Consonant identification in consonant-vowel-consonant syllables in speech-spectrum noise

David. L. Woods[a)]
Department of Neurology, UC Davis and VANCHCS, 150 Muir Road, Martinez, California 95553

E. William Yund, Timothy J. Herron, and Matthew A. I. Ua Cruadhlaoich
Research Service, VANCHCS, 150 Muir Road, Martinez, California 95553

(Received 20 March 2009; revised 16 December 2009; accepted 22 December 2009)

Identification functions of 20 initial and 20 final consonants were characterized in 9600 randomly sampled consonant-vowel-consonant (CVC) tokens presented in speech-spectrum noise. Because of differences in the response criteria for different consonants, signal detection measures were used to quantify identifiability. Consonant-specific baseline signal-to-noise ratios (SNRs) were adjusted to produce a d' of 2.20 for each consonant. Consonant identification was measured at baseline SNRs (B), at B−6, and at B+6 dB. Baseline SNRs varied by more than 40 dB for different consonants. Confusion analysis revealed that single-feature place-of-articulation errors predominated at the highest SNR, while combined-feature errors predominated at the lowest SNR. Most consonants were identified at lower SNRs in initial than final syllable position. Vowel nuclei (/ɑ/, /i/, or /u/) significantly influenced the identifiability of 85% of consonants, with consistent vowel effects seen for consonant classes defined by manner, voicing, and place. Manner and voicing of initial and final consonants were processed independently, but place cues interacted: initial and final consonants differing in place of articulation were identified more accurately than those sharing the same place. Consonant identification in CVCs reveals contextual complexities in consonant processing.
[DOI: 10.1121/1.3293005]

PACS number(s): 43.71.Es, 43.71.Sy, 43.71.An [MSS] Pages: 1609–1623

I. INTRODUCTION

Although most spoken syllables contain multiple consonants, consonant identification of large consonants sets in multi-consonant syllables has received relatively little study. The current experiment investigated the identification of 20 initial and 20 final consonants in a pseudorandomly sampled set of 9600 consonant-vowel-consonant (CVC) tokens presented in speech-spectrum noise.

A. Quantifying consonant identification

In their classic study, Miller and Nicely (1955) found that the identifiability of different consonants presented in noise varies substantially: some consonants (e.g., nonsibilant fricatives) were difficult to identify even at high signal-to-noise ratios (SNRs), while others (e.g., sibilants) were accurately identified at much lower SNRs. For example, /ʃ/ hit rates at SNRs of −12 dB were similar to /ð/ hit rates at SNRs of +6 dB. Their results also suggested that comparisons of consonant identifiability using hit rate alone are potentially confounded by response biases associated with different consonants. For example, they found that hit rates for /v/ were much higher than hit rates for /ð/. However, at low and intermediate SNRs, the majority of /v/ responses were elicited by other consonants, particularly /ð/. Indeed, CVs containing /ð/ elicited more false /v/ responses than /ð/ hits. Similarly, Wang and Bilger (1973) found that CVs containing /ð/ elicited nearly twice as many false /v/ responses as /ð/ hits, while false /ð/ responses were rarely elicited by CVs containing /v/. Such asymmetric confusion patterns suggest that subjects have significant response biases: among confusable consonants, ambiguous phonological cues are much more likely to elicit the report of some consonants (e.g., /v/) rather than others (e.g., /ð/).

As a result, hit rate alone can misrepresent consonant identification accuracy. Metrics are needed that incorporate information from both hit and false response rates. Signal detection theory (Green and Swets, 1974) provides two such metrics: d', a measure of receiver sensitivity that quantifies the discriminability of sensory information, and beta, a criterion metric that quantifies the amount of sensory evidence needed to elicit a particular response. Therefore, the current experiments incorporated signal detection measures to quantify consonant identification performance following similar methods used in previous studies (Müsch and Buus, 2001b, 2001a).

B. Consonant identifiability in noise

Several previous studies have investigated the identification of small consonant sets in CVC syllables. Boothroyd and Nittrouer (1988) developed matched sets of 120 words and 120 nonsense syllables by combining ten initial consonants, ten vowels, and ten final consonants. Although they did not analyze the identification of individual consonants, their results suggested that initial consonants, final consonants, and vowels were processed independently in nonsense syllables. Benkí (2003) presented Boothroyd and Nittrouer's

[a)]Author to whom correspondence should be addressed. Electronic mail: dlwoods@ucdavis.edu

(1988) nonsense CVC tokens to 37 undergraduates and analyzed consonant-confusion patterns at four different SNRs spanning a range of 9 dB. Because of the small number of CVC tokens, separate subjects were tested at each SNR to avoid token repetition. Benkí (2003) found large differences in the identifiability of different consonants. For example, the sibilant /s/ was almost perfectly identified even at the most difficult SNR, whereas the non-sibilant fricative /f/ was frequently misidentified even at the highest SNR. These results imply that different consonants require markedly different SNRs for identification. However, because Benkí (2003) used only ten consonants, confusions between many potentially confusable consonant pairs (e.g., /s/ and /ʃ/) could not be examined. Moreover, because only 240 of 1000 possible CVC combinations were used, it was impossible to independently analyze the processing of initial consonants, final consonants, and vowels. Redford and Diehl (1999) used a completely randomized set of CVCs created by using the consonants /p/, /t/, /k/, /f/, /θ/, /s/, and /ʃ/) at initial and final positions combined with the three vowels /ɑ/, /i/, and /u/. The CVCs were presented at mid-sentence in different sentence contexts at SNRs of +15 dB. Significant differences were again found in consonant identification accuracy with non-sibilant fricatives proving particularly difficult to identify (e.g., /f/ and /θ/).

Previous studies of consonant identification in CVCs used small consonant sets and thus may have overestimated the identifiability of many consonants. For example, the unvoiced plosives (/p/, /t/, and /k/) are often confused with each other, but may also be confused with voiced plosives, affricates, /h/, and other unvoiced fricatives. The study of consonant identification in larger consonant sets has been largely restricted to single consonant (CV and VC) syllables. Miller and Nicely (1955) analyzed the identification of 16 common initial consonants in CVs with the vowel /ɑ/. There were five female participants, four of whom identified consonants in the CVs spoken by the fifth. Talkers rotated on successive blocks so that a total of approximately 7000 different tokens were presented at seven different SNRs ranging from −18 to +12 dB. Confusion matrices for 4000 CV token presentations (approximately 250 presentations of each consonant) were obtained at each SNR and revealed large differences in the identifiabilities of different consonants. For example, nasals were more accurately identified at −12 dB than non-sibilant fricatives (e.g., /θ/) at +6 dB. The percentage of consonants correctly identified improved with SNR, with mean hit rates increasing from 7.8% at −18 dB to 90.6% at +12 dB, producing an average performance/SNR (P/S) slope of 2.8%/dB. They found that the slopes of P/S functions were generally steepest around SNRs near mid-performance (e.g., 50% correct) levels. However, differences in the P/S slopes were evident for different consonants. For example, plosives had steeper P/S slopes (e.g., /t/ 5.9%/dB) than non-sibilant fricatives (e.g., /θ/ 2.3%/dB).

Wang and Bilger (1973) studied consonant identification in broadband noise using two different CV and VC lists, each with 16 consonants. They found large differences in consonant identification for both CVs and VCs, with non-sibilant fricatives again proving particularly difficult to identify. Phatak and Allen studied consonant identification in both speech-spectrum noise (Phatak and Allen, 2007) and white noise (Phatak *et al.*, 2008) using CVs containing 16 consonants paired with four vowels. Nasal consonants proved easy to identify in white noise but hard to identify in speech-spectrum noise. Again, non-sibilant fricatives proved particularly difficult to identify in both noise conditions. For example, /ð/ hit rates in speech-spectrum noise at 0 dB SNR were similar to /z/ hit rates at −22 dB SNR.

C. The identification of leading and trailing consonants

Several previous studies of CVCs have reported that initial consonants are more accurately identified than final consonants. For example, Benkí (2003) found lower error rates for initial than final consonants at all SNRs tested. Redford and Diehl (1999) also found that initial consonants were more accurately recognized than final consonants across most listening conditions. However, Redford and Diehl (1999) noted that the initial consonant advantage was absent for some consonants (e.g., /θ/) and was influenced by vowel context, being enhanced in syllables containing /ɑ/. Similar initial consonant advantages have been found in comparisons of CVs and VCs (Wang and Bilger, 1973; Dubno and Levitt, 1981). However, a recent study investigating consonant identification in CVs and VCs found the opposite result: final consonants were identified more accurately than initial consonants in six-talker speech babble (Cutler *et al.*, 2004). These unexpected results may have reflected differences in initial and final consonant masking levels. The babble-noise masker was created by manually selecting 1 s segments of continuous speech from each of six different talkers and then combining the six speech segments. This procedure may have resulted in a better alignment of the speech-babble amplitude envelopes at babble onset so that greater masking would have occurred for initial than final consonants.

D. Analyzing phonetic features

Consonant identification requires the accurate analysis of place, manner, and voicing features. Considerable evidence suggests that these features differ in their discriminability in noise. For example, Dubno and Levitt (1981) studied consonant confusions in speech-spectrum noise in 11 different CV and VC lists. In each list, seven, eight, or nine different consonants were presented with a single vowel (/ɑ/, /i/, or /u/) in speech-spectrum noise. They found higher rates for place than for manner errors, and higher rates for place errors than combined place+manner (P+M) errors. Because voiced and unvoiced consonants were presented in separate lists, voicing errors could not be evaluated. Other studies using larger consonant sets including both voiced and unvoiced consonants also found that place errors were more common than manner errors, and that both place and manner errors were more common than voicing errors (Miller and Nicely, 1955; Wang and Bilger, 1973; Levitt and Resnick, 1978; Helfer and Wilber, 1990; Gelfand *et al.*, 1992; Phatak and Allen, 2007; Phatak *et al.*, 2008).

