

[3] Mayo, L.H., Florentine, M., and Buus, S., "Age of second-language acquisition and perception of speech in noise", *J. Speech Hear. Res.*, 40: 686-93, 1997.

[4] Best, C.T., "A direct realist view of cross-language speech perception", in W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language speech research*, pp. 171-204, York Press, Timonium, MD, 1995.

[5] Flege, J.E., "Second language speech learning: Theory, findings, and problems," in W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research*, pp.233-277, York Press, Timonium, MD, 1995.

[6] Gussenhoven, C., "Dutch", in *Handbook of the International Phonetic Association*, pp. 74-77, Cambridge University Press, Cambridge, UK, 1999.

[7] Ladefoged, P., "American English", in *Handbook of the International Phonetic Association*, pp. 41-44, Cambridge University Press, Cambridge, UK, 1999.

[8] Weber, A. and Smits, R., "Consonant and vowel confusion patterns by American English listeners", in *Proceedings of the 15th International Congress of Phonetic Sciences*, pp. 1437-1440, Palau de Congressos, Barcelona, Spain, 2003.

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

[10] Weber, A., and Cutler, A., "Lexical competition in non-native spoken-word recognition", *J. Memory and Language*, 50: 1-25, 2004.