Original Paper

Phonetica 2003;60:129–157
DOI: 10.1159/000071450

Received: June 5, 2002
Accepted: October 15, 2002

Analysis of English Nonsense Syllable Recognition in Noise

José R. Benkí

Department of Linguistics, University of Michigan, Ann Arbor, Mich., USA

Abstract

English nonsense consonant-vowel-consonant syllables were presented at four different signal-to-noise ratios for recognition. Information theory methods are used to analyze the response data according to segment type and phonological feature, and are consistent with previous studies showing that the consonantal contrast of voicing is more robust than place of articulation, word-initial consonants are more robust than word-final consonants, and that vowel height is more robust than vowel backing. Asymmetrical confusions are also analyzed, indicating a bias toward front vowels over back vowels. The results are interpreted as parts of phonetic explanations for synchronic and diachronic phonological patterns.

Introduction

While much phonetic research uses speech materials recorded or presented under ideal conditions, listeners deal with degraded speech or imperfect listening conditions on a regular basis, if not all the time for some listeners. A number of studies have investigated speech perception in noise [Fletcher, 1953; see review by Allen, 1994; Miller et al., 1951; Savin, 1963; Luce and Pisoni, 1998; Alwan et al., 2001]. As observed in these studies, due to the redundant nature of the speech signal, listeners are able to cope with degradation but do make misperceptions. As pointed out in a classic report of perceptual confusions [Miller and Nicely, 1955], the pattern of misperceptions is not random and can offer valuable information for our understanding of speech perception. The present study is a report of some analyses of such misperceptions by American English listeners.

Miller and Nicely [1955] and Wang and Bilger [1973] report analyses of consonant confusions, while Pickett [1957] reports an analysis of vowel confusions. Both consonant studies presented English consonant-vowel (CV) syllables in noise for identification; Wang and Bilger [1973] presented vowel-consonant (VC) syllables as well. Participants identified the consonant in each syllable in a closed-set task. Voicing and manner of articulation were the most robust contrasts in both studies, resulting in few confusions, while place-of-articulation confusions were more frequent. Pickett [1957]

KARGER

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2003 S. Karger AG, Basel
0031–8388/03/0602–0129
$19.50/0
Accessible online at:
www.karger.com/journals/pho

José R. Benkí
Department of Linguistics
University of Michigan
4080 Frieze Building, Ann Arbor, MI 48109-1285 (USA)
Tel. +1 734 763 9173, Fax +1 734 936 3406
E-Mail benki@umich.edu

reports a preponderance of confusions between vowels differing in backing (chromaticity) relative to vowels differing in height.

The present study goes further than each of these previous studies by reporting onset consonant, vowel, and coda consonant confusion data in an open-response task by naive listeners, a more realistic setting than the closed-set task performed by the experienced listeners in the previous studies. Sommers et al. [1997] show that closed-set tests, while useful in some regards, may induce perceptual strategies on the part of listeners that may not reflect normal communicative situations.

The experiment presented here is part of an in-depth investigation of lexical context effects using the materials from the first experiment in Boothroyd and Nittrouer [1988]. In the study by Boothroyd and Nittrouer [1988], two phonetically balanced sets of 120 English CVC nonsense syllables and 120 meaningful words were presented to listeners at four different signal-to-noise (S/N) ratios. The same materials, newly recorded, were used in the present investigation. The present paper reports the results for the nonsense CVC syllables but not the meaningful CVC syllables [for a report of the word recognition data see Benkí, 2003].

To preview the results, confusion data for voicing, manner of articulation, and place of articulation in syllable onsets are largely consistent with those reported by Miller and Nicely [1955], Wang and Bilger [1973], and Pickett [1957] as quantified by Miller and Nicely's [1955] measure of proportion of information transmitted. Consonantal contrasts are less robust in the coda than in the onset, and vowel height is more robust than the backing contrast, also measured by proportion of information transmitted. In addition, there was an asymmetrical pattern of confusions between certain phonemes, particularly back vowels heard as their front counterparts but not the reverse, also consistent with Pickett [1957]. Finally, syllables closed by nasals, voiced stops, and /l/ were perceived as containing no final consonants significantly more often than syllables closed by voiceless stops or the fricatives /s z/. These results should be useful for a number of other areas of speech research, including automatic speech recognition, where one of the most serious obstacles to improvement is the problem of degraded speech [Lippmann, 1997], as well as synchronic and diachronic phonology. The implications for synchronic and diachronic phonology in particular are presented in the 'discussion' section.

Method

The procedure for Boothroyd and Nittrouer's [1988] experiment 1, in which participants identified CVC nonsense and word syllables at different noise levels, was followed as closely as possible, except that stimulus presentation and response collection were done online.

Participants

Forty-three young adults were recruited from an undergraduate introductory linguistics course at the University of Michigan and participated for course credit. All were native speakers of English and reported no known hearing problems.

Stimuli

The same lists of CVC syllables developed by Boothroyd and Nittrouer [1988], consisting of 120 words and 120 nonsense syllables, were used for this study and are listed in the 'Appendix', although only the responses to the 120 stimuli in their nonsense syllable list are presented here. Both the word and nonsense syllable lists were phonetically balanced such that the phonemes in the

Table 1. Distribution of onset-vowel (CV) and vowel-coda (VC) combinations in the stimuli

	Vowels									
	a	aɪ	o	u	i	ɪ	e	ɛ	æ	total
Onsets										
p	0	1	3	2	1	2	1	1	1	12
b	4	1	1	1	2	1	1	1	0	12
m	3	2	2	1	2	1	0	1	0	12
h	4	2	0	1	1	0	1	1	2	12
t	3	1	0	2	2	0	3	0	1	12
k	2	1	0	1	1	2	2	2	1	12
s	1	1	3	1	1	1	1	2	1	12
d	4	0	1	1	0	2	0	1	3	12
r	1	2	1	2	1	2	2	1	0	12
l	4	1	1	0	2	0	1	2	1	12
Total	26	12	12	12	13	11	12	12	10	
Codas										
p	1	2	0	1	0	1	2	2	3	12
m	0	1	1	2	2	2	2	3	0	13
t	4	1	2	1	1	2	1	0	0	12
k	4	1	1	3	0	0	1	1	1	12
s	4	2	1	1	3	0	0	0	1	12
d	5	1	2	0	1	1	2	0	0	12
g	0	2	2	1	3	2	1	1	0	12
z	0	0	0	0	3	1	2	3	3	12
n	2	0	2	1	0	2	1	1	2	11
l	6	2	1	2	0	0	0	1	0	12
Total	26	12	12	12	13	11	12	12	10	

sets of 10 initial consonants /b p d t k s h m l r/, 10 vowels /i ɪ e ɛ æ u o ɔ a aɪ/, and 10 final consonants /p d t g k s z m n l/ were fairly evenly distributed in the word and nonsense syllable lists. As /ɔ/ and /a/ are not phonemic for many speakers of American English, they were both counted as /a/. The word-final obstruents were released. Table 1 shows the numerical distributions of the following and preceding phonemes of each onset, vowel, and coda.

Each item was read by the author, a native speaker of American Midwest English, in the carrier phrase 'You will write … please' in a sound-treated room and was recorded to DAT with a Realistic Highball microphone and a Tascam DA-30 digital tape deck at a sampling rate of 48 kHz. The recording of each item embedded in the carrier phrase was converted to a WAV file at the same sampling rate and stored on computer disk. The overall level of the entire carrier phrase containing the target stimulus was adjusted so that the peak amplitudes of all stimuli were matched.

Procedure

The experiment was run using software running in the Matlab (version 6.1) environment on four Windows NT laptop computers in an anechoic chamber. Signal-dependent (though uncorrelated) noise [Schroeder, 1968] was added online. Floor (near-chance) or ceiling (near-perfect) performance is not useful for evaluating perceptual confusions. Pretesting with 6 of the participants determined four S/N ratios (–14 dB, –11 dB, –8 dB, and –5 dB) that would result in performance covering the range between 5 and 95% correct for both phonemes and whole syllables. The remaining 37 participants were randomly assigned to one of these S/N ratios (11 participants at –14 dB, 9 participants at –11 dB, 9 participants at –8 dB, and 8 participants at –5 dB). Only the nonsense syllable results from the

37 participants who performed the task at one of the four selected S/N ratios are reported here. All participants were instructed in writing that they would be listening to real and nonsense CVC syllables of English presented in a carrier phrase mixed with noise, and were to type what they heard using standard English orthography for both the words and nonsense items. A brief list of examples of English orthography for spelling nonsense items was provided.

The stimuli in their carrier phrases were presented for identification binaurally via AKG headphones with the volume set to a comfortable listening level. Following a practice block of 5 items, the stimuli were presented for identification, blocked according to lexical status (word or nonsense syllable). At the beginning of each experiment, the experiment software randomly determined the order of 12 word and 12 nonsense blocks. As a result, word and nonsense blocks did not necessarily alternate. Each block consisted of 10 trials, which were randomly selected from the word or nonsense stimuli as appropriate. At the beginning of each block the computer informed the participant whether the next block of 10 items were either words or nonsense items, how many blocks remained, and then waited for the participant to press the enter key before presenting the first stimulus. Participants typed their responses using the keyboard, and could correct their responses for the errors before pressing the enter key, but could not request the stimulus to be played again. A half-second after the enter key was pressed, the computer presented the next stimulus. Between 1 and 4 participants were run at a time, with each participant at their own computer station. Each session lasted about 30 min. Only the nonsense syllable results are reported here.

Phonemic Analysis

The typed responses of the participants were preliminarily analyzed into phonemes with the text-to-phoneme program t2p [Lenzo, 1998]. The preliminary phonemic analyses were compared to the corresponding original responses and corrected as needed.

Each phoneme response was scored as correct if it matched the corresponding stimulus phoneme, and incorrect otherwise, with the following adjustments. First, /a/ and /ɔ/ were counted as matching vowels. In their stimulus list preparation and response analysis, Boothroyd and Nittrouer [1988] regarded the vowels /a/ and /ɔ/ as distinct phonemes, and these distinctions were maintained in the preparation of the stimuli for the present study. However, these vowels are merged in the English spoken by most of the participants, and were therefore counted as the same vowel for scoring purposes.

Although the participants were told that all of the target items would consist of CVC nonsense syllables, there were occasional deletion misperceptions in which onsetless or open syllables were reported. Deletions of C_1 were relatively few, but a significant number of C_2 deletions were reported. The missing phonemes in both of these types of cases are counted as 'nul' responses in the confusion matrices below.

Additionally, there were occasional epenthetic misperceptions consisting of a consonant cluster (5% of trials at –14 dB, 6% of trials at –11 dB, 5% of trials at –8 dB, and 2% of trials at –5 dB). A misperception consisting an epenthetic consonant C_e was counted as a single 'cluster' response, rather than a half response for each of the two reported consonants, which would result in a nonstandard confusion matrix. The cluster response category was assigned to the closest stimulus segment, which was straightforwardly C_1 for $C_eC_1VC_2$ responses or C_2 for $C_1VC_2C_e$ responses. However, an epenthetic consonant *between* a stimulus consonant and vowel could be considered a misperception of either the consonant or the vowel. Accordingly, when both of the stimulus segments adjacent to C_e were transcribed correctly, half of such cases were scored as clusters for the consonant and the other half of such cases were scored as clusters for the vowel. If only one of the adjacent segments to C_e was transcribed correctly, then the correct response was scored as correct and the incorrect response as a cluster.

It might be noted that these scoring procedures for epenthetic consonants underestimate the correctly perceived phonological information. For example, scoring the stimulus /t/ as a cluster response for the response /stig/ to stimulus /tig/ does not acknowledge the perception of /t/. On the other hand, these cases are errors, and scoring the /t/ as correct would overestimate the correctly perceived material. Given that only 5% of the trials involve cluster responses in the present data, this issue might be better investigated by a study more specifically designed to induce such confusions.