E. Vowel influences on consonant identifiability

Previous studies also found that vowels influence consonant identifiability. For example, Wang and Bilger (1973) found that consonant identification varied with vowel: consonant report was most accurate for syllables containing /ɑ/, intermediate for syllables with /i/, and lowest for syllables containing /u/. Moreover, vowel context affected the identifiability of leading and trailing consonants differentially: leading consonants were more accurately identified in syllables containing /ɑ/ and trailing consonants were more accurately identified in syllables containing /i/. Dubno and Levitt (1981) also found an overall superiority of consonant processing in syllables containing /ɑ/. In particular, nasals were identified approximately twice as accurately in syllables containing /ɑ/ than in syllables containing /i/ or /u/. In contrast, affricates were most accurately reported in syllables containing /i/. In addition, they found an interaction between place of articulation and vowel: front and middle consonants were less accurately reported in syllables containing /i/ than /u/, whereas the reverse pattern was found for back consonants. Finally, in CVCs, Redford and Diehl (1999) found that vowel effects differed for initial and final consonants within a syllable: initial consonants were most accurately reported in syllables containing /ɑ/, whereas final consonants were most accurately reported in syllables containing /u/.

F. Interactions in identifying initial and final consonants

Interactions in the processing of initial and final consonants in CVCs have not been systematically examined. However, consonant-consonant interactions are known to alter the processing of place-of-articulation features over syllable-length silent gaps (Mann and Repp, 1981). This raises the possibility that interactions in the processing of initial and final consonant features might be observed in CVCs, for example, if stimulus-specific neuronal adaptation (Woods and Elmasian, 1983) reduced the salience of repeated consonant features.

G. The current experiment

The goal of the current experiment was to characterize the identification of consonants in CVCs that included 20 initial and final consonants combined with three vowels (/ɑ/, /i/, and /u/). In order to estimate average consonant confusions, performance measures were obtained from the pseudorandom sampling of 9600 different CVC syllable tokens. The large consonant set permitted the examination of an extensive set of potential consonant confusions. Signal detection theory was used to measure identification performance for each consonant over a consonant-specific range of SNRs adjusted to equate overall identification performance for different consonants. This permitted the characterization of both baseline identifiability and P/S functions for each consonant in initial and final syllable positions. This further allowed comparisons of the identifiabilities of consonant groups sharing similar place, manner, or voicing features as well as an analysis of the patterns of feature-processing errors (single-feature place, manner, voicing errors, and combined-feature errors) in initial and final syllable positions at different SNRs. Finally, the use of CVCs permitted an analysis of the effects of the vowel nucleus on the identifiability of initial and final consonants as well as the evaluation of possible interactions in the processing of initial and final consonants within a CVC.

II. METHODS

The present report describes consonant identification and consonant confusions observed in 34 560 trials, obtained from 16 young normal-hearing subjects who each participated in three separate 1 h testing sessions. Additional details about test-retest reliability, the relationship of consonant identification performance to audiometric thresholds, interactions in the processing of initial and final consonants in word and nonsense syllable tokens, and the relationship of consonant identification performance to sentence reception thresholds are provided elsewhere (Woods *et al.*, 2010).

A. Subjects

Sixteen young subjects (eight female and eight male, ages 18–30 yrs) with normal hearing (thresholds ≤20 dB hearing level at 250–4000 Hz)[1] each participated in three testing sessions within 11 days.

B. Syllable tokens

The CVC list included 1200 syllables constructed from the exhaustive combination of 20 initial consonants, 20 final consonants, and three vowels (/ɑ/, /i/, and /u/). Twenty-one consonants (/b/, /d/, /g/, /r/, /l/, /ŋ/, /n/, /m/, /v/, /ð/, /z/, /ʤ/, /ʧ/, /ʃ/, /s/, /θ/, /f/, /p/, /t/, /k/, and /h/) were used, 19 in both initial and final syllable positions, /h/ occurring in only the initial position, and /ŋ/ in only the final position. The token corpus was created by first recording 4800 syllable sets from each of four phonetically trained talkers (two male and two female) using an AKG C-410 head mounted microphone in an Industrial Acoustics Company (New York, NY) sound booth. The four talkers had been raised in different parts of the United States (two in the Midwest and two in California) and had slightly different American English speech patterns. Syllables were digitized at 16 bit resolution and 44.1 kHz sampling rate using MATLAB (The MathWorks, Inc., Natick, Ma). The complete syllable sets were reviewed by one of the authors (E.W.Y.) and the two best exemplars of each syllable (2400 tokens) were selected from each talker's corpus. Then, two listeners with normal hearing independently reviewed each of the 9600 syllables in the absence of masking noise to assure the intelligibility of all tokens. Whenever the intelligibility test failed, a new exemplar was substituted and further testing was performed among laboratory staff to assure the intelligibility of the substituted tokens. Within the entire 9600 token corpus syllable durations ranged from 350 to 890 ms (mean: 636 ms). For each token, 100 ms at the center of each vowel was identified by manual review to establish the time interval over which the noise masking levels would be linearly adjusted (on the dB scale) to provide appropriate masking levels for initial and final consonants differing in intrinsic identifiability (Fig. 1).

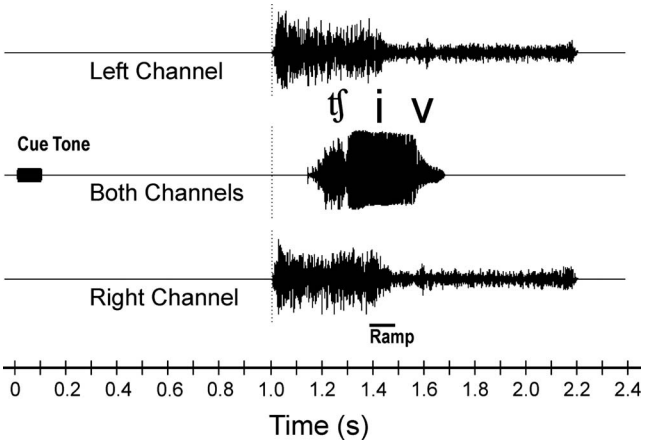

FIG. 1. Trial structure. Trials were cued by a 1.0 kHz tone. After 1.0 s two independent 1200 ms noise bursts were presented through the left and right loudspeakers. CVCs were presented simultaneously through both loudspeakers at random intervals after noise-burst onset. Noise amplitudes were modulated over a 100 ms interval during the mid-vowel segment of the CVC to provide appropriate masking levels for different initial and final consonants.

C. Speech-spectrum noise adjustment

Talker-specific speech-spectrum noise was used to mask each CVC token. The average spectrum for each talker was first obtained by averaging the spectra of all CVC tokens for that talker. This spectrum was then used to create a finite impulse response filter that was used to filter broadband white noise. Each filtered-noise file was trimmed of the first 0.5 s and cut into 100 different noise segments of 1200 ms duration each, which were then randomly sampled during the testing sessions.

D. Stimuli and procedure

Testing was performed in a 2.44×2.44 m single-walled, sound-attenuating testing room. The interior walls were covered by 2.5 cm thick acoustic foam resulting in ambient third-octave noise levels less than 20 dB sound pressure level (SPL) from 250 to 4000 Hz. In anticipation of future studies of hearing-impaired subjects with hearing aids, stimuli were presented through loudspeakers (M-Audio Studiophile AV 40). Immediately before the first CVC session, subjects were briefed with written and oral instructions and received ~5 min of training in identifying CVCs presented without masking noise.

During each session, each subject was presented with 720 different CVCs selected by constrained random sampling of the 9600 tokens. The CVCs were grouped by talker into 30-trial blocks with 24 talker-specific blocks presented during each experiment. Each trial (Fig. 1) began with a tone burst cue (100 ms, 1.0 kHz tone, 70 dB SPL). After a delay of 1.0 s, two randomly selected talker-specific noise bursts were presented independently through the left and right speakers, along with a single CVC presented synchronously from both loudspeakers. CVC onset times were randomized with respect to noise-burst onset with the constraint that each CVC began at least 100 ms after the onset of the noise bursts and ended at least 100 ms before their offset.

On each trial, the listener verbally reported the CVC token, referring, when necessary, to a list of acceptable initial and final consonants and vowels posted in the testing room. Responses were spoken into a microphone and phonetically transcribed using a modified QWERTY keyboard by an investigator (M.A.I.U.C.) listening through headphones in an adjacent room. Responses were spoken in quiet and subjects were queried via intercom when responses were invalid or poorly enunciated.[2] Subjects were given the option of repeating trials in cases of attentional lapse or noise interference (e.g., coughing, etc.). Each inter-trial interval of approximately 2 s included the time needed for syllable transcription plus a small delay (0.5 s) before the delivery of the warning tone signaling the next trial. Trials occurred at a rate of approximately 15/min so that each 720-syllable test required about 48 min in addition to rest breaks that occurred at each subject's discretion. Presentation software (NEUROBEHAVIORAL SYSTEMS, version 12.0) was used for stimulus delivery, masking noise and signal level adjustment, response monitoring, and d' calculations.

Syllable intensity was randomly roved from 70 to 75 dB SPL in 1 dB steps. Psychometric functions were measured for each initial and final consonant at three different SNRs: B−6, B, and B+6 dB relative to a baseline SNR (B) that was specific to each initial and final consonant. Consonant-specific SNR levels were established in preliminary experiments (see below). The SNR level (i.e., B−6 dB, B, or B +6 dB) varied randomly across trials.

During each test session, 720 tokens were randomly selected without repetition from the syllable corpus of 9600 tokens. Selection was constrained so that each initial and final consonant was presented exactly 12 times at each SNR. The 12 tokens included four syllables containing each of the three vowels /ɑ/, /i/, and /u/, with each syllable spoken by a different talker. CVC selection was selected based on the initial consonant, vowel, and final consonant so that each consonant was presented at initial and final syllable positions on 36 trials (i.e., 12 at each SNR) and each vowel was presented on 240 trials (80 at each SNR). Each token in the corpus had an equal probability of being presented. Following talker and CVC selection, one of the two CVC tokens spoken by that talker was randomly selected. This procedure resulted in the presentation of 240 tokens (60 from each talker) at each of the three baseline SNR levels (B−6, B, and B+6 dB). Because of the limited set of vowels and the low rate of vowel errors, only consonant identification was scored.