Table 2. Confusion matrices for onsets

		p	t	k	s	h	f	b	d	g	m	n	r	l	w	j	Null	Other
–14 dB	p	88	11	11	0	6	2	4	5	0	0	0	0	0	1	0	3	1
	t	49	60	16	0	1	1	0	0	0	1	0	0	0	0	0	1	3
	k	22	13	86	0	2	1	0	0	0	0	0	0	0	0	0	2	6
	s	0	0	0	122	0	2	0	0	0	0	0	0	0	0	0	2	6
	h	13	7	0	0	61	37	1	0	0	1	0	1	0	0	0	8	3
	b	0	0	0	0	0	1	99	6	1	0	1	3	6	1	2	5	7
	d	0	1	0	0	0	1	19	80	21	0	0	0	0	0	2	1	7
	m	2	1	0	0	0	1	2	0	0	53	17	2	31	2	11	5	5
	r	0	0	0	0	0	0	0	1	0	5	6	30	51	22	11	2	4
	l	1	0	0	0	0	0	1	0	0	6	9	7	86	12	8	2	0
–11 dB	p	74	12	4	0	9	0	3	0	0	1	0	0	0	0	0	1	4
	t	15	79	11	0	0	0	0	1	0	0	0	0	0	0	0	0	2
	k	14	10	67	1	2	0	0	0	0	1	0	0	1	1	0	5	6
	s	0	0	1	93	1	1	0	0	0	0	0	0	0	0	0	2	10
	h	7	2	2	1	79	7	0	0	0	1	1	0	0	1	0	4	3
	b	0	0	1	0	2	1	77	4	1	3	0	3	4	3	0	6	3
	d	0	1	0	0	0	1	10	66	20	1	0	0	1	1	1	3	3
	m	0	0	1	0	2	0	2	0	0	45	13	3	27	3	6	3	3
	r	1	0	0	0	0	0	1	0	0	10	1	17	46	22	6	1	3
	l	0	0	1	0	0	0	2	0	0	6	2	4	75	10	2	2	4
–8 dB	p	95	3	3	0	3	0	1	0	0	0	0	0	0	0	0	3	0
	t	12	87	4	0	1	0	0	1	0	0	0	0	0	0	0	3	0
	k	9	6	84	0	1	0	0	0	0	0	0	0	0	0	0	4	4
	s	0	0	0	104	0	0	0	0	0	0	0	0	0	0	0	3	1
	h	6	0	2	0	93	3	1	0	0	0	0	0	0	0	0	3	0
	b	1	0	0	0	0	0	90	0	0	2	1	3	1	0	2	4	4
	d	0	1	0	0	0	0	8	83	7	0	1	1	0	0	1	3	3
	m	0	0	0	0	0	1	0	0	0	65	5	3	14	1	6	7	6
	r	0	1	0	0	0	0	1	0	0	1	2	41	34	17	1	8	2
	l	0	0	0	0	0	0	0	0	0	8	1	0	83	10	1	2	3
–5 dB	p	87	4	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1
	t	4	91	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	k	2	0	89	0	0	0	0	0	0	0	0	0	1	0	0	2	2
	s	0	0	0	96	0	0	0	0	0	0	0	0	0	0	0	0	0
	h	0	1	0	0	94	0	0	0	0	0	0	0	0	0	0	0	0
	b	0	0	0	0	0	1	82	0	0	1	0	1	1	4	1	1	4
	d	0	0	0	0	0	0	1	91	4	0	0	0	0	0	0	0	0
	m	0	0	0	0	0	0	2	0	0	67	5	6	9	2	2	0	3
	r	0	0	1	0	0	0	0	0	0	0	0	74	14	6	1	0	0
	l	0	0	0	0	0	0	0	0	0	3	0	1	84	7	0	0	1

Results

The response data are presented as confusion matrices in tables 2–4 by S/N ratio and segment type (onset, vowel, and coda), pooled across participants. Because participants were not limited to the stimulus set in their responses, these confusion

Table 3. Confusion matrices for vowels

		a	æ	aɪ	ɛ	e	o	ɪ	i	u	ʌ	Null	Other
–14 dB	a	227	17	13	0	1	4	1	0	2	11	4	6
	æ	26	68	0	6	4	0	2	1	0	2	1	0
	aɪ	7	2	92	4	5	0	2	3	1	4	5	7
	ɛ	11	7	2	57	17	8	1	1	1	24	1	2
	e	2	5	4	2	91	12	13	1	0	1	1	0
	o	9	5	11	7	45	34	10	0	0	5	3	3
	ɪ	1	2	0	0	22	1	87	1	2	3	1	1
	i	1	1	1	4	1	0	11	94	20	2	4	4
	u	0	0	2	3	6	0	13	60	41	0	3	4
–11 dB	a	213	7	2	0	0	2	0	0	2	2	4	2
	æ	9	72	0	4	2	0	0	0	0	0	2	1
	aɪ	2	0	84	2	2	1	11	0	0	2	1	3
	ɛ	6	3	1	67	14	0	1	0	1	14	0	1
	e	4	12	3	3	76	1	2	2	0	0	5	0
	o	6	3	2	5	31	42	4	3	1	3	4	4
	ɪ	0	2	0	6	10	0	77	0	0	2	2	0
	i	0	0	0	7	2	0	10	93	2	0	1	2
	u	3	2	0	4	5	2	12	42	27	2	2	7
–8 dB	a	210	5	3	0	0	6	0	0	1	4	3	2
	æ	10	75	0	2	0	0	1	0	0	0	2	0
	aɪ	0	0	96	0	1	0	4	4	0	0	1	2
	ɛ	2	1	0	93	3	0	1	1	0	3	3	1
	e	1	3	8	2	83	2	2	0	0	1	6	0
	o	12	0	2	1	19	62	0	1	1	5	2	3
	ɪ	0	0	0	1	4	0	89	1	0	0	4	0
	i	1	0	0	6	0	0	1	99	5	1	4	0
	u	0	0	0	2	3	1	2	34	49	0	8	9
–5 dB	a	199	5	1	0	7	1	0	0	0	0	1	2
	æ	10	67	0	1	2	0	0	0	0	0	0	0
	aɪ	0	0	87	0	0	0	5	4	0	0	0	0
	ɛ	0	0	1	88	0	0	4	2	0	0	1	0
	e	2	1	1	3	79	0	1	1	0	0	0	0
	o	5	0	0	0	8	71	0	0	3	7	1	1
	ɪ	0	0	2	1	1	0	83	0	0	0	1	0
	i	0	0	1	4	0	0	1	96	1	1	0	0
	u	1	0	2	0	0	1	5	6	70	4	0	7

matrices are not square. A column was added for each phoneme that garnered 5% or more of the responses to any single stimulus phoneme. The consonants /f ɡ n w j/ were added to the onset matrices, /ʌ/ was added to the vowel matrices, and /f b v r/ were added to the coda matrices. Any other phonemes that received fewer than 5% of responses to any single stimulus phoneme are lumped into the ‘other’ column. The cluster responses are also included in the ‘other’ totals. Deletion misperceptions are totaled in the ‘null’ column.

Table 4. Confusion matrices for codas

		p	t	k	s	f	b	d	g	z	v	m	n	r	l	Null	Other
–14 dB	p	58	49	21	0	0	1	0	2	0	0	0	0	0	0	0	1
	t	49	60	18	0	0	1	0	0	0	0	0	1	0	0	2	1
	k	65	24	33	0	1	0	2	4	0	0	0	0	0	1	0	2
	s	0	3	1	64	19	0	0	1	9	1	0	0	0	1	5	28
	d	6	4	2	0	1	10	44	15	0	12	0	8	7	4	15	4
	g	5	3	1	1	0	8	26	28	1	26	1	5	0	1	14	12
	z	1	1	0	6	0	1	4	2	51	47	0	0	0	0	8	11
	m	0	3	0	0	1	0	6	4	2	19	5	31	3	4	51	14
	n	0	1	0	0	0	1	7	6	1	10	8	37	1	7	33	9
	l	2	4	0	0	0	1	5	2	0	5	1	2	14	65	23	8
–11 dB	p	50	39	12	0	2	0	0	0	0	0	0	0	0	0	3	2
	t	37	53	10	0	1	0	0	0	2	0	0	0	0	0	3	2
	k	44	24	32	0	0	0	0	1	0	0	0	0	0	0	2	5
	s	0	8	2	60	8	0	0	1	15	0	0	0	0	0	3	11
	d	5	3	2	0	0	11	46	7	0	4	0	2	1	1	19	7
	g	2	0	1	0	0	13	27	14	2	13	4	11	0	1	13	7
	z	1	0	0	0	1	0	0	5	71	19	0	0	0	0	5	6
	m	1	1	0	0	0	1	4	1	3	3	6	37	0	7	42	11
	n	0	1	0	0	0	0	2	1	1	3	13	35	0	0	30	13
	l	0	0	1	0	0	0	5	0	0	1	2	2	6	63	14	14
–8 dB	p	78	14	8	0	0	0	0	1	0	0	0	0	0	0	5	2
	t	25	71	6	0	0	0	0	0	0	0	0	1	0	0	4	1
	k	40	18	44	0	0	0	0	2	0	0	0	0	0	0	2	2
	s	0	0	1	77	2	0	0	0	14	0	0	0	0	0	3	11
	d	1	4	0	0	0	8	59	10	1	1	1	0	1	0	14	8
	g	1	2	1	0	0	16	27	22	0	8	0	2	0	0	14	15
	z	0	0	0	5	0	1	0	1	89	4	0	0	0	0	4	4
	m	0	0	0	0	0	1	2	1	1	2	17	44	1	1	39	8
	n	0	0	0	0	0	0	2	2	0	4	4	44	0	0	24	19
	l	0	1	0	0	0	0	1	1	0	1	1	2	4	84	7	6
–5 dB	p	71	14	7	0	0	0	0	0	0	0	0	0	0	0	2	2
	t	4	91	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	k	22	7	66	0	0	0	0	0	0	0	0	0	0	0	0	1
	s	0	0	0	67	0	0	0	0	20	0	0	0	0	0	0	5
	d	0	2	0	0	0	2	79	1	0	0	0	0	0	0	5	7
	g	0	0	1	0	0	9	23	40	0	1	2	7	0	0	2	11
	z	0	0	0	9	0	0	0	0	90	0	0	0	0	0	1	0
	m	1	0	0	0	0	0	1	0	0	0	33	48	0	2	14	5
	n	1	0	0	0	0	0	0	0	0	0	2	63	0	0	11	11
	l	0	0	0	0	0	0	1	0	0	0	0	0	2	91	2	0

Presentation of these results in the form of confusion matrices assumes independent recognition of segments, or at least that there were no systematic CV, VC, or even CVC effects. This may not be a valid assumption. For example, a high first formant (F_1) of vowels is known to function as a cue for vowel height but also voicelessness in a following stop [Whalen, 1989; Nearey, 1990, 1997]. However, the whole-syllable proportion correct scores of the present data are well predicted by the proportion correct of the segments under an assumption of independent segmental recognition as reported by Benkí [2003] using a j-factor analysis [Boothroyd and Nittrouer, 1988].

Assuming segmental independence, then, the data in the confusion matrices were analyzed in three ways. First, proportion of information transmitted was computed for segments and phonological features and is plotted in figures 1–3. Second, the degree of asymmetry between confusable phonemes was evaluated. Finally, the extent to which listeners classified phonemes using responses not present in the set of stimulus phonemes, particularly null responses to coda consonants, is presented in a Pearson chi-square test of association.

Intelligibility Measures

Information theoretic measures of intelligibility [Shannon, 1948; Cover and Thomas, 1991] were used to quantify performance of the listeners for segments and phonological features, following Miller and Nicely [1955]. These measures are useful for comparing the robustness of contrasts that differ in the number of possible values, and have been used by other studies of human speech perception [e.g., van Son and Pols, 1999; Smits, 2000]. The error rate and relative proportion of information transmitted is presented for segments and selected phonological features. Additionally, the conditional entropy of the stimulus, a measure of the amount of confusion present in the stimulus, is presented for segments.

These measures can be calculated as follows. First, the amount of information transmitted T is given by

$$T = -\sum_{i,j} p_{ij} \log_2 \frac{p_i\, p_j}{p_{ij}}$$

in which p_i is the probability of input i, p_j is the probability of response j, and p_{ij} is the probability of input i and response j together. For n observations in a given confusion matrix, the row sums of the confusion matrix divided by n give the input probabilities, the column sums divided by n give the response probabilities, and the individual cell entries divided by n give the probability of input i and response j together.

Miller and Nicely [1955] normalize T by the stimulus entropy H(x) to yield T_{rel}, the relative proportion of information transmitted, where

$$H(x) = -\sum_{i} p_i \log_2 (p_i)$$

is the maximum number of bits per stimulus that could be transmitted. This normalization is appropriate for a square confusion matrix in which each observed response category corresponds to a stimulus category, and the stimulus entropy is therefore (approximately) equal to the response entropy. However, the open-set response format of the present study resulted in more response categories than stimulus categories as indicated by the confusion matrices in tables 2–4, with 5 nonstimulus consonants garnering 5% or more responses for any one stimulus (including ‘null’ for codas but not for onsets), and /ʌ/ garnering a number of vowel responses. Normalizing by H(x) would therefore overestimate T_{rel}, so T was normalized by the more appropriate response entropy

$$H(x) = -\sum_{j} p_j \log_2 (p_j).$$

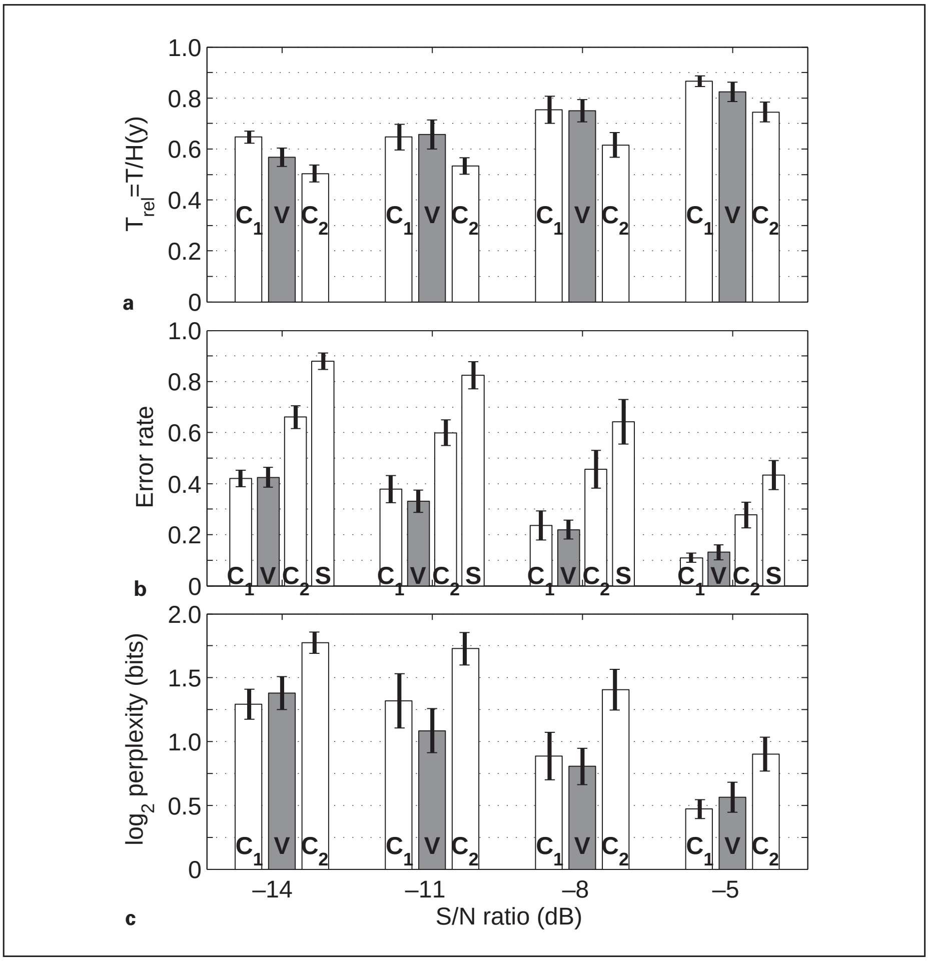

Fig. 1. Relative proportion of information, error rate, and $\log_2$ perplexity by segment type as a function of S/N ratio. In this and in subsequent plots, the values are averages over the participants at each S/N ratio, and error bars represent 95% confidence intervals. **a** Relative proportion of information per segment, $T_{rel} = T/H(y)$, for onsets (C_1), vowels (V, shaded), and codas (C_2). **b** Error rates for onsets (C_1), vowels (V, shaded), codas (C_2), and whole syllables (S). **c** $\log_2$ perplexities for onsets (C_1), vowels (V, shaded), and codas (C_2).

Averaged over S/N ratio, the response entropies are 3.62 bits (standard deviation SD = 0.08) for onsets, 3.18 bits (SD = 0.02) for vowels, and 3.62 bits (SD = 0.07) for codas. These response entropies are slightly higher than the stimulus entropy H(x) of 3.32 bits for both onsets and codas, and 3.10 bits for vowels. The proportion of information transmitted per phoneme, T_{rel}, is plotted as a function of S/N ratio in figure 1a for onsets, vowels, and codas, averaged across participants. Error bars represent 95% confidence intervals, $\pm 1.96(SD/\sqrt{N})$ with standard deviation SD for *N* participants, for this and subsequent plots.

Table 5. Feature analysis used for computing information entropies

Feature	Values	Members
Voicing, consonants	voiced	/b d ɡ z v m n ŋ r l w j ð ʒ dʒ/
	voiceless	/p t k s f θ h ʃ tʃ/
Place of articulation, consonants	labial	/b p m v f/
	coronal	/d t n s z θ ð r l/
	dorsal	/ɡ k ŋ ʃ ʒ tʃ dʒ/
Manner of articulation, consonants	stop/affricate	/p b d t k ɡ tʃ dʒ/
	fricative	/v f d ð t θ s z ʃ ʒ h/
	sonorant	/m n ŋ w j r 1/
Height, vowels	high	/i ɪ u ʊ/
	mid	/e ɛ o ʌ/
	low	/a æ/
Backing, vowels	front	/i ɪ e ɛ æ/
	back	/u ʊ o ʌ a/

The conditional entropy of the stimulus $H_y(x) = H(x) - T$, called the *log_2 perplexity* by van Son and Pols [1999] and *equivocation* by Shannon [1948], is also a useful measure of the efficiency of an information channel. As the complement of the amount of information transmitted, the $\log_2$ perplexity is correlated with the error rate and quantifies the confusion of the signal in terms of bits per stimulus, i.e., how much information present in the stimulus is missing in the received signal. It is directly related to the perplexity S(x), widely used in evaluating automatic speech recognition systems [e.g., Bahl et al., 1983], as $S(x) = 2^{Hy\,(x)}$. The error rates and $\log_2$ perplexities for each type of segment are plotted in figure 1c, along with the whole syllable error rate.

As shown in figure 1a by T_{rel}, the average onset consonant contains as much or greater information than the average vowel, and both contain greater information than the average coda consonant for all S/N ratios. In particular, onsets carry information at a rate that is 15% points higher than codas for the range of S/N ratios in this study. The error rate and $\log_2$ perplexity displayed in figure 1c are correlated with each other and inversely correlated with T_{rel}, as expected.

The robustness in noise of phonological features can also be compared using the metric of proportion of information. Height and backing are compared for vowels, while voicing, place of articulation, and manner of articulation are compared for consonants. A simplified version of Miller and Nicely's [1955] feature analysis for consonantal place and manner of articulation was adopted with three places of articulation (labial, coronal, and dorsal) and three manners of articulation (plosive/affricate, fricative, and sonorant), as shown in table 5. Confusion matrices were derived for each of these features from the overall confusion matrices, omitting all trials with null or cluster responses from consideration, and trials with /aɪ/ as a response or stimulus. Trials with /h/ as a response or stimulus were excluded from the place of articulation analyses as well.

Relative proportion of information transmitted was calculated from the feature confusion matrices following the same method as described above for segments, with $T_{rel} = T/H(y)$, normalizing by the response entropy instead of the stimulus entropy for

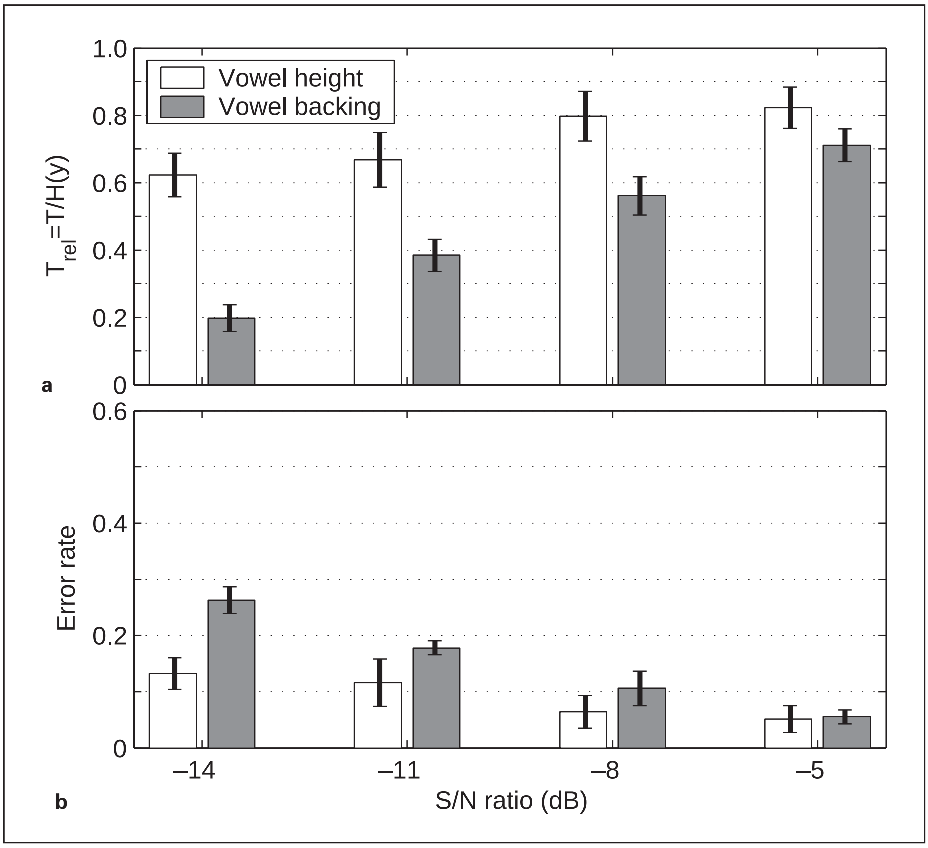

Fig. 2. Relative proportion of information and error rate for vowel height and backing features as a function of S/N ratio. **a** Relative proportion of information per feature, T_{rel} = T/H(y), for height and backing (shaded). **b** Corresponding error rates.

consistency with the segment results. Averaged over S/N ratio, the response entropies H(y) are 1.58 bits (SD = 0.003) for vowel height, 0.96 bits (SD = 0.02) for vowel backing, 1.36 bits (SD = 0.01) for onset place of articulation, 1.26 bits (SD = 0.09) for coda place of articulation, 1.0 bits (SD = 0.0003) for onset voicing, 0.99 bits (SD = 0.008) for coda voicing, 1.49 bits (SD = 0.006) for onset manner of articulation, and 1.46 bits (SD = 0.03) for coda manner of articulation. Because the feature confusion matrices are square, these response entropies are fairly close to the stimulus entropies H(x) of 1.58 bits for vowel height, 0.99 bits for vowel backing, 1.34 bits for onset place of articulation, 1.38 bits for coda place of articulation, 1.0 bits for onset voicing, 0.99 bits for coda voicing, 1.49 bits for onset manner of articulation, and 1.45 bits for coda manner of articulation. Figure 2 shows proportion of information transmitted for the vowel features and figure 3 shows the consonant results.