E. Baseline SNR adjustments to equate consonant identifiabilities

Two preliminary experiments were used to estimate the baseline SNRs needed to equate the identifiabilities of different consonants. In the first experiment, four members of the laboratory staff were tested repeatedly (22, 1 h sessions) while consonant-specific SNR levels were adjusted independently in initial and final syllable positions to equate the hit rates for different consonants. This approach revealed significant differences in the response criteria for different consonants. Therefore, signal detection theory was used to com-

TABLE I. SNRs used for the presentation of initial and final consonants in the main experiment.

	b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
Initial	10.3	3.8	6.9	2.5	10.5	⋯	5.7	7.8	10.6	19.6	−2.6	−0.4	0.9	1.1	−7.4	19.7	9.1	12.7	−1.5	6.9	15.5
Final	13.0	11.6	13.8	−0.3	4.4	19.3	17.5	17.5	21.4	32.6	−1.8	4.5	0.4	0.3	−7.9	17.6	11.1	12.3	3.2	7.5	⋯

pute d' for each consonant using hit and false alarm (FA) rates. FAs were defined as consonant responses that occurred when a different consonant had been presented (e.g., a /v/ response when /ð/ was actually presented). An examination of the confusion matrices suggested that FAs did not occur randomly to all possible consonants, but rather occurred in response to a small group of potentially confusable consonants. For example, /p/ FAs occurred in response to /t/, /k/, /h/, /b/, and /g/ but were rarely observed in response to other consonants. On average, significant FAs (exceeding approximately 1% of total responses including hits) were produced by approximately 6.5 other consonants. Therefore, the FA rate was calculated by assuming a pool of 6.5 FA-producing stimuli. While assuming that a fixed pool size of FA-producing consonants was somewhat arbitrary, it eliminated the problem of estimating the FA-pool size for consonants identified using different criterion levels (see discussion below).

The results of the first preliminary study were used to adjust SNR levels to produce an estimated mean d' of 2.6 for each consonant that were used in the second preliminary experiment. Seven young naive volunteers with normal hearing participated after giving informed consent following local Institutional Review Board regulations. Each subject was tested on three occasions for a total of 21, 1 h sessions. Calculated d' values in the second preliminary experiment averaged 2.52 (72.9% correct) for initial consonants and 2.58 (74.8% correct) for final consonants.

The results of the second preliminary experiment were then used to make finer B-level adjustments in order to minimize variations in the identifiabilities of different consonants in the main experiment. In addition, the target d' was reduced to 2.20 (approximately 65% correct) to measure steeper portions of P/S functions. Additional baseline adjustments were also incorporated to equate performance for syllables spoken by different talkers (syllables spoken by female talkers were reduced by 1.8 dB on average) and for syllables containing different vowels (syllables with /i/ were reduced by 3.0 dB and those containing /ɑ/ were reduced by 1.2 dB relative to those containing /u/). The mean SNR values used for each consonant in the main experiment are shown in Table I. B values averaged 6.6 dB for initial consonants and 9.9 dB for independently adjusted final consonants. The range of B values needed to equate identifiabilities spanned 27.1 dB for initial consonants and 40.5 dB for final consonants.

F. Statistical analysis

The data were analyzed with analysis of variance (ANOVA) for repeated measures using the open-source CLEAVE program (T. J. Herron). The original degrees of freedom are reported for each test with the significance levels adjusted using the Box/Greenhouse–Geisser correction for inhomogeneity of variance when appropriate (Greenhouse and Geisser, 1959). In these cases, the original degrees of freedom are reported along with corrected significance (p) levels. Because of the large number of ANOVA comparisons, a relatively strict criterion ($p<0.01$) was used to evaluate statistical significance.

III. RESULTS

A. Quantifying consonant identification

On average, subjects correctly identified 64.1% of initial consonants and 65.2% of final consonants. However, although SNRs had been adjusted to equate d' values for all consonants, hit rates varied substantially for both initial consonants (range: 43.5%–82.5%) and final consonants (range: 43.2%–85.0%). Differences in hit rate were frequently accompanied by corresponding differences in FA rate. For example, the hit rate for /v/ was nearly twice as high as the hit rate for /s/ while the FA rate for /v/ was 8.8 times greater than the FA rate for /s/. Thus, these results revealed systematic differences in consonant response criteria similar to those observed in previous studies (Miller and Nicely, 1955; Wang and Bilger, 1973; Dubno and Levitt, 1981). The covariation in hit and FA rates was the primary reason that signal detection theory was used to analyze the results.

B. Consonant identifiability in noise

Analysis was first performed on initial and final consonants, averaged over vowels, speakers, and final or initial consonants, respectively. Thus, the d' value for each initial consonant reflected 1728 responses from the 16 subjects to 480 different randomly sampled CVC tokens beginning with

TABLE II. Observed d' and beta values for each consonant in initial and final positions.

		b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
Initial	d'	2.08	2.15	2.10	2.14	2.12		2.18	2.23	1.95	2.11	2.10	1.88	2.24	2.33	2.12	2.44	2.32	2.37	2.03	2.36	2.26
	beta	1.72	1.72	1.47	2.15	0.47		2.23	1.82	0.26	2.10	1.27	1.23	1.00	2.90	2.67	1.40	0.87	1.31	2.05	1.36	0.00
Final	d'	2.03	2.14	1.98	2.35	2.29	2.11	2.28	2.41	2.15	1.90	2.17	2.12	2.14	2.18	2.24	2.36	2.20	2.20	2.13	2.15	
	beta	1.27	1.47	1.15	2.59	2.08	2.10	0.85	1.05	−0.07	1.50	2.12	0.91	1.62	1.91	2.88	0.79	1.30	1.64	1.25	1.19	

TABLE III. Estimated SNRs (in dB) needed to produce comparable identifiability ($d'=2.20$) of all consonants in the main experiment. SEM=standard error of the mean.

		b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
Initial	Mean	11.1	4.0	7.3	2.8	11.1		5.8	7.6	12.7	21.9	−2.0	1.1	0.6	0.2	−6.9	13.7	7.8	11.6	−0.8	6.2	15.1
	SEM	0.4	0.4	0.3	0.4	0.4		0.3	0.4	0.5	0.7	0.3	0.4	0.3	0.5	0.6	0.6	0.5	0.5	0.4	0.4	0.4
Final	Mean	14.1	11.9	15.2	−1.4	3.4	19.8	16.8	15.6	22.0	38.6	−1.7	4.9	0.7	0.4	−8.4	15.8	11.1	12.3	3.6	7.8	
	SEM	0.4	0.5	0.5	0.4	0.4	0.5	0.3	0.3	0.6	0.7	0.5	0.4	0.4	0.5	0.5	0.7	0.5	0.4	0.3	0.2	

that consonant and including 576 tokens presented at each SNR (B−6, B, and B+6). Table II shows d' and beta values for each consonant. The observed d' values averaged 2.18 for initial consonants (range 0.37d' units) and 2.19 for final consonants (range 0.48d' units). These values were similar to the targeted performance level ($d'=2.20$).

A repeated measures ANOVA was used to analyze the d' values for the 19 consonants that occurred in both the initial and final syllable positions. This analysis revealed that the mean SNRs of initial and final consonants had been effectively equated by the consonant- and position-specific SNR adjustments made on the basis of preliminary studies (position effect $[F(1,15)=0.45]$). However, the SNR adjustments had not fully corrected for small residual differences in the identifiabilities of different consonants, as reflected in a significant main effect of consonant $[F(18,270)=6.76,\ p<0.0001]$ reflecting differences in mean d' values for different consonants that ranged from 1.96 (/ð/) to 2.38 (/θ/).

The B levels needed to equate consonant identifiability at the $d'=2.20$ level based on the three-point psychometric functions of each consonant are shown in Table III. These values differed from the B values actually used (Table I) by an average of 0.09 dB (range −6.0 to +6.0 dB, excluding the final /ð/ and initial /θ/, the range was −1.9 to +2.3 dB). Standard errors of the SNR means across subjects averaged 0.62 dB. This suggests that the estimated B values needed to equate the identifiability of different consonants shown in Table III were generally precise to within less than 2.0 dB. Estimated B values spanned a range of 28.8 dB for initial consonants and 47.0 dB for final consonants.

B values varied with manner of articulation, with sibilants and affricates requiring the lowest SNRs, followed at increasing SNRs by liquids, plosives, nasals, and non-sibilant fricatives. In general, unvoiced consonants were identified at lower SNRs than voiced consonants. However, B values varied considerably even among consonant sharing similar manner and voicing. For example, among unvoiced plosives, /t/ was identified at SNRs about 10 dB lower than /p/, as was previously noted by Phatak and Allen (2007).

Beta values are also presented in Table II. There were no significant correlations between d' and beta for initial $[r(19)=0.04]$ or final $[r(19)=0.11]$ consonants. ANOVA for repeated measures was used to analyze beta values for the 19 consonants. There was a significant effect of consonant $[F(18,270)=20.79,\ p<0.0001]$ that reflected large differences in response criteria of different consonants. For example, subjects produced more than eight times more /v/ than /ʃ/ false alarms, reflected in respective beta values of 0.09 and 2.40.

Table IV shows the psychometric functions of d' and beta for the different consonants. Both beta and d' increased with increasing SNRs, but there was an insignificant correlation between their rates of increase ($r=0.26$). A repeated measures ANOVA was used to analyze the slopes of the d' P/S for the 19 consonants that occurred in both initial and final positions with consonant and position as factors. The position factor failed to reach significance. However, there was a highly significant main effect of consonant $[F(18,270)=23.68,\ p<0.0001]$. Non-sibilant fricatives (/v/, /ð/, /f/, and /θ/) had P/S functions with substantially shallower slopes (mean 0.08d'/dB) than did other consonant classes including plosives (0.15 d'/dB), sibilant fricatives (0.16d'/dB), nasals (0.16d'/dB), liquids (0.17d'/dB), and affricates (0.19d'/dB). Some of the differences in slope reflected the fact that certain consonants (e.g., the non-sibilant fricatives) remained hard to identify even at the highest SNRs. This was reflected in a highly significant correlation between the B values of each consonant and the slope of its P/S function $[r=-0.70,\ t(18)=5.82,\ p<0.001]$.