Vowel height is more robust than backing over the range of S/N ratios used in this experiment. For vowel height, T_{rel} rises from 0.6 to 0.8, while for backing T_{rel} rises from just under 0.2 to 0.7. The consonant results are largely consistent with Miller and

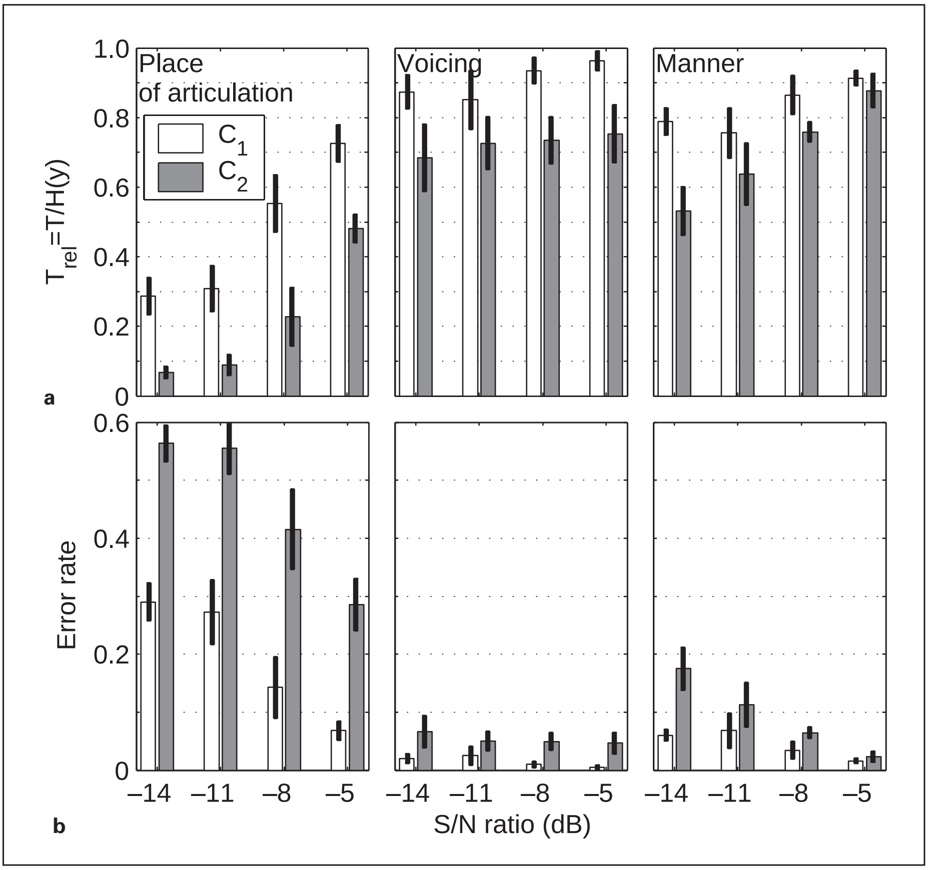

Fig. 3. Relative proportion of information and error rates for onset (C_1) and coda (C_2, shaded) voicing, place of articulation, and manner of articulation features as a function of S/N ratio. **a** Relative proportion of information per feature, T_{rel} = T/H(y). **b** Corresponding error rates.

Nicely [1955] and Wang and Bilger [1973]. Voicing in consonants is fairly robust and is only slightly affected by noise in onset position for the S/N ratios used here. Manner of articulation is slightly less robust than voicing. Place of articulation, on the other hand, is severely affected by noise, with T_{rel} for place being about 60% points lower than for voicing at –14 dB in onsets and about 20% points lower at –5 dB. Both voicing and place are more affected by noise in codas than in onsets, with T_{rel} in codas being between 10 and 20% points lower than the corresponding value in onsets at each S/N ratio.

The slightly higher T_{rel} measures for voicing than for manner of articulation should be viewed as a tentative result, given the simplified feature analysis used here and the uneven distribution of the different manners of articulation in the stimuli. Note that the confidence intervals for voicing and manner of articulation overlap for both onsets and codas.

Relative error rates for features (fig. 2b, 3b) are consistent with T_{rel}, with an inverse correlation between the two for all of the features. However, T_{rel} is a better quantitative measure for purposes of comparing different contrasts, as it presents the amount of information transmitted normalized by the amount of information potentially present in the response. For example, the vowel backing error rate for vowel backing is only slightly higher than that of height. However, the error rate does not take into account the fact that backing is only a binary contrast while height is a ternary contrast (in these data). The T_{rel} metric takes this into account, and correctly diagnoses vowel height as more robust in noise than backing. For consonants, T_{rel} diagnoses the binary contrast of voicing as more robust than the ternary contrast of place of articulation.

Asymmetrical Confusions

As might be inferred from the feature analysis, phonemes of certain phonological natural classes (e.g., voiceless stops, voiced stops, high vowels, mid vowels, etc.) are confusable. Among the confusable phonemes, asymmetries exist such that certain phonemes are substantially favored over others. This pattern can be observed from the confusion matrices, and can be quantified with the signal detection theory (SDT) [Macmillan and Creelman, 1990] bias measure c (criterion).

In order to calculate c, the constant ratio rule [Clarke, 1957] was assumed in order to extract 2 × 2 confusion matrices from the overall confusion matrices. The constant ratio rule says that response frequency ratios are not affected by the number of stimuli. Because this is a rather strong assumption and given the relatively small scale of the present study, the numerical values and statistical tests of c presented below should be viewed as tentative, but they do provide a useful assessment of the presence and direction of bias between phonemes.

Unlike the more commonly reported SDT bias measure β, the bias measure c is statistically independent of measures of sensitivity such as d′, and is given by

$$c = -0.5[z(H) + z(F)]$$

using z cores of hit and false alarm rates extracted from the main confusion matrix. For the 2 × 2 confusion matrix of two phonemes x and y, the diagonal entries are the response frequencies for x and y from the diagonal of the original confusion matrix, which become the hits and correct rejections. The response frequency for x given stimulus y from the original matrix becomes the miss frequency for the 2 × 2 matrix, and the response frequency for y given stimulus x becomes the false alarms. The marginal totals from the resulting 2 × 2 confusion matrix are used to compute the hit rate H and false alarm rate F.

Error rates are low enough at –5 dB that ceiling effects make any bias measures suspect, so c was calculated for confusable pairs of phonemes for the response frequencies from the –14, –11, and –8 dB confusion matrices. All pairs of stimulus phonemes for which at least 10% of one were categorized as the other were included in this analysis, pooling over all participants in the lowest three S/N ratios. For the most part, this criterion selected pairs of consonants with the same voicing and manner of articulation but different places of articulation, and vowels of the same height but different backing. For these pairs of phonemes, table 6 reports c averaged over the three S/N ratios with standard deviations, and the p values of a t test of whether the average c significantly differs from zero. The final column, lexical errors, has counts of

Table 6. SDT bias values of c for pairs of confusable phonemes, averaged over the lowest three S/N ratios with standard deviations

	Pair	Average c	SD	p	Lexical errors	
Onsets	t p	0.3138	0.2074	0.16	t: 0	p: 32
	k p*	0.2756	0.0611	0.02	k: 1	p: 20
	r l*	1.0710	0.1354	<0.01	r: 8	l: 0
Vowels	æ a*	0.3872	0.0523	<0.01	æ: 13	a: 33
	u i*	0.8191	0.2430	0.04	u: 1	i: 31
	o e*	0.7778	0.1715	0.02	o: 6	e: 28
Codas	t p	0.0491	0.1019	0.56	t: 72	p: 5
	k p*	0.5623	0.0490	<0.01	k: 24	p: 13
	k t*	0.3706	0.0728	0.02	k: 2	t: 33
	ɡ d	0.5544	0.1878	0.05	ɡ: 15	d: 23
	m n*	0.9453	0.0704	<0.01	m: 4	n: 60

Negative values of c indicate a bias in favor of the first phoneme of the pair, while positive values indicate a bias in favor of the second phoneme of the pair, and zero indicates no bias. The p values indicate whether the average c significantly differs from zero in a two-tailed t test. Statistically significant ($p < 0.05$) favored phoneme pairs are starred (*). The final column, lexical errors, provides the number of incorrect responses of a particular phoneme that result in a meaningful word, given a stimulus containing the other member of the pair.

how many of the incorrect responses resulted in meaningful words, and is discussed below.

Among the voiceless stops /p t k/, /p/ is favored over /k/ in onset position, i.e., /k/ is classified as /p/ more often than /p/ is classified as /k/. In coda position, /k/ is disfavored relative to both /t/ and /p/. For the codas /d ɡ/ there is a statistically nonsignificant trend favoring /d/ over /ɡ/. Among the sonorant consonants, /l/ is favored over /r/ in onset position, and in coda position, /n/ is favored over /m/.

There is a high rate of confusions between front and back vowels of the same height. The asymmetry in nonlow vowels here, shown by the biases favoring /i/ and /e/, is that the back tense vowels /u/ and /o/ are more of then classified as /i/and /e/, respectively, than the reverse. The opposite pattern appears for the low vowels /a/ and /æ/, for which the back vowel /a/ is favored quite strongly over the front vowel /æ/. The pattern among the vowels strongly suggests an acoustic-phonetic explanation involving asymmetries of F_2, discussed in section 4.

Alternatively, these asymmetries may arise from the confusable phonemes having different relative frequencies or being contained by different numbers of entries in the mental lexicons of the listeners [e.g., Stemberger, 1992]. A frequency or lexical-based account would be supported by the difference in relative frequency or lexical representation between a pair of confusable phonemes being correlated with the observed bias c. Two databases were used to quantify differences in relative frequency and lexical representation: an online version of *Webster's Seventh Collegiate Dictionary* [Nusbaum et al., 1984] with the Kucera and Francis [1967] frequency of usage data (word occurrence per million words), and French's et al. [1930; summarized by Fletcher, 1953, p. 95] report of phoneme occurrences per 100 phonemes with the consonants coded for syllable position, based on an 80,000-word telephone conversation

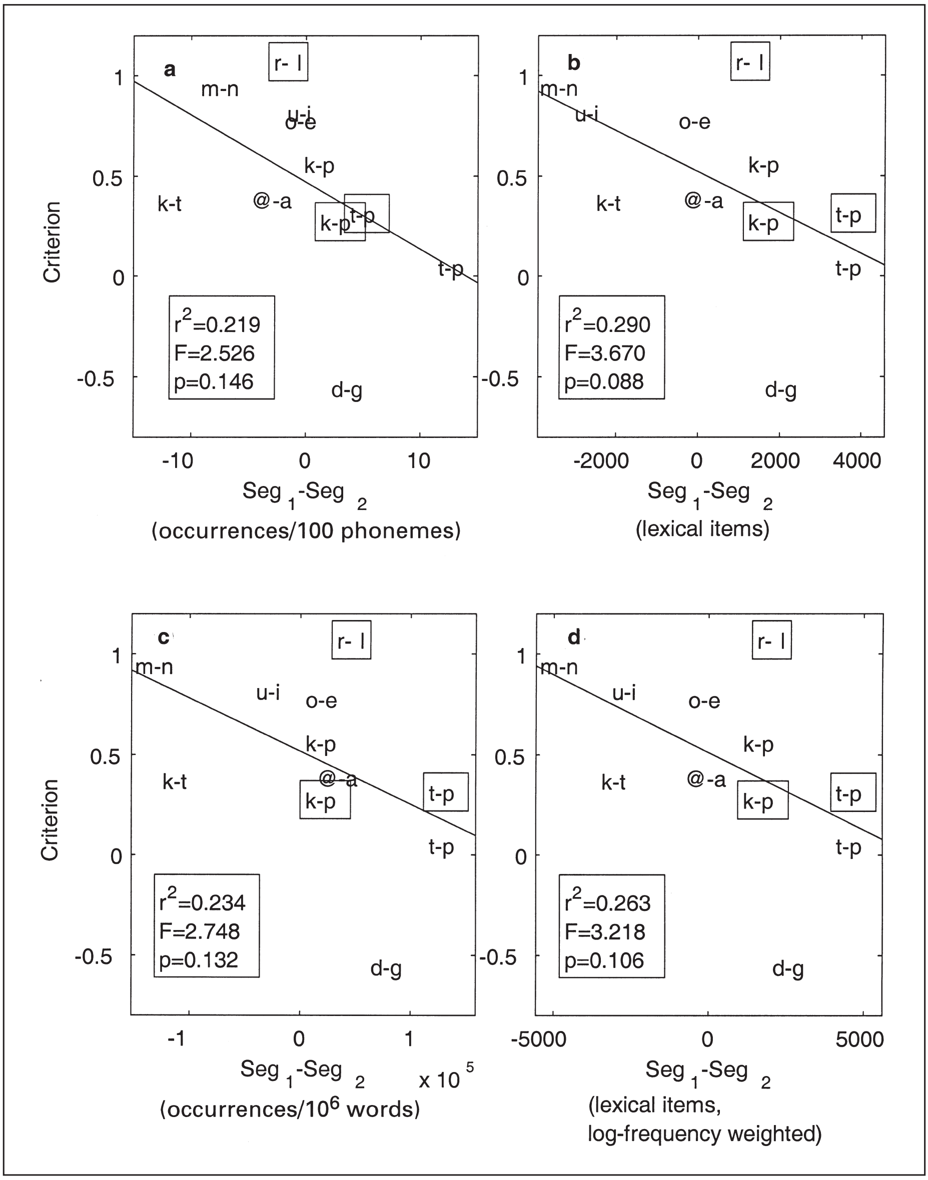

Fig. 4. Criterion c as a function of frequency and lexicality difference measures for confusable pairs of phonemes. Each panel displays c of the phoneme pairs in table 6 as a function of the difference between the segmental frequency measures of the members of the pair. Panel **a** uses phoneme frequency data from French et al. [1930], while panels **b–d** use lexical and phoneme frequency data from Kucera and Francis [1967] and Nusbaum et al. [1984]. See the text for how the frequency measures were computed. Each panel contains the line of best fit from a linear regression, along with r^2, F, and p statistics. Coda pairs are enclosed by boxes, and '@' is used as a label for the vowel /æ/.

Table 7. Contingency table of the frequencies of null responses for coda consonants

Coda consonant	All other responses	Null responses	Expected null responses	Standardized residuals
p	340	8	40.1	–5.069
t	339	9	40.1	–4.911
k	344	4	40.1	–5.701
s	337	11	40.1	–4.595
z	331	17	40.1	–3.648
d	300	48	40.1	1.248
ɡ	307	41	40.1	0.142
m	245	132	43.4	13.436
n	232	87	36.8	8.287
l	304	44	40.1	0.616

Column 1 contains the total number of responses other than ‘null’ for each consonant, pooled across all participants at the –14 dB, –11 dB, and –8 dB S/N ratios. Column 2 contains the corresponding total ‘null’ responses for each coda consonant. Column 3 shows the estimated expected null response frequencies assuming nonassociation between null response rate and coda consonant, and column 4 shows the standardized residuals. A Pearson chi-square test of association ($X^2_P = 415.8$, d.f. = 9, $p < 0.001$) shows a significant association between coda consonant and null response rate. For familywise $\alpha = 0.01$ for 10 tests, a significant deviation from the expected null response rate is indicated for standardized residuals $|z| \geq 3.2905$.

corpus. Four measures of relative frequency were computed for each phoneme in table 6, with the difference between members of each pair used as independent variables for linear regression analyses. These measures were (a) number of occurrences per 100 phonemes [French et al., 1930], (b) number of lexical items containing the phoneme [Nusbaum et al., 1984], (c) number of occurrences per million words [Kucera and Francis, 1967; Nusbaum et al., 1984], and (d) the sum of the log-transformed frequencies of the lexical items containing the phoneme [Kucera and Francis, 1967; Nusbaum et al., 1984], Measure (d) is a log-transformed version of measure (b).

For each pair in table 6, the difference between segments on measures (a)–(d) was used as the independent variable in a linear regression with bias c as the dependent variable, and plotted in figure 4. As shown by the similarity in the ordering of the points along the abscissa of each plot, the difference measures are correlated with each other, indicating a high degree of correspondence between frequency of occurrence and lexicality as measured by number of lexical items containing a phoneme. However, only weak support at best for a phoneme or lexical frequency-based account for the asymmetries is offered by linear regression analyses. While the r^2 values indicate that some variance is explained, only the difference of measure (b), number of lexical items containing the phoneme, reached significance at the $p < 0.10$ level ($r^2 = 0.29$, $F = 3.67$, $p = 0.09$). While this analysis supports further investigation of lexicality as a factor in producing phoneme biases in perception, lexicality does not appear to be the only factor, particularly for vowels. Furthermore, the small sample size of the present data prevents strong confirmation or disconfirmation of frequency or lexical-based accounts of phoneme bias.

Another prediction of the lexical frequency-based account of phoneme bias is that meaningful words should show up as errors consistent with the phoneme biases. For example, given nonsense stimulus /raɪɡ/, a listener might report /raɪd/ ‘ride’, con-

Table 8. Response rates for nonstimulus responses for selected phonemes, pooled over the –14 dB, –11 dB, and –8 dB confusion matrices

	Stimulus phoneme	Response phoneme	Response rate
Onsets	r	w	0.18
	m	n	0.10
	h	f	0.14
Vowels	ɛ	ʌ	0.12
Codas	m	v	0.06
	z	v	0.20
	l	r	0.07

sistent with the phoneme bias. Using the same online version of *Webster's Seventh Collegiate Dictionay* [Nusbaum et al., 1984] as an approximation of the participants' lexicons, 12% of all responses, or 18% of all errors (pooling over all participants), were meaningful words fitting this pattern. The lexical errors column in table 6 lists the number of such lexical errors for each member of the phoneme pairs. For example, the lexical error entry for the first pair, onsets /t p/, is t:0 and p:32, indicating that there were no /t/ responses to stimulus /p/ that resulted in meaningful words, but 32 /p/ responses to stimulus /t/ that resulted in meaningful words. Like the linear regressions in figure 4, the lexical error data neither confirm nor reject lexical bias as a source of phoneme biases. The lexical errors are consistent with the difference in lexical items containing each phoneme (fig. 4b) for only 6 of the 11 phoneme pairs (onset pair /k p/, vowel pair /u i/, coda pairs /t p/, /k t/, /g d/, and /m n/). Again, as the data sample has become quite small at this point, stronger conclusions will not be put forth.

Responses That Are Not Present in the Stimulus Categories

Although participants were informed that all the syllables would consist of a CVC sequence, the open-set response nature of the task allowed them to provide responses that included phonemes that were not possible stimuli, including null responses for coda consonants. Null responses for onsets were negligible. The null response rates for the coda consonants, pooled over the confusion matrices at –14 dB, –11 dB, and –8 dB, are evaluated by a Pearson chi-square test of association in table 7.

There is a significant association between coda consonant type and null response rate (X_P^2 = 415.8, d.f. = 9, p < 0.001). Standardized residuals indicate which consonants were more or less likely to be classified as null. Nasals /mn/ were significantly more likely to be transcribed as null than the lateral and the voiced stops /l d g/ at the familywise $\alpha = 0.01$ level. Note that because there were 13 /m/ codas and 11 /n/ codas (instead of the 12 for the other coda segments, table 1), the expected null response rates are slightly different for these response categories. Fricatives and voiceless stops /s z p t k/ were significantly less likely than the liquid and the voiced stops /l d g/ to be transcribed as null, also at the familywise $\alpha = 0.01$ level.

A number of nonstimulus response phonemes garnered significant numbers of responses. Table 8 shows the nonstimulus response rates pooled from the –14 dB, –11 dB, and –8 dB confusion matrices for selected phonemes. The confusion rates shown in table 8 are consistent with the feature analysis above, showing consonantal place of articulation confusions within a natural class, and vowel backing confusions. An exception is the bilabial nasal coda /m/ classified as the voiced labiodental fricative /v/.

Summary of Results

To summarize, consonantal onsets carry information at a per phoneme rate T_{rel} that is about 15% points higher than codas across the range of S/N ratios in this experiment, and vowels carry information at rates slightly lower than consonantal onsets. For consonants, the binary contrast of voicing is much more robust in noise than the ternary contrast of place of articulation, with T_{rel} for voicing exceeding that for place of articulation by 42% points in onsets and 54% points in codas, averaged across all S/N ratios. For vowels, the binary backing contrast is less robust in noise than the ternary height contrast, with T_{rel} for height exceeding that for backing by 26% points, averaged across all S/N ratios.

Among certain confusable phonemes, listeners displayed a bias toward certain members of different pairs. Listeners classified /k/ more often as /p/ than the reverse in both coda and onset position, and /k/ as /t/ more often than /t/ as /k/ in coda position. Similarly, between onsets /l/ and /r/, listeners were biased toward /l/, and between codas /m/ and /n/ were biased toward /n/.

Among the vowels, listeners more frequently classified the nonlow tense back rounded vowels /u o/ as their front unrounded counterparts /i e/ than the reverse, but for the confusable low vowels /a/ and /æ/, the back vowel /a/ was favored over its front counterpart.

The coda consonants fell into three groups with respect to likelihood of being classified as null. The nasals /m n/ were significantly more likely than the liquid and voiced stops /l d g/ to be perceived as absent, which were in turn significantly more likely than the fricatives and voiceless stops /z s p t k/ to be perceived as absent.

Discussion

As mentioned at the beginning of this paper, patterns of perceptual confusions can inform our understanding of phonology in both synchronic and diachronic terms. Hume and Johnson [2001] discuss three ways in which the nature of human speech perception constrains phonology. First, listeners may be the source of sound change, as in Ohala's [1981] proposal of hypo- and hypercorrection for coarticulation. Second, all other things being equal, languages favor robust contrasts over contrasts that are difficult to perceive. Third, as proposed by Steriade [2001], the elements within a phonological alternation must satisfy some constraints on perceptual similarity.

In this section, the results are discussed and applied as phonetic explanations for attested phonological phenomena, in particular along the three lines pointed out by Hume and Johnson [2001]. While the definitiveness of the claims is constrained by the small size of the present study, it should be noted that the results are fairly clear and are consistent with previous studies insofar as comparisons are possible.

Advantage of Height over Backing in Vowels in Noise

The present finding of greater perceptual robustness of vowel height over backing is consistent with previous studies showing that vowel height is more robust than backing in white noise, in /b V b/ context and in Harvard PB words for English listeners [Pickett, 1957] and in isolation for Dutch listeners [Nooteboom, 1968; Lindblom, 1975/1979]. This advantage of height over backing is likely to arise from both acoustic and auditory factors. The contrast between high, mid, and low vowels is primarily conveyed by differences in the frequency of F_1 and concomitant bandwidth differences, while the backing contrast is primarily conveyed by differences in the frequency of F_2

and in part by F_3. F_1 is more intense than the other formants and is therefore the major factor in determining a vowel's overall intensity, which is inversely correlated with the frequency of F_1 [Fant, 1956, 1960]. As a result, the greater intensity of F_1 makes it more salient to listeners than the other formants. Furthermore, listeners can judge vowel height not only on the basis of the frequency of F_1 but on the overall intensity of the vowel as well, providing additional redundancy in the presence of noise [Lindblom, 1975/1979].

There is a possible auditory basis for the advantage of height over backing as well. As shown by seminal experiments by Young and Sachs [1979], spectral peaks such as formants induce neural synchrony, or phase-locking, in a wide tonotopic span of auditory nerve fibers. This neural synchrony is broader for F_1 in particular (and frequency peaks below 1 kHz in general) than for the higher formants, and has been replicated closely in models of the auditory periphery [Jenison et al., 1991]. The spread of neural synchrony to auditory nerve fibers with higher CFs effectively suppresses synchrony to energy in spectral valleys, providing enhanced redundancy and resulting sensitivity for the main spectral peak (i.e., F_1) in the face of noise [Kiang and Moxon, 1974; Greenberg, 1988].

The phonological consequences are both diachronic and synchronic. Under the *sociophonetic adaptation hypothesis* [Lindblom et al., 1995], that sound change is an adaptation of a communicative system to social, perceptual, and articulatory-acoustic constraints, the world's languages should show a bias in favor of F_1-based contrasts in vowels, consistent with Hume and Johnson's [2001] point that robust contrasts are favored over less reliable ones. This prediction is borne out: crosslinguistic studies of vowel inventories have observed that the number of height contrasts is equal to or greater than the number of front-back contrasts in any given language [Crothers, 1978]. The early version of adaptive dispersion [Liljencrantz and Lindblom, 1972] has been improved by subsequent versions that give a greater weight to F_1-based contrast [Lindblom, 1986].

One interpretation of the advantage of height over backing is that vowels differing only in backing are perceptually more similar than vowels differing only in height. With Steriade's [2001] proposal that phonological neutralizations are subject to constraints of perceptual similarity between underlying and surface representations, a prediction of this interpretation is that backing alternations among vowels should be represented in the world's languages. Turkish is one well-known example [e.g., Kenstowicz and Kisseberth, 1979], with suffixal vowels harmonizing to root vowels in backing (and rounding). K'ichee', a Mayan language spoken in Westem Guatemala, is another, with the vowels /i u/ alternating in a number of morphemes [Barrett, 1999].