C. The identification of leading and trailing consonants

In agreement with previous reports (Wang and Bilger, 1973; Dubno and Levitt, 1981; Redford and Diehl, 1999), initial consonants were collectively detected at 3.3 dB lower SNRs than final consonants. However, this initial consonant advantage was not observed for all consonants: liquids were

TABLE IV. Slopes of psychometric functions of d' and beta for each consonant in initial and final position.

		b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
Initial	d'/dB	0.16	0.24	0.24	0.22	0.13		0.16	0.16	0.12	0.04	0.18	0.21	0.13	0.14	0.15	0.04	0.09	0.15	0.23	0.22	0.15
	Beta/dB	0.11	0.08	0.06	0.18	0.02		0.07	0.09	0.00	0.06	−0.03	0.07	−0.02	0.09	0.14	0.02	0.00	0.04	0.15	0.01	0.07
Final	d'/dB	0.16	0.19	0.16	0.14	0.09	0.18	0.11	0.11	0.09	0.05	0.20	0.20	0.19	0.19	0.08	0.09	0.17	0.17	0.17	0.19	
	Beta/dB	0.05	0.03	0.07	0.08	0.08	0.07	−0.01	0.01	−0.01	0.09	0.03	−0.06	0.05	0.17	0.05	0.06	0.11	0.09	0.09	0.08	

TABLE V. Confusion matrices for initial consonants averaged across subjects, SNRs, and voices. Each consonant was delivered on 1728 trials.

	b	d	g	r	l	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
b	1001	12	24	17	50	2	32	266	16	4	4	2	0	5	15	69	58	7	16	128
d	32	1059	110	28	73	26	28	74	36	27	32	6	4	13	18	15	18	32	30	67
g	36	76	1097	26	62	10	26	78	20	16	22	2	3	5	11	17	41	6	33	141
r	39	35	85	924	206	27	62	114	17	29	20	16	2	12	4	15	27	12	20	62
l	28	9	11	50	1334	19	48	150	28	2	1	1	3	2	0	6	10	1	1	24
n	17	28	33	45	303	917	170	77	18	11	12	3	0	4	7	5	8	4	10	56
m	39	5	18	43	203	106	1073	127	9	6	3	1	1	2	4	7	13	1	10	57
v	57	10	12	37	61	6	15	1360	102	11	0	1	0	0	11	17	8	2	3	15
ð	17	36	6	1	193	2	1	499	861	41	1	1	0	2	54	8	0	2	0	3
z	14	51	26	12	38	10	20	54	30	1155	66	21	7	53	27	27	16	32	25	44
ʤ	12	85	78	24	53	9	23	23	9	45	1056	138	15	7	9	12	12	24	49	45
ʧ	3	7	13	2	7	2	1	4	2	15	253	1222	67	6	3	5	15	42	43	16
ʃ	5	15	9	3	12	2	4	3	1	23	278	451	813	23	5	4	10	12	17	38
s	11	24	15	10	24	6	14	39	10	424	47	52	21	734	42	64	25	58	39	69
θ	1	3	0	0	0	0	0	15	9	3	0	0	0	7	1224	455	3	6	0	2
f	25	2	0	0	4	2	2	120	8	2	1	3	0	5	196	1333	9	1	8	7
p	9	3	7	5	6	1	7	7	0	0	1	0	0	0	1	15	1264	16	58	328
t	22	47	27	10	42	8	24	34	10	39	36	41	7	21	23	25	92	880	165	175
k	12	9	30	3	7	4	7	12	2	9	6	1	1	4	8	26	116	41	1253	177
h	11	0	4	0	7	0	4	13	2	0	2	2	0	1	3	20	176	7	51	1425

detected at lower SNRs in final consonant position. There also remained small position × consonant interactions for d' [$F(18,270)=3.82$, $p<0.001$] after consonant- and position-specific B adjustments. For some consonants, d' values were higher in initial than final syllable position (e.g., /p/, /k/, and /ʃ/) but the opposite pattern was seen for other consonants (e.g., /ð/, /v/, and /s/). Beta values varied significantly with position [$F(1,15)=25.93$, $p<0.0001$] due to stricter criteria for initial than for final consonants (1.58 vs 1.46). Beta measures also showed a significant position × consonant interaction [$F(18,270)=15.89$, $p<0.0001$], reflecting the fact that some consonants showed different response criteria in initial and final syllable positions (e.g., /n/ and /l/).

Tables V and VI show the confusion matrices for the 34 560 initial and final consonants presented in the main experiment. The patterns of confusion resemble those reported in previous studies (Miller and Nicely, 1955; Wang and Bilger, 1973; Phatak and Allen, 2007). Across all consonants, FA rates in excess of 2% were produced by an average of 4.6 alternatives and FAs in excess of 1% were produced by an average of 7.2 alternatives. The remaining 11.8 response alternatives (particularly those differing in multiple phonetic

TABLE VI. Confusion matrices for final consonants averaged across subjects, SNRs, and voices. Each consonant was delivered on 1728 trials.

	b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k
b	1116	94	91	1	11	6	17	29	184	56	1	4	2	2	0	19	19	60	10	6
d	103	1114	135	9	12	10	38	12	97	87	6	36	0	0	2	29	7	16	9	6
g	122	91	1140	5	7	9	9	15	128	54	2	6	0	1	0	18	14	33	17	57
r	27	49	112	925	128	25	44	19	74	14	38	65	19	24	12	30	17	32	37	37
l	51	22	84	123	1026	33	48	42	141	24	7	23	2	6	2	16	16	20	24	18
ŋ	7	7	14	4	42	877	390	281	69	22	0	1	0	1	0	7	3	2	1	0
n	11	15	11	7	14	90	1327	159	54	26	0	3	1	0	0	4	4	2	0	0
m	17	7	6	2	28	78	182	1328	56	15	3	0	0	0	0	2	2	1	0	1
v	64	9	17	3	7	3	7	7	1468	131	3	0	0	0	0	3	4	1	1	0
ð	10	9	3	1	3	1	2	3	714	940	23	1	0	0	0	13	4	0	1	0
z	25	55	62	16	11	10	24	21	70	32	948	144	40	34	61	32	16	21	72	34
ʤ	21	68	66	0	11	13	11	14	38	13	15	1269	75	29	4	14	12	12	27	16
ʧ	9	3	21	3	2	6	6	9	8	6	10	285	1066	146	9	17	9	10	65	38
ʃ	6	6	29	2	4	14	15	7	17	8	13	221	214	1008	40	19	5	18	39	43
s	21	20	57	7	9	11	32	21	42	22	177	82	59	106	746	76	44	39	111	46
θ	1	2	2	0	0	0	1	0	26	19	3	0	0	3	9	1314	337	1	7	3
f	6	2	6	0	1	1	4	2	67	16	3	1	2	0	9	394	1178	9	10	17
p	112	12	21	3	6	1	8	14	18	7	1	2	11	5	0	32	49	1095	114	217
t	6	40	32	5	6	6	15	12	17	8	8	37	37	12	10	46	29	99	1171	132
k	16	12	59	2	8	5	12	11	10	4	4	9	15	8	4	60	48	149	90	1202

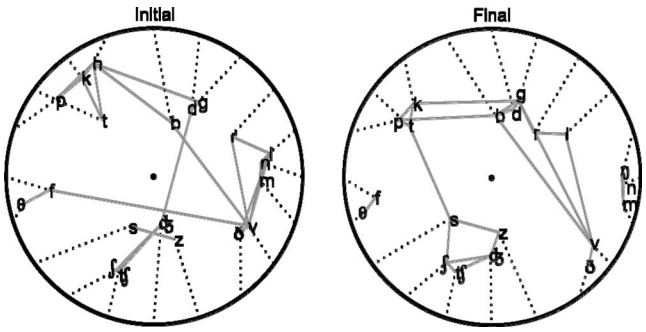

FIG. 2. Confusion clusters for initial and final consonants. Left: Consonant confusion clusters for initial consonants visualized using barycentric clustering. Initial consonants were first equally spaced around the confusion circle reflecting phonetic classification of voicing, manner, and place so that starting locations (clockwise from top center) were /b/, /d/, /g/, /r/, /l/, /n/, /m/, /v/, /ð/, /z/, /ʤ/, /ʧ/, /ʃ/, /s/, /θ/, /f/, /p/, /t/, /k/, and /h/, with /h/ adjacent to /b/. Then, each consonant was displaced from its initial position (dashed line) using iterated weighted averaging of its position with that of every other consonant based on the frequency of confusions between each consonant pair. Consonants that were frequently confused (e.g., /v/ and /ð/) cluster together at intermediate locations that also reflect confusions of one or both consonants in the pair with all other consonants. Gray lines connect the 12.5% of consonant pairs that showed greatest confusions. Right: Consonant confusion clusters for final consonants with starting position (top left-center) /b/, /d/, /g/, /r/, /l/, /ŋ/, /n/, /m/, /v/, /ð/, /z/, /ʤ/, /ʧ/, /ʃ/, /s/, /θ/, /f/, /p/, /t/, and /k/, with /k/ adjacent to /b/.

features from the target consonant) produced very low FA rates. Both FAs (in the columns) and misses (along the rows) were concentrated near the correct-response diagonal, indicating a preponderance of place errors. Other common confusions reflected manner errors conserving place and voice (e.g., /b/-/v/ and /b/-/m/ confusions) and voicing errors conserving place and manner (e.g., /b/-/p/ and /ʧ/-/ʤ/ confusions).

The confusion matrices presented in Tables V and VI reveal that consonants fell into confusable clusters of varying sizes. For some consonants, confusions were largely confined to consonants that shared manner (e.g., nasals in final consonant position). In contrast, other consonants showed confusions with larger consonant sets (e.g., initial /v/ with /b/, /ð/, /m/, /l/, /r/, and /f/). Figure 2 shows the clustering of consonant confusions, visualized using the barycentric algorithm developed by Cohen (2009) to characterize interrelated clusters of internet users. Consonants were initially placed in equidistant positions around a unit circle based on voicing, manner, and place features using the *a priori* consonant ordering shown in Tables V and VI. Unvoiced and voiced consonants were initially positioned on left and right sides of the circle, connected at the bottom across the voiced and unvoiced affricates and connected at the top across voiced and unvoiced plosives. Then, the location of each consonant was computed as an average of its current position, weighted by its hit rate, and the position of every other consonant weighted by number of false responses to that consonant. Two iterations were used to generate the cluster plots shown in Fig. 2. As a result of these iterations, each consonant was displaced from its initial position (dotted lines in Fig. 2) toward the locations of consonants with which it was confused.

Consonant pairs producing the most frequent confusions (top 12.5%, corresponding to total bi-directional confusion rates in excess of approximately 6% of hits) are shown connected by solid gray lines. For example, in final syllable position (Fig. 2, right) nasals were frequently confused with each other but rarely confused with any other consonants, so they moved to an intermediate position near /n/ on the circle edge. In contrast, in the initial syllable position (Fig. 2, left), the voiced sibilant /z/ was frequently confused with the unvoiced sibilant /s/ but also showed moderate rates of confusion with /t/, /ð/, /r/, and /d/, resulting in its displacement toward the center of the circle (Fig. 2, left).

The magnitude of the total displacement of consonants from their initial positions reflects both the initial consonant locations and the observed consonant confusions. For example, if two confusable consonants were initially placed in opposite positions on the confusion circle, both would move substantial distances toward the circle center. In contrast, if two confusable consonants were initially placed in adjacent positions, only small displacements would occur between their locations near the circle boundary. Thus, the overall magnitude of consonant displacement reflects in part the degree to which the initial consonant positions accurately predicted the confusions observed. In the current experiment, initial consonant placements were determined based on place, manner, and voicing features. Feature analysis (see below) revealed that place confusions occurred more frequently than manner confusions, and that manner confusions occurred more frequently than voicing confusions. Therefore, voiced and unvoiced consonants were placed on opposite sides of the circle. Within each voicing group, consonants were then segregated by manner. Finally, within each manner and voicing grouping, consonants were ordered by place of articulation. Consonants whose manner also defined that they are voiced (nasals and liquids) were centered within the voiced-consonant group. At the borders separating different manners, adjacent consonants were positioned to share place and voicing. Similarly, at the borders separating different voicings, adjacent consonants were positioned to share place and manner.

Permutation testing was used to evaluate how accurately the *a priori* consonant positions reflected the observed pattern of consonant confusions. This was accomplished by comparing the total consonant displacement distance with that produced by 10^7 unique random consonant positioning schemes. The *a priori* placement resulted in less displacement than 99.9998% of the random position schemes tested, with no significant differences observed between the magnitude of displacement obtained with the *a priori* positions and the magnitude of displacement obtained with the optimal random placement. This suggests that the initial consonant positions accurately captured the actual confusion clustering evident in the data.

Figure 2 shows the patterns of consonant confusions observed for initial and final consonants. Among initial consonants, the unvoiced plosives /p/, /t/, and /k/ were frequently confused with each other and also with /h/, and thus formed a relatively tight cluster in the upper right of the confusion circle. A relatively high frequency of /p/-/h/ confusions resulted in both /p/ and /h/ being attracted toward a location

intermediate between their initial positions. Because /t/ was occasionally confused with the sibilants and affricates located in the lower right quadrant of the circle, both /t/ and the sibilant-affricate cluster were displaced toward the circle center. These sibilant-affricate confusions were less common for /p/, /k/, and /h/, so these consonants remained closer to the circle circumference than /t/ in the unvoiced-plosive cluster.

Figure 2 shows clear confusion clustering of initial unvoiced plosives, voiced plosives, unvoiced non-sibilant fricatives, sibilants and affricates, voiced fricatives, and liquids and nasals. A few initial consonants (e.g., /v/, /b/, /h/, and /ʤ/) were frequently confused with consonants in more distant clusters. For example, voicing confusions were relatively frequent for /ʤ/-/ʧ/ pairs, and both of these consonants were confused with /ʃ/, /s/, and /z/. Similarly, voicing confusions occurred between /f/, /v/, /b/, and /h/, and manner confusions occurred between /b/ and /v/. In addition, /v/ was frequently confused with both nasals and liquids and was tightly clustered with /ð/, reflecting frequent confusions between this consonant pair. In general, confusion clusters were similar for initial and final consonants with a few exceptions. In the final consonant position, unvoiced plosives formed a tighter cluster in the absence of /h/, and nasals were particularly well discriminated from all other consonants including both voiced non-sibilant fricatives and liquids.

D. Analyzing phonetic features

Figure 3 shows the incidence of single-feature errors (place, manner, and voicing) and combined-feature errors for initial consonants (top) and final consonants (bottom) at the three different SNRs. Table VII shows the average number of consonants in each of the different feature-error categories.

ANOVA for repeated measures was first used to analyze error rates for different phonetic features with position, SNR, and feature (place-only [P], manner-only [M], voicing-only [V], P+M, P+V, M+V, and P+M+V) as factors. This analysis revealed no significant effect of position in the syllable, but revealed the expected main effect of SNR [$F(2,30)=1524.02$, $p<0.0001$]. There was also a large main effect of feature [$F(6,90)=467.28$, $p<0.0001$]: P errors were most frequent (10.2% of trials), followed by P+M (8.6%), M (6.6%), and V (3.3%) errors. The relatively high incidence of P+M errors has been reported previously (Dubno and Levitt, 1981) and reflected in part the fact that there were nearly twice as many consonants in the P+M category as the M category, and more than four times as many consonants in the P+M category as in the P category (Table VII). The remaining combined-feature errors generally occurred on less than 3% of trials except at the most difficult SNR. In addition, there were significant position × feature [$F(6,90)=85.85$, $p<0.0001$] and SNR × feature [$F(12,180)=138.01$, $p<0.0001$] interactions that were explored in additional ANOVAs.

The next ANOVA compared the incidence of different single-feature errors (P, M, and V). This analysis showed a main effect of feature [$F(2,30)=427.76$, $p<0.0001$], primarily reflecting the increased overall incidence of P errors

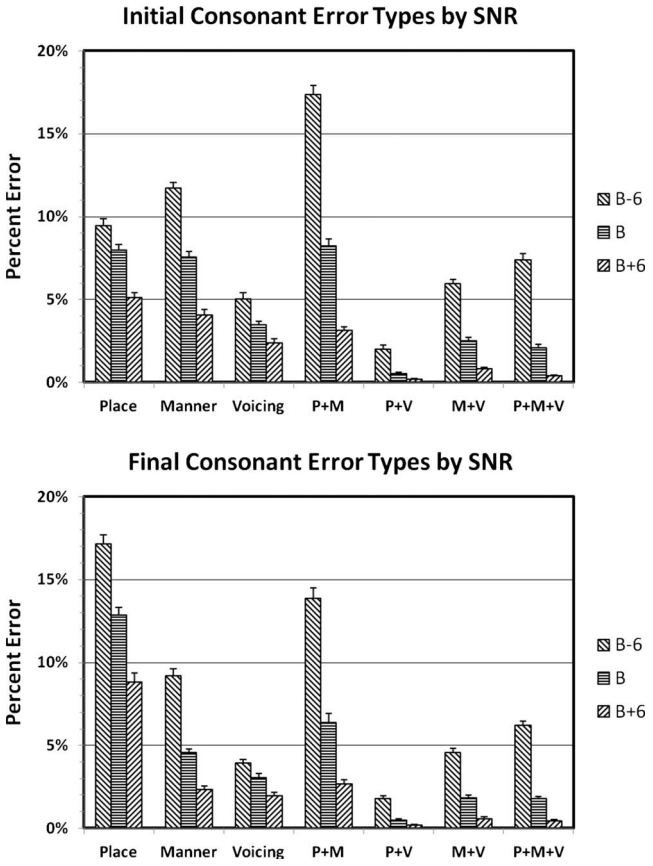

FIG. 3. Error rates for different phonetic features. Place, manner, voicing, P+M=place+manner, etc. Error bars show standard errors of the mean.

in comparison with M and V errors. There was also a main effect of position [$F(1,15)=22.18$, $p<0.0005$] due to fewer single-feature errors for initial than final consonants that was accompanied by a correspondingly greater incidence of combined-feature errors among initial consonants. Finally, there was a highly significant position × feature interaction [$F(2,30)=153.90$, $p<0.0001$], reflecting the fact that M and V errors were more common for initial than final consonants while P errors showed the opposite pattern.

In order to evaluate the changes in feature processing that accompanied increases in SNR, error rates for each feature and position were normalized relative to the total number of errors of each type. The SNR × feature interaction was significant [$F(12,180)=96.13$, $p<0.0001$], indicating that SNR differentially affected error rates for different features. An additional specific analysis examining only single-feature errors showed a significant effect of feature [$F(2,30)=53.02$, $p<0.000\,01$], reflecting the fact that M and V errors declined more rapidly than P errors with increases in SNR. Finally, additional *post-hoc* testing demonstrated that combined-feature errors declined more rapidly than P, M, or V errors ($p<0.001$ for all comparisons).

E. Vowel influences on consonant identifiability

The effects of vowel context on consonant processing were analyzed by comparing consonant identification performance as a function of the accompanying vowel for the 19 consonants that appeared in both initial and final syllable

TABLE VII. Average number of consonants sharing different features and combinations of features with each target consonant.