On the other hand, there are many examples of height alternations, such as Polish [Steele, 1973; Steiber, 1973] and Yawelmani Yokuts (which shows both height and backing alternations) [Newman, 1944]. However, examples and counterexamples, used in an inductive fashion, are insufficient to settle the point, since a variety of factors are at play in the case of any one language. Eventually, it will be up to a comprehensive theory, using such principles as the perceptual salience of height over backing – in a deductive fashion – to predict crosslinguistic phonological patterns.

Consistent with the examples of languages with height alternations, studies of vowel identification without noise maskers have found height errors to be as frequent [English listeners: Fairbanks and Grubb, 1961] or more frequent [Dutch listeners:

Klein et al., 1970] than backing errors. However, the results of these studies are consistent with the interpretation of the present results and other studies of vowels in noise, given that the neural synchrony and acoustic advantages of F_1 are most present in the presence of noise. Indeed, the methodology of the present study, open-response identification of stimuli in noise, in many ways is a more realistic assessment of normal communicative situations [Lindblom, 1975; Jenison et al., 1991] than forced-choice identification of stimuli in quiet.

Vowel Fronting

The pattern of asymmetrical confusions of the vowels (/i e a/ favored over /u o æ/) bears a striking resemblance to well-represented patterns of sound change in vowel backing, first observed by Sweet [1888], and reviewed by a number of subsequent studies [Labov et al., 1972; Labov, 1994; Gordon and Heath, 1998; Cox, 1999]. Labov [1994, p. 116] restates Sweet's [1888] original proposals as three axiomatic principles regarding vowel chain shifts, the third of which is that back vowels move to the front, later revised as 'tense vowels move to the front along peripheral paths, and lax vowels move to the back along nonperipheral paths [Labov, 1994, p. 200]. The direction of vowel movement described by Labov's third vowel chain shift principle is precisely the asymmetrical pattern of confusions observed in the high and mid vowels of the present study: the tense back vowel phonemes /u o/ were often classified as their front counterparts of the same height /i e/, respectively. The same asymmetry favoring /i e/ over /u o/ is explicitly noted by Pickett [1957, footnote 6]. Like many other phonetic and phonological alternations, the historical patterns of vowel change discussed by Labov appear to have a basis in acoustic-phonetic factors.

A phonetic explanation for the patterning of vowels in the present study and in the aforementioned axiomatic principles along the backing dimension might lie in the combination of (1) differences in the contextual variability exhibited by particular vowels, and (2) the relative perceptual weakness of F_2 documented in the present study and in previous research. Natural variability of vowels has been proposed as a potential explanation for the direction of the historical trends by Goldstein [1983], on the basis of articulatory synthesizer data suggesting that front vowels will naturally vary in F_1, while back vowels will vary in both F_1 and F_2. A review of the coarticulation literature [Recasens, 1999] indicates that the F_2 variability of back vowels may be stronger than suggested by Goldstein [1983]. The tongue dorsum constriction in back vowels is particularly susceptible to perturbation due to adjacent consonants, more so than front vowels, with coronal consonants having the greatest effect. Various studies have documented the resulting F_2 overshoot (i.e., fronting) [English: Stevens and House, 1963; Hillenbrand et al., 2001; Swedish: Lindblom, 1963; Catalan: Recasens, 1985, 1991]. Note that in the present study, the increased accuracy in vowel classification with S/N ratio indicates that any centralization of the vowels in the stimuli of the present study was not abnormal but within the range expected by the listeners.

In contrast to the back vowels, the vowel /i/ is especially acoustically stable in the face of articulatory variability [Stevens, 1989], and to a lesser extent the other front vowels are as well. Back vowels show much more centralization than front vowels, and /i/ shows the least variability of all.

However, the case of the bias in favor of /a/ over lax /æ/ may not be explainable in coarticulatory terms. While lax vowels, because they tend to be shorter, may have their F_2 target frequencies more affected by consonantal gestures than those of tense vowels, leading to more F_2 variation as a class [Recasens, 1999], the vowel /æ/, regarded as lax by Labov [1994] and many phonologists, is in fact the longest vowel reported by Hillenbrand et al. [2001, table 1] at 266.6 ms, well above an average of all vowels of 198 ms. Other studies of American English [Peterson and Barney, 1952; Hillenbrand et al., 1995] report similar durational facts. The classification of American English lax /ɛ/ as /ʌ/, however, may yield to a coarticulatory explanation, since it is truly short with an average duration of 176.1 ms according to Hillenbrand et al. [2001].

On the perceptual side, the logarithmic nature of the auditory Bark scale will make a difference in F_2 in the back vowel region, where F_2 frequency is low, more salient to listeners than an equal F_2 Hertz in the front vowel region, where F_2 frequency is high. Thus, even if coarticulation in Hertz for F_2 were equal for front vowels and back vowels, the perceptual shift will be greater for back vowels than front vowels. This explanation may apply to the confusability between /a/ and /æ/, with the additional factor of the more crowded F_2 space for low values. As acknowledged above, however, the direction of the asymmetry remains unexplained.

In sum, an explanation for the strong biases in favor of /i e/ could be the result of normal coarticulation on /u o/. Listeners were unable to compensate for this coarticulation because of uncertainty of the consonantal context and the relative perceptual weakness of the F_2 dimension. On a longer time scale, the same factors could result in the historically attested directions of vowel movement, along the lines of Ohala's [1981] hypocorrection scenario.

Advantage of Voicing and Manner of Articulation over Place of Articulation

The same reasons for the robustness of the vowel height dimension relative to the backing dimension may also explain the higher T_{rel} for voicing and manner of articulation than for place of articulation in consonants. While all three contrasts are conveyed by a number of acoustic cues, the temporal pattern of low-frequency energy onset is one of the most important cues for the perception of voicing [Liberman et al., 1958]. Manner of articulation is similarly conveyed primarily by the temporal pattern of intensity change between the consonant and the vowel, mostly conveyed and determined by low-frequency components of the speech signal with the exception of certain fricatives [Miller and Nicely, 1955]. On the other hand, the transitions of F_2 and F_3 are responsible for the perception of place of articulation [Cooper et al., 1952; Miller and Nicely, 1955]. An additional reason for greater robustness of voicing and manner of articulation could be that the masking noise in the present study has primarily spectral and intensity-based effects, but does not affect durational cues, such as voice onset time (VOT) or preconsonantal vowel duration, important cues for the voicing contrast [Lisker, 1957; Lisker and Abramson, 1964].

The *synchrony capture hypothesis* suggests that a shift in neural synchrony from F_2 to F_1 in the transition from a voiceless obstruent to a vowel is a cue for voicelessness [Kluender et al., 1995]. Recall that an intense F_1 can cause mid CF auditory nerve fibers to synchronize to the frequency of F_1 instead of synchronizing to the frequency of a spectral peak closer to their CF. While recent results have failed to confirm the synchrony capture hypothesis [Lotto and Kluender, 2002], the broader synchrony response for low-frequency signals than for mid- and high-frequency signals likely provides more reliable perception in noise of contrasts conveyed by lower frequencies (such as voicing and manner of articulation) than by higher frequencies (such as place of articulation), along the lines described above for vowel height over backing.

There is evidence for a durational cue for place of articulation which might be robust in noise, with listeners sensitive to covariation in VOT with place of articulation [Miller, 1977; Benkí, 2001; but cf. Kluender, 1991]. However, results from the categorization of CV stimuli varying in both VOT and F_3 transition pattern among /ba pa ɡa ka/ indicate that such sensitivity affects the perception of voicing more than the perception of place of articulation [Benkí, 1998].

Advantage of Onsets over Codas

The present results indicate that onsets have a perceptual advantage over codas in noise, consistent with Wang and Bilger [1973]. In that study, intelligibility was higher for the first of two sets of 16 CV syllables than two sets of 16 VC syllables, one of which was phonemically matched to the first CV set. The intelligibility of the second CV set was the same as the two VC sets, likely a result of the second CV set containing the phonemes /h h^w w/.

There are two important ways in which the present results show a perceptual advantage for onsets over codas. First, onsets have a T_{rel} about 15% points higher than codas across the range of S/N ratios (fig. 1). This increase is reflected in the feature analysis, with higher T_{rel} in onsets than codas for voicing and manner of articulation and most of all for place of articulation (fig. 3). Second, particular codas are deleted more often than onsets (which as a group are deleted negligibly), namely /m n l b ɡ/, the sonorants and the voiced stops (table 7).

The advantage of codas over onsets appears to result from factors in the production as well as the perception of speech. There are a number of articulatory advantages for onsets with acoustic consequences. Release bursts may favor onsets over codas to the extent that they signal important place of articulation information to listeners [Stevens and Blumstein, 1981]. Although they were present for the coda stimuli used in this study, coda release bursts are generally weaker than onset bursts [Krakow, 1989]. Early devoicing gestures such as opening or constricting the glottis for voiceless stops will reduce extent of the VC formant transitions [Manuel, 1991], which are important cues for place of articulation. In nasals, velum lowering and oral closure are simultaneous in onsets, but velum lowering significantly precedes oral closure of codas, resulting in a crisp transition from onset nasal to vowel but a blurred transition from vowel to coda nasal [Krakow, 1989, 1993], making the coda nasal harder to distinguish from the vowel. Additionally, /n/ has greater linguopalatal contact at the beginnings of prosodic domains, such as the syllable, than at the ends [Fourgeron and Keating, 1997]. In general, syllable-initial articulations are more stable and precise than syllable-final articulations [see reviews by Krakow, 1999; Ohala and Kawasaki, 1984].

These articulatory and acoustic advantages for onsets are complemented by aspects of the auditory system that result in greater sensitivity for onsets than for codas [see discussion in Wright, 2001]. Abrupt rises in signal energy, such as at syllable onset, cause a disproportionate increase in auditory-nerve fiber firing rate for about 50 ms before saturation, effectively amplifying this portion of the speech signal [Kiang et al., 1965; Smith, 1979]. This effect has been found to occur for consonants in CV sequences [Sinex and Geisler, 1983; Delgutte and Kiang, 1984a, b] but not to occur for consonants in VC sequences [Sinex, 1995].

Other studies besides the present one also find the formant transitions of CV sequences to be more readily perceived than VC formant transitions. VC transitions are generally less effective than matched CV transitions in signaling place of articulation for VC_1C_2V-type stimuli [Fujimura et al., 1978; Repp, 1978, 1983; Ohala, 1990], although Kingston [in press] reports differences in place of articulation. Wright [2001] reports that place of articulation is more robust for matched CV stimuli than for VC stimuli in noise, with the releases edited out of both sets of stimuli.

These disadvantages for codas provide additional empirical evidence for a phonetic explanation of the crosslinguistic prevalence of open syllables. The generalization that every language has open syllables and some languages only have open syllables, observed by Jakobson [1941], is a central linguistic generalization that phonological theories have sought to explain [e.g., Clements and Keyser, 1983; Prince and Smolensky, 1993]. Under the sociophonetic adaptation hypothesis that phonological structure will likely be selected if favored on multiple phonetic and/or social grounds, the prediction is that sound change may result in the maintenance of onsets but the loss of codas. The multiple phonetic disadvantages for codas mean that it may not be necessary for phonological theories to stipulate an innate preference on the part of human language for open syllables in order to explain this trend [Ohala, 1990].

In addition to the avoidance or deletion of coda consonants, neutralization is another possible phonological configuration resulting from the perceptual disadvantages of consonantal contrasts in the coda such as voicing, manner of articulation, and place of articulation. Indeed, neutralization of both place of articulation and laryngeal contrasts in syllable and word-final position are well-attested patterns in the world's languages [for recent reviews see Steriade, 1997; Lombardi, 2001]. The application of the present results to other languages is interesting in that while English has voicing, place of articulation, and manner contrasts in word-final consonants, figure 3 shows that perceptual confusions of features in word-final position are more common than in initial position, as might be expected for a language without these contrasts in word-final position.

The pattern of coda deletions observed in this study, with /m n/ deleted more often than /l d ɡ/, which are in turn deleted more often than /p t k s z/, may also reflect the same phonetic factors that give rise to the specific patterns of coda deletion observed in the world's languages. The voiceless stops are less likely than the voiced stops to be deleted by listeners in coda position probably because release bursts are more intense for voiceless stops than for voiced stops [Lisker, 1957]. The fricatives /s z/ are less likely to be deleted on similar grounds, as their frication noise is particularly intense.