	Place	Manner	Voicing	P+M	P+V	M+V	PMV
Initial	1.30	2.50	0.70	5.30	1.10	2.60	5.50
Final	1.30	2.50	0.70	5.55	0.90	2.60	5.40

positions. The results were analyzed using repeated measures ANOVA with subjects, vowel, position, and consonant as factors. The SNRs of syllables containing different vowels had been adjusted in preliminary studies to reduce vowel-specific differences in overall mean syllable identifiability. As a result, the vowel main effect failed to reach significance [$F(2,30)=2.62$, $p<0.10$]. However, there was a highly significant vowel×position interaction [$F(2,30)=42.40$, $p<0.0001$]. This reflected the fact that initial consonants were identified more accurately than final consonants in syllables containing /ɑ/, whereas final consonants were identified more accurately than initial consonants in syllables containing /i/. Initial and final consonants were identified with comparable accuracy in syllables containing /u/.

There was also a highly significant vowel×consonant interaction [$F(36,540)=45.67$, $p<0.0001$], as shown in Table VIII. Separate ANOVAs analyzing vowel effects for each of the 19 consonants showed that d' measures for 16 of the 19 consonants (excepting /ʧ/, /d/, and /k/) were significantly influenced by the accompanying vowel. Vowel-related changes in d' ranged from 0.37 to 1.46 for different consonants and were generally highly significant [the average over all consonants was $F(2,30)=44.23$, $p<0.0001$]. For example, /m/ was identified much more accurately in syllables containing /ɑ/ than in syllables containing /i/ or /u/, producing vowel-related d' differences of 1.46. This would be equivalent to a 9.1 dB change in SNR estimated based on the P/S slope for /m/ ($0.16d'$/dB).

Figure 4 shows identification performance for initial and final consonants displayed on a vowel triangle with the distance of each consonant from each vowel apex inversely scaled by its d' value when presented with that vowel. Consonants with similar d' values for all vowels are located near the center of the triangle (e.g., /ʧ/). Figure 4 shows that initial nasals were preferentially identified with a /ɑ/, and better identified with /u/ than /i/, while final nasals were preferentially identified with a /ɑ/, but better identified with /i/ than /u/. Initial non-sibilant fricatives were poorly identified with /i/ and therefore clustered on the /ɑ/-/u/ boundary. In final syllable position, the fricatives /v/, /f/, and /θ/ showed less vowel specificity, while /ð/ was very poorly identified with both /a/ and /i/. Similarly, /t/ was poorly identified with /ɑ/ in both consonant positions, while /r/ was poorly identified with /u/ as an initial consonant and poorly identified with /ɑ/ as a final consonant.

Confusion matrices for initial and final consonants in syllables containing different vowels are included in Supplementary Tables 1–6 (see Supplementary material). Figure 5 shows the cluster confusion analyses as a function of vowel. An examination of the vowel effects on the confusion matrices clarifies the nature of the vowel-related changes in consonant confusions. For example, the confusion clusters of nasal consonants (/m/, /n/, and in final consonant position, /ŋ/) show that initial nasals are moderately confused with each other and also with the liquid /l/ when presented with /ɑ/. In contrast, when initial nasals were presented with /i/ or /u/ they were more often confused with each other (i.e., they are closer together in the confusion circle) and also confused with /l/, /r/, /v/, and /ð/. Similarly, final nasals presented with /ɑ/ are tightly clustered near the circle periphery, reflecting a low incidence of confusions with non-nasal consonants. They are also widely spaced from each other, reflecting enhanced within-nasal discrimination. Final nasals presented with /i/ remain distinct from other confusion clusters, but show increased within-nasal confusions. Final nasals presented with /u/ show a high frequency of within-nasal confusions as well as additional confusions with /l/ and /v/. Similarly, the poor discriminability of the final /ð/ when presented with /ɑ/ and /i/ can be seen to reflect its very poor discriminability from /v/. The discriminability of /v/ and /ð/ is improved when these consonants are presented with /u/.

Additional analyses were performed to examine the effects of vowel context on the processing of consonants with different manners, with vowel, position, and manner as factors. This analysis revealed a highly significant interaction between vowel and manner [$F(8,120)=125.44$, $p<0.0001$]. Sibilants, liquids, and plosives were identified most accurately in syllables containing /i/, non-sibilant fricatives were identified most accurately in syllables containing /u/, and nasals were identified most accurately in syllables containing /ɑ/.

The next analysis examined vowel context effects on

TABLE VIII. d' scores for initial and final consonants in syllables containing different vowels.

		b	d	g	r	l	ŋ	n	m	v	ð	z	ʤ	ʧ	ʃ	s	θ	f	p	t	k	h
Initial	/ɑ/	2.24	1.90	2.25	2.23	2.72		3.04	3.12	1.97	2.35	1.88	1.67	2.27	2.16	1.96	2.41	2.33	2.16	1.57	2.41	2.28
	/i/	2.20	2.31	2.13	2.51	2.00		1.45	1.62	1.60	1.52	2.42	1.86	2.08	2.41	2.21	2.01	1.85	2.94	2.28	2.38	2.38
	/u/	1.94	2.28	2.04	1.74	1.78		2.27	2.01	2.27	2.48	1.95	1.98	2.22	2.34	2.06	3.08	2.95	2.17	2.20	2.19	2.03
Final	/ɑ/	2.34	2.18	1.98	1.82	2.33	2.56	2.86	3.32	2.11	1.46	1.97	1.88	2.02	1.86	1.81	2.08	1.92	2.00	1.77	1.88	
	/i/	2.01	2.01	2.42	2.69	2.87	2.32	2.02	2.25	2.05	1.61	2.38	2.32	2.09	2.33	2.43	2.27	2.20	2.38	2.28	2.31	
	/u/	1.82	2.21	1.78	2.49	1.90	1.58	2.16	1.97	2.35	2.82	2.20	2.19	2.25	2.31	2.02	2.81	2.40	2.12	2.31	2.26	

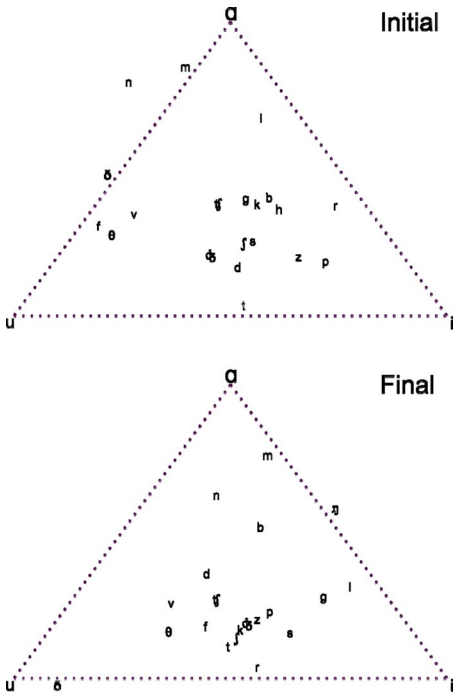

FIG. 4. (Color online) Consonant-vowel interactions shown on a triangular map of vowel space with the three vowels presented in the current experiment at the apices. The distance of each consonant from each vowel apex is inversely related to the d' measure for that vowel-consonant combination. Top=initial consonants; bottom=final consonants. The few consonants without significant vowel preferences (e.g., /ʧ/) are found near the center of the triangle. Consonants with strong vowel preferences lie closer to the preferred vowel (e.g., nasals are close to /ɑ/).

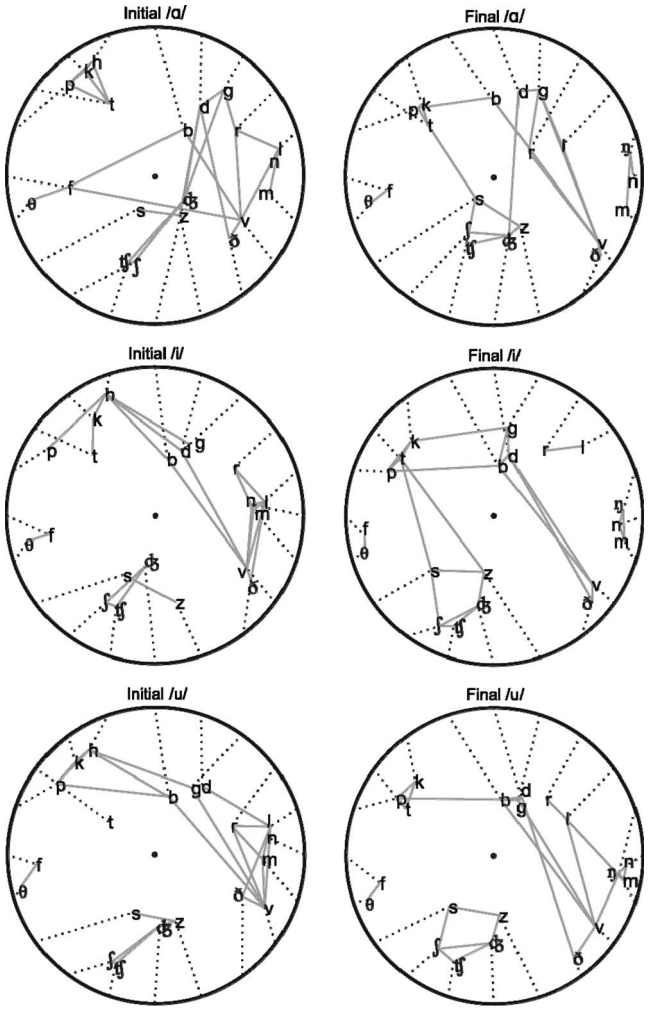

FIG. 5. Confusion clusters for initial (left) and final (right) consonants as a function of vowel. See Fig. 2 for further details.

place of articulation. There was a highly significant vowel $\times$place interaction $[F(4,60)=34.58,\ p<0.0001]$. Front consonants were perceived most accurately in syllables containing /ɑ/, middle consonants were perceived most accurately in syllables containing /u/, and back consonants were perceived most accurately in syllables containing /i/. There was also a significant interaction between vowel, syllable position, and place $[F(4,60)=6.02,\ p<0.002]$, indicating that vowel context effects on place differed at initial and final syllable positions.

Finally, there was also a significant vowel$\times$voicing interaction $[F(2,30)=57.97,\ p<0.0001]$. Unvoiced consonants were most accurately identified in syllables containing /u/, whereas voiced consonants were most accurately identified in syllables containing /ɑ/. There was no significant difference in vowel effects on the voicing of consonants at initial and final syllable positions $[F(2,30)=0.93, \text{NS}]$.

Vowel effects on feature processing were analyzed by evaluating the relative incidence of various types of feature errors in the vowel-specific confusion matrices presented in Supplementary Tables 1–6 (see supplementary material). An omnibus ANOVA indicated a highly significant interaction between vowel and error-type (P, M, V, P+M, P+V, M+V, and PMV) $[F(12,180)=33.70,\ p<0.0001]$ that was explored in additional ANOVAs. There were no significant main effects of vowel on the incidence of V errors $[F(2,30)=0.12]$ and only a trend toward a reduction in M errors with the vowel /i/ $[F(2,30)=3.36,\ p<0.06]$. However, there was a highly significant vowel main effect on the incidence of P errors $[F(2,30)=78.52,\ p<0.0001]$. P errors were increased in syllables containing /i/ relative to those containing /ɑ/ and /u/. P errors also showed a significant vowel$\times$position interaction $[F(2,30)=21.03,\ p<0.0001]$. Inter-vowel differences increased in the final syllable position due in large part to the overall increase in P errors in the final syllable position described above. An analysis of combined-feature errors also revealed a significant vowel main effect $[F(2,30)=53.81,\ p<0.0001]$ on the relative incidence of P+M errors. P+M errors were reduced in syllables containing /i/ relative to those containing /ɑ/ and /u/, without a significant vowel$\times$position interaction.

F. Interactions in identifying initial and final consonants

We evaluated the effects of consonant context using repeated measures ANOVAs with subject, vowel, consonant position, and matching of initial and final consonant features as factors. Separate analyses were performed for manner, place, and voicing features, each evaluated at the $p<0.01$ level of significance. There were no significant effects of matching on manner $[F(1,15)=1.58]$ or voicing $[F(1,15)=3.99,\ p<0.07]$. However, there was a highly significant

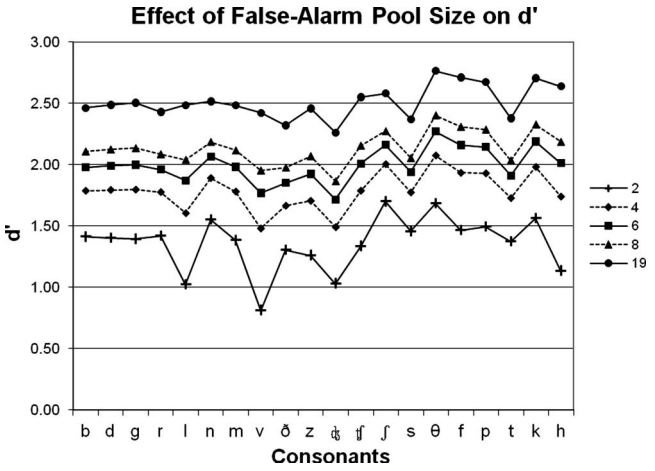

FIG. 6. d' values for initial consonants calculated with different false-alarm-pool sizes.

effect of matching on place $[F(1,15)=41.70,\ p<0.0001]$: subjects showed increased accuracy when initial and final consonants differed in place of articulation. This effect did not differ as a function of consonant position.

IV. DISCUSSION

A. Quantifying consonant identification

Both hit rate and d' were examined as metrics of consonant identification performance. Hit rates were confounded by large differences in response criteria associated with nearly ten-fold differences in the FA rates of different consonant responses. Miller and Nicely (1955) found similar results: up to ten-fold differences in the FA rates for different consonants at certain SNRs. Therefore, d' appeared to be a better metric for evaluating consonant identifiability because of its correction for differences in response criteria.

Unfortunately, there is no standard method of calculating d' under conditions where many potentially confusable signals are presented in noise. In the standard signal detection theory paradigm, the only trials presented are signal-plus-noise (associated with hits) or noise-alone (associated with FAs). However, in the current study, FAs assigned to a particular consonant response were usually generated by a number of different consonants with which it shared phonological and acoustic features. For example, /v/ FAs occurred on 29% of /ð/ trials, but on less than 0.2% of ʃ trials. These results suggest that some, but not all, consonants could be sensibly considered plausible FA-producing distractors. In our analysis we included only the plausible FA-producing consonants in the FA pool used to determine N, the divisor for the FA rate calculation. As noted in Sec. II, the size of the FA pool was estimated to be 6.5 consonants based on preliminary studies. This FA-pool size corresponded to consonants producing FA rates in excess of approximately 1.3% and was slightly greater than a FA-pool size (5.3) based on consonants that shared two phonetic features (e.g., manner and voicing) with each response.

Figure 6 shows d' values for initial consonants calculated using different FA-pool sizes ranging from 2 to 19 consonants. Increasing the FA-pool size decreases in FA rate and correspondingly increases the d'. The increase in d' values is greatest for consonants with high FA rates, e.g., /v/, /l/, and /h/ in Fig. 6. Although the d' curves differ somewhat at extreme FA-pool sizes, they have similar shapes over intermediate FA-pool sizes, indicating that the current results are largely insensitive to changes in the FA-pool size over realistic intermediate ranges.

B. Consonant identifiability in noise

The current experiment assessed listeners' abilities to identify 20 initial and 20 final consonants using spoken responses that included the majority of potential consonant confusions. To assure that confusion patterns were representative of average articulation patterns, syllables were randomly sampled from a set of 9600 tokens with equal sampling of each talker, vowel, and consonant at each SNR. Because the results of the current study are based on the random sampling of a very large CVC token set, they provide insights into the identification of American English consonants in typically articulated CVCs.

It is well known that some consonants are much easier to identify in noise than others (Wang and Bilger, 1973; Dubno and Levitt, 1981; Dubno *et al.*, 1982; Helfer and Huntley, 1991; Gelfand *et al.*, 1992; Benkí, 2003; Cutler *et al.*, 2004; Phatak and Allen, 2007; Phatak *et al.*, 2008). Previous studies suggest that SNR levels must be adjusted by 18 dB (Miller and Nicely, 1955) to 24 dB (Phatak and Allen, 2007) to equate hit rates across all consonants in 16-consonant sets. We found that even larger differences in SNR were needed to equate consonant identifiability in a 20-consonant set of initial (28.8 dB) and final (47.0 dB) consonants. The increased range of SNRs needed to equate consonant identifiability likely reflects in part the increased number of possible consonant confusions. The addition of the consonants /l/, /r/, /ʧ/, /h/, and /ŋ/ in the 20-consonant sets increased potential confusions for many consonants (particularly /ð/), reducing their discriminability and correspondingly increasing their required baseline SNR levels.

The pattern of results shown in Table III suggests that a subset of consonants (sibilants, affricates, liquids, and some plosives) provides the majority of consonant information in difficult listening situations (e.g., absolute SNRs less than 6 dB). Consonant information becomes available from less easily identified consonants only at higher SNRs. The results also revealed surprisingly large differences in identifiability of different consonants within the same phonetic class. For example, among initial plosives, baseline SNRs for /p/ and /t/ differed by 12.4 dB, while baseline SNRs for /b/ and /d/ differed by 7.1 dB. Phatak and Allen (2007) observed similar differences in the relative identifiability of these consonant pairs. These differences in plosive identifiability may reflect the greater high-frequency spectral energy in /t/ and /d/ plosive bursts with respect to /p/ and /b/ (Halle *et al.*, 1957), resulting in less effective masking by low-frequency speech-spectrum noise of the sort used in the current study.

Although the average psychometric functions obtained in this experiment had mean slopes that were similar to those reported in previous studies using nonsense CVCs (Boo-

throyd and Nittrouer, 1988), there were large differences in the P/S slopes of different consonants. Not surprisingly, harder-to-identify consonants (i.e., those that required higher baseline SNRs) had shallower slopes. This was due in part to the fact that harder-to-identify consonants (e.g., non-sibilant fricatives) remain difficult to identify even at high SNRs (Wiley *et al.*, 1979). In contrast, easier-to-identify consonants (e.g., sibilants) showed steeper psychometric functions. The correlation between baseline SNRs and P/S slopes suggests that consonant information from easy-to-identify consonants continues to provide the majority of consonant information in conversational speech at moderate noise levels.

C. The identification of leading and trailing consonants

Overall, we found that initial consonants were identified at consistently lower SNRs than final consonants, replicating the results of previous studies using smaller token sets (Wang and Bilger, 1973; Dubno and Levitt, 1981; Redford and Diehl, 1999; Benkí, 2003). These differences likely reflected slight increases in the amplitude and duration of initial consonants with respect to final consonants (Redford and Diehl, 1999). However, the magnitude of the initial consonant advantage differed for different consonant manners. Nasals showed particularly large initial consonant advantages as in previous reports (Dubno and Levitt, 1981; Repp and Svastikula, 1988). In contrast, liquids showed a final consonant advantage that has likewise been previously noted (Cutler *et al.*, 2008). Although d' had been equated for initial and final consonants, we found a significant position-related difference in response criteria: initial consonants were identified with stricter criteria than final consonants. In part, this may have reflected differences in the 20th consonant used in initial and final consonant sets: /h/ was a more potent FA-generator than /ŋ/.

Cluster analysis was used to visualize the confusions among consonants in initial and final syllable positions. This analysis revealed that the particular *a priori* categorization of consonants based on voice, manner, and place features provided a valid starting point for consonant-confusion analysis. Although consonants generally segregated into confusable clusters sharing manner and voicing, cluster analysis revealed considerable overlap between nasals and liquids and also between sibilants and affricates. Distant clusters were connected through "hub" consonants (e.g., /h/, /v/, /b/, and /ʤ/) that were frequently confused with consonants with distant clusters characterized by different manner and/or voicing features. Cluster analysis showed that confusion clusters were similar in initial and final consonant positions, with some clusters better segregated in initial syllable position and other clusters better segregated in final syllable position.