The coda deletion results may be applied to Steriade's [2001] proposal that phonological neutralizations are subject to constraints of perceptual similarity between the underlying and surface representations. The perceptual weakness of the place of articulation contrast in coda nasals has been documented [Hura et al., 1992; Beddor and Evans-Romaine, 1995; Wright, 2001], providing a basis for nasal place assimilation or neutralization in coda or word-final position. The present coda deletion data provide an empirical basis for considering open syllables to be perceptually most similar to syllables closed by nasals, somewhat similar to syllables closed by /l/ and voiced stops, and least similar to syllables closed by fricatives and voiceless stops. For nasals in particular, the slower spectral change of codas relative to onsets, due to velum lowering during the vowel [Krakow, 1989, 1993], may make them most susceptible to perceptual deletion [Beddor and Evans-Romaine, 1995], partly explaining the historical pattern of word- or syllable-final nasal deletion [Hajek, 1997].

However, because all of the coda consonants in the present study are also word-final, the results may not distinguish among the fine-grained explanatory proposals for the weakness of codas and their tendency to assimilate under specific conditions. Such accounts include phonologically oriented accounts claiming that contrasts may be licensed on a prosodic (syllable position) basis [e.g., Itô, 1986; Beckman, 1998; Lombardi, 2001]. More phonetically oriented accounts include Steriade's [1997] cue licensing proposal, in which contrasts are made when their acoustic-phonetic cues are perceptually salient, the recent perceptual similarity proposal [Steriade, 2001], and the proposal that speakers use consonantal releases to signal prosodic boundaries for the purpose of word segmentation [Kirk, 2001; Kingston, in press].

It is tempting to connect the result that voiceless stop codas /p t k/ are less likely to delete than voiced stop codas /d g/ with the crosslinguistic direction of voicing neutralization, although /z/ did not pattern like the stops. The overwhelming pattern for languages with voicing neutralization of obstruents in word- or syllable-final position is neutralization to the voiceless obstruent. However, deletion of voiced stops (but not voiceless stops) as a stage in the loss of a voicing contrast in codas is not attested. What the present data may instead indicate is the perceptual stability of voiceless stops, and strident fricatives in general, in coda or word-final position.

Asymmetrical Confusions in Consonants

No comprehensive explanation or generalization seems apparent for the consonant confusion asymmetries favoring onsets /p/ over /k/, /l/ over /r/ or codas /t/ over /k/, /p/ over /k/, or /n/ over /m/. As discussed above, the lexicality and relative frequencies of confusable phonemes do not provide a comprehensive account, but may be worth further investigation. The phonological proposal of radical underspecification with [coronal] as the default place of articulation [Paradis and Prunet, 1991] also fails to provide a simple explanation. While the nasals are consistent with underspecification, the oral stops are not: /p/ is favored just as much if not more than /t/, and there is no clear bias toward /d/.

One possibility for the biases for /t/ and /p/ over /k/ is that the characteristic velar mid-frequency concentration of energy during the release and its multiple bursts [Stevens and Blumstein, 1981; see recent review in Plauché, 2001] could be masked by the noise. Without these unique acoustic-phonetic cues, /k/ perceptually resembles /p/ and /t/, but /p/ and /t/ in the presence of noise would not be mistaken for /k/. Perceptual studies show support for such an explanation in the case of the observed asymmetric pattern favoring /ti/ over /ki/ [Plauché et al., 1997; Chang et al., 2001; Plauché, 2001]. In the account given in those reports, listeners may occasionally fail to perceive a mid-frequency burst when presented with /ki/, resulting in the perception of /ti/, but listeners will rarely if ever hallucinate a mid-frequency burst when presented with /ti/, and will therefore rarely report /ki/. However, it is not likely that a high front vowel context could account for all of the observed /p/ and /t/ responses for /k/, as there are only three stimuli with the vowels /i ɪ/ next to a /k/ (/kiz kɪp kɪg/). A follow-up study with a more focused sampling of onset-vowel and vowel-onset combinations would be worthwhile to further investigate this issue.

Conclusion

Data collection in open-response set experiments in the past has been very tedious and expensive. With online stimulus presentation and response collection, these experiments are much easier to conduct. A number of results reported originally on the basis of forced-choice (closed-response set) experiments using highly trained listeners have been replicated here with an open-response set using relatively naive listeners, such as the advantages of voicing over place of articulation in consonants, word-initial consonants over word-final consonants, and the height contrast over the front-back contrast in vowels. In addition, the open-response format made some patterns observable that would not be in a closed-set task, including the word-final consonant deletion rates and some confusion asymmetries.

Nearly all of the present results, including the bias toward the front vowels /i e/ over the back vowels /u o/, have been observed in previous studies, and have bases in acoustic, articulatory, and auditory factors. Because the present methodology models speech perception in a normal communicative situation more realistically than forced-choice experiments, the results provide even stronger evidence than previous reports for these perceptual advantages and asymmetries being important factors in determining common crosslinguistic phonological patterns.

Acknowledgments

Special thanks to Katherine Cassady and Sherene Flemmings for running the experiment. A preliminary version of this work was presented in ‘Recognition of Word and Nonsense Syllables in Noise’, a poster at the 143rd meeting of the Acoustical Society of America, Pittsburgh, June 2002. I am grateful to colleagues Rusty Barrett, Pam Beddor, Teresa Satterfield, and attendees of the ASA meeting, particularly Terry Nearey, for useful comments and discussion. The reviews of an anonymous reviewer and Madeleine Plauché improved this paper enormously. All errors are mine.

Appendix

These are the CVC nonsense stimuli used by Boothroyd and Nittrouer [1988] and in the present study. About 24 of the 120 original Boothroyd and Nittrouer nonsense syllable stimuli may be actual lexical items for some speakers of American English, either because of an /a/-/ɔ/ merger (e.g., /sak/ *sock*), the items may be familiar proper names (e.g., /lɛn/ *Len*), or in the case of one item, the possibility of a morphologically complex word was not considered (e.g. /kiz/ *keys*). These items were kept in the present study and responses to them were analyzed along with the rest of the data.

maɪɡ
kɔs
hem
dɪt
tup
bak
pod
sɛm
læl
riz
rit
sez

bep
raɪk
lim
mɪm
put
sog
tæz
das
hɔd
kɛl
reg
tuk

kɛz
pɪd
sum
raɪɡ
ten
lol
mat
dɔk
bis
hæp
lɛn
dæp

biɡ
tæt
pɛm
ros
kal
hun
dɪz
lep
sɔk
maɪd
pæz
his

ruk
tet
laɪl
bɔp
pɪm
kæz
don
had
mis
sɛɡ
rem
lɔk

Appendix (continued)

bɔd	mɔt	paɪt	mɛk	mul
kum	kɪp	bɪm	som	dæs
lɛp	bod	siɡ	tal	hɛz
pon	hɑs	hɑl	lɔt	tid
haɪs	læn	kez	ked	pot
dæk	saɪl	rus	baɪp	ban
tiɡ	dɛz	tɔd	duɡ	kaɪp
mal	pim	mok	rɪn	sɪɡ
puk	kiz	ted	dɔs	haɪm
lit	sæn	bɛp	moɡ	ral
kɪɡ	pek	taɪs	dɔd	miz
rɛm	hæp	lan	sot	bul

References

Allen, J.: How do humans process and recognize speech? IEEE Trans. Speech Audio Processing *2:* 567–577 (1994).

Alwan, A.; Zhu, Q., Lo, J.: Human and machine recognition of speech sounds in noise. Proc. 13th World Multiconf. on Systems, Cybernetics, and Information, pp. 218–223 (2001).

Bahl, L.R.; Jelinek, F.; Mercer, R.L.: A maximum likelihood approach to continuous speech recognition. IEEE Trans. Pattern Analysis Machine Intelligence PAMI-5, pp. 179–190 (1983).

Barrett, R.: A grammar of Sipakapense Maya; PhD diss. University of Texas, Austin (1999).

Beckman, J.N.: Positional faithfulness; PhD diss. University of Massachusetts, Amherst (1998).

Beddor, P.S.; Evans-Romaine, D.: Acoustic-perceptual factors in phonological assimilations: a study of syllable-final nasals. Riv. Ling. *7:* 145–174 (1995).

Benkí, J.R.: Evidence for phonological categories from speech perception; PhD diss. University of Massachusetts, Amherst (1998).

Benkí, J.R.: Place of articulation and first formant transition pattern both affect perception of voicing in English. J. Phonet. *29:* 1–22 (2001).

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

Boothroyd, A.; Nittrouer, S.: Mathematical treatment of context effects in phoneme and word recognition. J. acoust. Soc. Am. *84:* 101–114 (1988).

Chang, S.C.; Plauché, M.; Ohala, J.: Markedness and consonant confusion asymmetries; in Hume, Johnson, The role of speech perception in phonology (Academic Press, San Diego 2001).

Clarke, F.R.: Constant-ratio rule for confusion matrices in speech communication. J. acoust. Soc. *29:* 515–520 (1957).

Clements, G.N.; Keyser, S.J.: CV phonology: a generative theory of the syllable (MIT Press, Cambridge 1983).

Cooper, F.S.; Liberman, A.M.; Delattre, P.; Gerstman, L.: Some experiments on the perception of speech sounds. J. acoust. Soc. Am. *24:* 597–606 (1952).

Cover, T.M.; Thomas, J.A.: Elements of information theory (Wiley, New York 1991).

Cox, F.: Vowel change in Australian English. Phonetica *56:* 1–27 (1999).

Crothers, J.: Typology and universals of vowel systems; in Greenberg, Ferguson, Moravcsik, Universals of human language, vol. 2: Phonology (Stanford University Press, Stanford 1978).

Delattre et al., 1955.

Delgutte, B.; Kiang, N.Y.S.: Speech coding in the auditory nerve. III. Voiceless fricative consonants. J. acoust. Soc. Am. *75:* 887–896 (1984a).

Delgutte, B.; Kiang, N.Y.S.: Speech coding in the auditory nerve. IV. Sounds with consonant-like dynamic characteristics. J. acoust. Soc. Am. *75:* 897–907 (1984b).

Fant, G.: On the predictability of formant levels and spectrum envelopes from formant frequencies; in For Roman Jakobson (Mouton, The Hague 1956).

Fant, G.: Acoustic theory of speech production. The Hague, Netherlands: Mouton (1960).

Fairbanks, G.; Grubb, P.: A psychophysical investigation of vowel formants. J. Speech Hear. Res. *4:* 203–219 (1961).

Fletcher, H.: Speech and hearing in communication (Van Nostrand, New York 1953).

Fourgeron, C.; Keating, P.A.: Articulatory strengthening at edges of prosodic domains. J. acoust. Soc. Am. *101:* 3728–3740 (1997).

French, N.R.; Carter, C.W.; Koenig, W.: The words and sounds of telephone conversations. Bell Syst. Tech. J. *9:* 290–324 (1930).

Fujimura, O.; Macchi, M.J.; Streeter, L.A.: Perception of stop consonants with conflicting transitional cues: a cross-linguistic study. Lg. and Speech *21:* 227–346 (1978).

Goldstein, L.: Vowel shifts and articulatory-acoustic relations; in Cohen, van den Broecke, Proc. 10th Int. Congr. Phonet. Sci., pp. 267–273 (Foris, Dordrecht 1983).

Gordon, M.; Heath, J.: Sex, sound symbolism, and sociolinguistics. Curr. Anthrop. *39:* 421–449 (1998).

Greenberg, S.: The ear as a speech analyzer. J. Phonet. *16:* 139–150 (1988).

Hajek, J.: Universals of sound change in nasalization. Publ. Philological Soc. 31 (Blackwell, Oxford 1997).

Hillenbrand, J.M.; Clark, M.J.; Nearey, T.M.: Effects of consonant environment on vowel formant patterns. J. acoust. Soc. Am. *109:* 748–763 (2001).

Hillenbrand, J.M.; Getty, L.A.; Clark, M.J.; Wheeler, K.: Acoustic characteristics of American English vowels. J. acoust. Soc. Am. *97:* 3099–3111 (1995).

Hume, E.; Johnson, K.: A model of the interplay of speech perception and phonology; in Hume, Johnson, The role of speech perception in phonology (Academic Press, San Diego 2001).

Hura, S.L.; Lindblom, B.; Diehl, R.: On the role of perception in shaping phonological assimilation rules. Lang. Speech *35:* 59–72 (1992).

Itô, J.: Syllable theory in prosodic phonology; PhD diss. University of Massachusetts, Amherst (Garland, New York 1986).

Jakobson, R.: Kindersprache, Aphasie und allgemeine Lautgesetze (Sprakvetenskapliga Sallskapets i Uppsala Forhandlingar 1940–1942, Almqvist and Wiksell, Uppsala 1941). Reprinted in Jakobson, Selected writings 1 (Mouton, The Hague, 1962).

Jenison, R.L.; Greenberg, S.; Kluender, K.R.; Rhode, W.S.: A composite model of the auditory periphery for the processing of speech based on the filter response functions of single auditory-nerve fibers. J. acoust. Soc. Am. *90:* 773–786 (1991).

Kenstowicz, M.; Kisseberth, C.: Generative phonology (Academic Press, San Diego 1979).

Kiang, N.Y.S.; Moxon, E.C.: Tails of tuning curves of auditory-nerve fibers. J. acoust. Soc. Am. *55:* 620–630 (1974).

Kiang, N.Y.S.; Watanabe, T.; Thomas, E.C.; Clark, L.F.: Discharge patterns of single fibers in the cat's auditory nerve. MIT Res. Monogr. No. 35 (MIT Press, Cambridge 1965).

Kingston, J.: Keeping and losing contrasts. Proc. 28th Annu. Meet. Berkeley Ling. Soc., Berkeley 2002 (Berkeley Linguistics Society, Berkeley, in press).

Kirk, C.J.: Phonological constraints on the segmentation of continuous speech; PhD diss. University of Massachusetts, Amherst (2001).

Klein, W.; Plomp, R.; Pols, L.C.W.: Vowel spectra, vowel spaces, and vowel identification. J. acoust. Soc. Am. *48:* 999–1009 (1970).

Kluender, K.R.: Effects of first formant onset properties on voicing judgments result from processes not specific to humans. J. acoust. Soc. Am. *90:* 83–96 (1991).

Kluender, K.R.; Lotto, A.J.; Jenison, R.L.: Perception of voicing for syllable-initial stops at different intensities: does synchrony capture signal voiceless stop consonants? J. acoust. Soc. Am. *97:* 2552–2567 (1995).

Krakow, R.: The articulatory organization of syllables: a kinematic analysis of labial and velar gestures; PhD diss. Yale University, New Haven (1989).

Krakow, R.: Nonsegmental influences on velum movement patterns: syllables, sentences, stress, and speaking rate; in Huffman, Krakow, Nasals, nasalization, and the velum (Phonetics and phonology V), (Academic Press, San Diego 1993).

Krakow, R.: Physiological organization of syllables: a review. J. Phonet. *27:* 23–54 (1999).

Kucera, F.; Francis, W.: Computational analysis of present day American English (Brown University Press, Providence 1967.

Labov, W.: Principles of linguistic change, vol. I: Internal factors (Blackwell, Cambridge 1994).

Labov, W., Yaeger, M.; Steiner, R.: A quantitative study of sound change in progress (US Regional Study, Philadelphia 1972).

Lenzo, K.: t2p: Text-to-phoneme converter builder software. Retrieved from <http://www-2.cs.cmu.edu/~lenzo/t2p/> (1998).

Liberman, A.M.; Delattre, P.C.; Cooper, F.S.: Some cues for the distinction between voiced and voiceless stops in initial position. Lang. Speech *1:* 153–167 (1958).

Liljencrantz, J.; Lindblom, B.: Numerical simulation of vowel quality systems: the role of perceptual contrast. Language *48:* 839–862 (1972).

Lindblom, B.: Spectrographic study of vowel reduction. J. acoust. Soc. Am. *35:* 1773–1781 (1963).

Lindblom, B.: Experiments in sound structure. Plenary address, 8th Int. Congr. Phonet. Sci., Leeds 1975. Rev. Phonét. appl. *51:* 154–189 (1979).

Lindblom, B.: Phonetic universals in vowel systems; in Ohala, Jaeger, Experimental phonology (Academic Press, New York 1986).

Lindblom, B.; Guion, S.; Hura, S.; Moon, S.-J.; Willerman, R.: Is sound change adaptive? Riv. Ling. *7:* 5–37 (1995).

Lippmann, R.P.: Speech recognition by machines and humans. Speech Commun. *22:* 1–15 (1997).

Lisker, L.: Closure duration and the intervocalic voiced-voiceless distinction in English. Language *33:* 42–49 (1957).

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

Lombardi, L.: Why place and voice are different: constraint-specific alternations in optimality theory; in Lombardi, Segmental phonology in optimality theory: constraints and representations (Cambridge University Press, Cambridge 2001).

Lotto, A.J., Kluender, K.R.: Synchrony capture hypothesis fails to account for effects of amplitude on voicing perception. J. acoust. Soc. Am. *111:* 1056–1062 (2002).

Luce, P.A.; Pisoni, D.B.: Recognizing spoken words: the neighborhood activation model. Ear Hear. *19:* 1–36 (1998).

Macmillan, N.A.; Creelman, C.D.: Detection theory: a user's guide (Cambridge University Press, Cambridge 1991).

Manuel, S.Y.: Some phonetic bases for the relative malleability of syllable-final versus syllable-initial consonants. Proc. 12th Int. Congr. Phonet. Sci., vol. 5, pp. 118–121 (University of Provence, Aix-en-Provence 1991).

Miller, G.A.; Heise, G.A.; Lichten, W.: The intelligibility of speech as a function of the context of the test material. J. exp. Psychol. *41:* 329–335 (1951).

Miller, G.A.; Niceley, P.E.: An analysis of perceptual confusions among some English consonants. J. acoust. Soc. Am. *27:* 338–352 (1955).

Miller, J.L.: Nonindependence of feature processing in initial consonants. J. Speech Hear. Res. *20:* 519–528 (1977).

Nearey, T.N.: The segment as a unit of speech perception. J. Phonet. *18:* 347–373 (1990).

Nearey, T.N.: Speech perception as pattern recognition. J. acoust. Soc. Am. *101:* 3241–3254 (1997).

Newman, S.: Yokuts language of California. Viking Fund Publ. Anthrop. vol. 2 (New York 1944).

Nooteboom, S.: Perceptual confusions among Dutch vowels presented in noise. IPO Annu. Progr. Rep., vol. 3, pp. 68–71 (Instituut voor Perceptie Onderzoek, Eindhoven 1968).

Nusbaum, H.C.; Pisoni, D.B.; Davis, C.K.: Sizing up the Hoosier mental lexicon: measuring the familiarity of 20,000 words. Research on Speech Perception Progress Report, No. 10 (Speech Research Laboratory, Psychology Department, Indiana University, Bloomington 1984).

Ohala, J.: The listener as a source of sound change. Papers from the parasession on language and behavior (Chicago Linguistics Society, Chicago 1981).

Ohala, J.J.: Alternatives to the sonority hierarchy for explaining segmental sequential constraints. Papers from the regional meetings, vol. 2, pp. 319–338 (Chicago Linguistic Society, Chicago 1990).

Ohala, J.J.; Kawasaki, H.: Prosodic phonology and phonetics. Phon. Yb. *1:* 113–127 (1984).

Paradis, C.; Prunet, J.-F.: The special status of coronals: internal and external evidence (Academic Press, New York 1991).

Peterson, G.E.; Barney, H.L.: Control methods used in a study of the vowels. J. acoust. Soc. Am. *24:* 175–184 (1952).

Pickett, J.M.: Perception of vowels heard in noises of various spectra. J. acoust. Soc. Am. *29:* 613–620 (1957).

Plauché, M.C.: Acoustic cues in the directionality of stop consonant confusions; PhD diss. University of California, Berkeley (2001).

Plauché, M.; Delogu, C.; Ohala, J.: Asymmetries in consonant confusion. Proc. Eurospeech '97: 5th Eur. Conf. on Speech Commun. Technol., vol. 4, pp. 2187–2190 (1997).

Prince, A.; Smolensly, P.: Optimality theory: constraint interaction in generative grammar (Rutgers University, New Brunswick and University of Colorado, Boulder, unpublished manuscript, 1993).

Recasens, D.: Coarticulatory patterns and degrees of coarticulatory resistance in Catalan CV sequences. Lang. Speech *28:* 97–114 (1985).

Recasens, D.: An electropalatographic and acoustic study of consonant-to-vowel coarticulation. J. Phonet. *19:* 177–192 (1991).

Recasens, D.: Lingual coarticulation; in Hardcastle, Hewlett, Coarticulation: theory, data and techniques (Cambridge University Press, Cambridge 1999).

Repp, B.H.: Perceptual integration and differentiation of spectral information of spectral cue for intervocalic stop consonants. Percept. Psychophys. *24:* 471–485 (1978).

Repp, B.H.: Bidirectional context effects in the perception of VC-CV sequences. Percept. Psychophys. *33:* 147–155 (1983).

Savin, H.: Word-frequency effects and errors in the perception of speech. J. acoust. Soc. Am. *35:* 200–206 (1963).

Schroeder, M.R.: Reference signal for signal quality studies. J. acoust. Soc. Am. *44:* 1735–1736 (1968).

Shannon, C.E.: A mathematical theory of communication. Bell Syst. tech. J. *27:* 379–423, 623–656 (1948).

Sinex, D.G.: Auditory nerve fiber representation of cues to voicing in syllable-final stop contexts. J. acoust. Soc. Am. *90:* 2441–2449 (1995).

Sinex, D.G.; Geisler, C.D.: Responses of auditory-nerve fibers to consonant-vowel syllables. J. acoust. Soc. Am. *73:* 602–615 (1983).

Smith, R.L.: Adaptation, saturation and physiological masking in single auditory-nerve fibers. J. acoust. Soc. Am. *65:* 166–178 (1979).

Smits, R.: Temporal distribution of information for human consonant recognition VCV utterances. J. Phonet. *28:* 111–135 (2000).

Sommers, M.S.; Kirk, K.I.; Pisoni, D.B.: Some considerations in evaluating spoken word recognition by normal-hearing, noise-masked normal-hearing, and cochlear implant listeners. I: The effects of response format. Ear Hear. *18:* 89–99 (1997).

Son, R.J.J.H. van; Pols, L.C.W.: Perisegmental speech improves consonant and vowel identification. Speech Commun. *29:* 1–22 (1999).

Steele, R.D.: The segmental phonology of contemporary standard Polish; PhD diss. Harvard University (1973).

Steiber, Z.: A historical phonology of the Polish language (Winter Universitätsverlag, Heidelberg 1973).

Stemberger, J.P.: Vocalic underspecification in English language production. Language *68:* 492–524 (1992).

Steriade, D.: Directional asymmetries in assimilation: a directional account; in Hume, Johnson, The role of speech perception in phonology (Academic Press, San Diego 2001).

Steriade, D.: Phonetics in phonology: the case of laryngeal neutralization (University of California, Los Angeles, unpublished manuscript; <http://web.mit.edu/linguistics/www/steriade.home.html>, 1997).

Stevens, K.N.: On the quantal nature of speech. J. Phonet. *17:* 3–45 (1989).

Stevens, K.N.; Blumstein, S.E.: The search for invariant acoustic correlates of phonetic features; in Eimas, Miller, Perspectives on the study of speech (Erlbaum, Hillsdale 1981).

Stevens, K.N.; House, A.S.: Perturbations of vowel articulations by consonantal context: an acoustical study. J. Speech Hear. Res. *6:* 111–128 (1963).

Sweet, H.: History of English sounds (Clarendon Press, Oxford 1888).

Wang, M.D.; Bilger, R.C.: Consonant confusions in noise: a study of perceptual features. J. acoust. Soc. Am. *54:* 1248–1266 (1973).

Whalen, D.W.: Vowel and consonant judgments are not independent when cued by the same information. Percept. Psychophys. *46:* 284–292 (1989).

Wright, R.: Perceptual cues in contrast maintenance; in Hume, Johnson, The role of speech perception in phonology (Academic Press, San Diego 2001).

Young, E.D.; Sachs, M.B.: Representation of steady-state vowels in the temporal aspects of the discharge patterns of the populations of auditory-nerve fibers. J. acoust. Soc. Am. *66:* 1381–1403 (1979).