D. Analyzing phonetic features

Phonetic features differed in intelligibility. Among single-feature errors, P confusions were most common, followed by M and V confusions. Moreover, P confusions persisted to a greater degree at higher SNRs than M and particularly V confusions as in previous reports (Grant and Walden, 1996). We also found a high incidence of P+M errors, particularly at low SNRs, consistent with previous studies (Dubno and Levitt, 1981). P+M errors and other combined-feature errors declined more rapidly than did single-feature errors with increasing SNR.

The relative salience of different phonetic features also differed significantly as a function of syllable position. Combined feature errors were more common in initial than final consonant position (possibly because of the inclusion of /h/), while single-feature errors showed the opposite pattern. Moreover, among single-feature errors, V and M confusions were more common among initial consonants while P confusions were more common among final consonants. Differences in the articulation of initial and final consonants may have contributed to these results (Redford and Diehl, 1999). In addition, there are differences in the perceptual operations associated with the identification of initial and final consonants. Subjects begin extracting information about the initial consonant prior to receiving vowel information and must identify the consonant and vowel concurrently. The parallel analysis of the initial consonant and vowel may increase multi-feature as well as M and V errors that depend on the accurate analyses of the vowel nucleus. In contrast, the increase in place errors in final consonant position may have reflected the relative reduction in M and V errors in the final position subsequent to vowel identification. Alternatively, uncertainty in the onset time of the initial consonant and the resulting failure to accurately detect its onset might have given rise to increased multi-feature errors. The temporal uncertainty of final consonant delivery would have been reduced because of the additional timing cues provided by the vowel and the start of the vowel-consonant transition.

E. Vowel influences on consonant identifiability

Overall differences in consonant identifiability in syllables containing different vowels had been eliminated by small vowel-specific adjustments in syllable intensity. However, different vowel effects were observed in initial and final syllable positions. Initial consonants were more accurately identified in syllables containing /ɑ/ while final consonants were more accurately identified in syllables containing /i/, as observed in previous reports (Halle *et al.*, 1957; Wang and Bilger, 1973; Redford and Diehl, 1999).

Further analysis revealed that vowel space was not independent of consonant space: highly significant vowel × consonant interactions for 85% of consonants. These interactions were generally similar for consonants sharing similar manner. For example, nasals were more accurately identified with /ɑ/, as observed previously (Dubno and Levitt, 1981; Repp and Svastikula, 1988), while sibilants, plosives, and liquids were most accurately identified in syllables containing /i/, and non-sibilant fricatives were most accurately perceived in syllables containing /u/. These effects may have reflected more salient formant transitions in preferred consonant-vowel pairs. Vowels also interacted with place of articulation. Front consonants were perceived most accurately in syllables containing the back vowel /ɑ/, middle consonants were perceived most accurately in syllables contain-

ing /u/, and back consonants were perceived most accurately in syllables containing the front vowel /i/. These effects may have reflected the increased salience contrasts associated with greater tongue movements and presumably longer duration formant changes.

Confusion-cluster analysis was used to visualize vowel-consonant interactions. This analysis revealed variations in the clarity of different phonetic cues as a function of the vowel-nuclei of syllables. Vowel-related improvements in consonant identification were usually accompanied by a reduction in place confusions within clusters as well as reductions in confusions with distant clusters. We also found systematic differences in the overall frequency of P confusions as a function of vowel: P confusions increased with /i/ in comparison with /ɑ/ or /u/, whereas P+M errors showed a corresponding decrease. One possible explanation is that /i/ facilitated manner processing and thus converted potential P+M errors into P errors.

F. Interactions in identifying initial and final consonants

We found evidence of significant interactions between the processing of initial and final consonants in CVCs: both initial and final consonants were identified more accurately when they differed in place of articulation. No such interactions were found for voicing or manner. One possible explanation is because the analysis of place is more difficult than the analysis of manner or voicing, particularly at the final consonant position, it requires more time for analysis. This would increase the possibility of consonant-consonant interactions in place processing over the relatively long inter-consonant intervals of CVCs (Christiansen and Greenberg, 2008). The reduced accuracy found when initial and final consonants share place may thus reflect adaptation of the place feature at relatively short intervals.

V. CONCLUSIONS

Consonant identification in CVCs presented in noise is a highly complex process that is subject to many contextual influences. Different consonants are identifiable over widely different SNR ranges, suggesting that a subset of consonants contributes disproportionately to speech understanding during difficult listening situations. Cluster analysis of confusions revealed that consonants are segregated into perceptually distinct clusters, exhibiting high intracluster confusion rates and generally low rates of intercluster confusions. Perceptual clusters typically consist of consonants with the same manner and voicing, with the exception of nasal-liquid and sibilant-affricate clusters. Highly significant differences were observed in the identification of initial and final consonants, and consonant identification was strongly modulated by the accompanying vowel. In addition, consonant-consonant interactions were observed in the processing of initial and final consonant place features. These results suggest that consonant processing within CVCs is a complex, non-linear process that is subject to contextual modulation by consonant position, intervening vowels, and even other consonants.

[1]Pure tone thresholds (over 500, 1000, and 2000 Hz) averaged 6.71 dB HL (±1.93) in the left ear and 4.80 dB HL (±1.52) in the right, with average 8000 Hz thresholds of 10.94 dB HL (±2.95) and 6.88 dB HL (±2.58) in the left and right ears, respectively.

[2]Experimenter response transcription was used in preference to subject transcription to maintain the naturalness of the listening task, minimize procedural learning effects, and avoid scoring biases that might be introduced by listeners untrained in the use of the phonetic alphabet.

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

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

Christiansen, T. U., and Greenberg, S. (**2008**). "Cross-spectral synergy and consonant identification," J. Acoust. Soc. Am. **123**, 3850.

Cohen, J. (**2009**). "Graph twiddling in a MapReduce world," Comput. Sci. Eng. **11**, 29–41.

Cutler, A., Garcia Lecumberri, M. L., and Cooke, M. (**2008**). "Consonant identification in noise by native and non-native listeners: Effects of local context," J. Acoust. Soc. Am. **124**, 1264–1268.

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

Dubno, J. R., Dirks, D. D., and Langhofer, L. R. (**1982**). "Evaluation of hearing-impaired listeners using a nonsense-syllable test. II. Syllable recognition and consonant confusion patterns," J. Speech Hear. Res. **25**, 141–148.

Dubno, J. R., and Levitt, H. (**1981**). "Predicting consonant confusions from acoustic analysis," J. Acoust. Soc. Am. **69**, 249–261.

Gelfand, S. A., Schwander, T., Levitt, H., Weiss, M., and Silman, S. (**1992**). "Speech recognition performance on a modified nonsense syllable test," J. Rehabil. Res. Dev. **29**, 53–60.

Grant, K. W., and Walden, B. E. (**1996**). "Evaluating the articulation index for auditory-visual consonant recognition," J. Acoust. Soc. Am. **100**, 2415–2424.

Green, D. M., and Swets, J. A. (**1974**). *Signal Detection Theory and Psychophysics* (Robert E. Krieger, Huntington, NY).

Greenhouse, S. W., and Geisser, S. (**1959**). "On methods in the analysis of profile data," Psychometrika **24**, 95–112.

Halle, M., Hughes, G. W., and Radley, J. P. (**1957**). "Acoustic properties of stop consonants," J. Acoust. Soc. Am. **29**, 107–116.

Helfer, K. S., and Huntley, R. A. (**1991**). "Aging and consonant errors in reverberation and noise," J. Acoust. Soc. Am. **90**, 1786–1796.

Helfer, K. S., and Wilber, L. A. (**1990**). "Hearing loss, aging, and speech perception in reverberation and noise," J. Speech Hear. Res. **33**, 149–155.

Levitt, H., and Resnick, S. B. (**1978**). "Speech reception by the hearing-impaired: Methods of testing and the development of new tests," Scand. Audiol. Suppl. **6**, 107–130.

Mann, V. A., and Repp, B. H. (**1981**). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. **69**, 548–558.

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

Müsch, H., and Buus, S. (**2001a**). "Using statistical decision theory to predict speech intelligibility. I. Model structure," J. Acoust. Soc. Am. **109**, 2896–2909.

Müsch, H., and Buus, S. (**2001b**). "Using statistical decision theory to predict speech intelligibility. II. Measurement and prediction of consonant-discrimination performance," J. Acoust. Soc. Am. **109**, 2910–2920.

Phatak, S. A., and Allen, J. B. (**2007**). "Consonant and vowel confusions in speech-weighted noise," J. Acoust. Soc. Am. **121**, 2312–2326.

Phatak, S. A., Lovitt, A., and Allen, J. B. (**2008**). "Consonant confusions in white noise," J. Acoust. Soc. Am. **124**, 1220–1233.

Redford, M. A., and Diehl, R. L. (**1999**). "The relative perceptual distinctiveness of initial and final consonants in CVC syllables," J. Acoust. Soc. Am. **106**, 1555–1565.

Repp, B. H., and Svastikula, K. (**1988**). "Perception of the [m]-[n] distinction in VC syllables," J. Acoust. Soc. Am. **83**, 237–247.

See supplementary material at http://dx.doi.org/10.1121/1.3293005 E-JASMAN-127-034003 for confusion matrices for individual vowels.
Wang, M. D., and Bilger, R. C. (**1973**). "Consonant confusions in noise: A study of perceptual features," J. Acoust. Soc. Am. **54**, 1248–1266.
Wiley, T. L., Strennen, M. L., and Kent, R. D. (**1979**). "Consonant discrimination as a function of presentation level," Audiology **18**, 212–224.
Woods, D. L., and Elmasian, R. (**1983**). "The habituation of human event-related brain potentials elicited by speech sounds and tones," Soc. Neurosci. Abstr. **9**, 365.
Woods, D. L., Yund, E. W., and Herron, T. J. (**2010**). "Measuring consonant identification in nonsense syllables, words and sentences," J. Rehab. Res. Dev. **47** (In Press).