Consonant confusions in noise: a study of perceptual features

Marilyn D. Wang*† and Robert C. Bilger

Bioacoustics Laboratory, Eye and Ear Hospital and School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15213

(Received 27 February 1973; revised 26 June 1973)

Consonant confusion matrices were obtained for four sets of CV and VC nonsense syllables presented both in quiet and in the presence of a masking noise. A sequential method of partitioning transmitted information for confusion matrices was developed and used to test the hypothesis that when the internal redundancy of feature systems is taken into account, certain articulatory and phonological features of consonants consistently account for transmitted information better than other, closely related, features. Results of the analyses indicate that for most confusion matrices several feature systems can be shown to account equally well for transmitted information, and that across syllable sets and listening conditions, there is little consistency in the identification of perceptually important features. The implication of these findings with respect to the existence of natural perceptual features for consonants is discussed.

Subject Classification: 9.5.

INTRODUCTION

It is difficult to find a perceptual study of consonants in which the relationships among stimuli are not analyzed in terms of phonological or articulatory features. Although most researchers share the view that feature perception plays a role in the discrimination of speech sounds, the feature analyses which they perform tend to fall into two classes depending on whether it is individual features or feature systems which are to be evaluated.

When data are presented in the form of confusion matrices, feature analysis usually consists of specification of a set of distinctive features and determination of the relative importance of these features in terms of a performance measure such as error rate or percentage of information transmitted. Feature analyses may also be based on triangular error matrices compiled from same–different responses to pairs of stimuli. In either case, it is assumed that a high level of performance on a particular feature is indicative of the perceptual relevance of the feature.

As a result of such analyses we know that individual features differ significantly in intelligibility,[1–6] and that the relative importance of features varies as a function of phonetic context[4,5] and conditions of signal distortion.[1] Although pairs of sounds which differ by the same number of features are not always equally discriminable,[6,7] on the average, as the number of feature differences between two sounds increases, discriminability improves.[6–10]

The second approach to feature analysis is exemplified by studies in which alternative feature systems are compared or in which a method such as multidimensional scaling is used to derive a set of perceptual features empirically. Here, emphasis is placed on the ability of the feature system as a whole to predict certain relationships among the stimuli, rather than on the levels of performance associated with the individual features within the system. For example, Wickelgren[11] used three articulatory feature systems to predict the relative frequency of consonant confusions in short-term memory,[12] and compared the systems in terms of the percentage of predictions confirmed by the data. Singh, Woods, and Becker[16] used a multiple-regression procedure to compare the ability of several feature systems to predict scaled interpoint distances among stimuli. In neither study was level of performance on specific features a factor in the evaluation of a feature system.

Within the last ten years, there has been a steady increase in the use of multidimensional scaling as a technique for discovering perceptual features.[13–20] Scaling is particularly attractive because it provides a method for obtaining non-arbitrary perceptual dimensions and because, in addition, it provides a measure of goodness-of-fit for the obtained set of dimensions.

In general, the results of scaling analyses support the notion that articulatory features, either singly, or in interactive combination, do, in fact, constitute valid perceptual dimensions. This conclusion, however, reflects the fact that scaling solutions invariably require considerable interpretation, and that linguistic features, as well as traditional articulatory features, provide a ready basis for that interpretation. Obtained spatial configurations have not been sufficiently stable across experiments and scaling methods to suggest definition of new and totally unanticipated perceptual features. Neither have they provided an empirical basis for preferring a particular feature definition over other closely related definitions. Since scaling methods are sensitive to non-perceptual considerations such as the choice of a distance metric[13–16] and differences in data collection procedures,[16] this instability of solutions is not surprising.

A question which has not been systematically explored, and which represents an untestable assumption in scaling analyses, is whether or not a natural set of perceptual features actually exists. What has been demonstrated repeatedly is that it is possible to describe performance in terms of sets of relatively arbitrary features. But if natural perceptual categories exist, and if features are to be considered more than descriptive labels for sets of related acoustic cues, then there must exist a set of features which, according to some criterion, is optimal for describing the perceptual relationships among stimuli.

The problem is one of demonstrating that if a great many potentially relevant features, including features suggested by scaling analyses, were combined in a large, highly redundant, feature system, that a particular subset of these (those corresponding most closely to the natural perceptual categories) would consistently emerge as optimal in accounting for perceptual data. The relative importance of the features within this subset need not remain invariant in all phonetic and experimental contexts, but the particular features belonging to the subset should not change substantially from one experiment to another.

Klatt[22] has argued that if a feature represents a natural set of perceptual categories, then perceptual confusions should be higher for sounds identical with respect to the feature than for sounds which differ on the feature. If we extend this by arguing that correct feature identification should contribute significantly to correct identification of the stimulus, we make explicit the logical basis for feature analyses such as those of Miller and Nicely[1] and Singh and Black.[2] That this has been recognized implicitly for some time is evidenced by the large number of studies which have taken this approach to feature analysis.

Although level of performance provides a logically sound basis for comparing individual features, particularly closely related features, it has not previously been used as a criterion for deriving or evaluating a feature system. This is due in part to the lack of a strategy for combining measures of feature performance. In Sec. II is described a sequential method of analyzing transmitted information which systematically identifies, from among a number of features, those on which performance is high, and which takes the internal redundancy of the features into account in doing so. This is accomplished by partialing out, in each iteration, the effects of features identified in earlier iterations. The analysis also allows us to determine what proportion of the total transmitted information is accounted for by the features identified as perceptually important. The procedure, and the rationale behind it, may be loosely interpreted as the information analogue of a stepwise multiple-regression analysis.

Table I. Composition of four syllable sets, each containing 48 syllables formed by combining 16 consonants with /i,ɑ,u/.

Syllable Set	Consonant phonemes
CV-1	/p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, ʒ, tʃ, dʒ/
VC-1	/p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, ʒ, tʃ, dʒ/
CV-2	/p, b, tʃ, dʒ, l, r, f, s, v, z, m, n, h, hʷ, w, j/
VC-2	/p, b, g, ŋ, m, n, f, θ, s, ʃ, v, ð, z, ʒ, tʃ, dʒ/

The aim of the study reported here was to determine, using this analysis, which features best account for performance in a consonant discrimination task, and to see whether the same features are best in a variety of contexts and listening conditions.

I. METHOD

A. Speech System

Stimuli for the experiments consisted of 129 recorded nonsense syllables. The items represent all phonologically permissible CV and VC syllables which can be formed by combining English consonants, liquids, and glides, with the vowels /i, ɑ, u/. An adult male speaker tape recorded five repetitions of each syllable, and the production judged best, using criteria of naturalness, loudness, and duration, was selected to be the stimulus. The stimuli were then rerecorded permanently on the vocabulary drum of a Cognitronics Speechmaker. This drum is organized into 189 addressable locations, each of which may contain a single speech sample of 511-msec maximum duration.

The Speechmaker is controlled by a special programming device which operates in conjunction with an ASR 33 Teletype. Briefly, its operation is as follows. Information for a single trial is coded on paper tape and read from the Teletype. The device then selects the appropriate stimulus from the Speechmaker, controls the timing of warning, observation, and answer intervals, receives consonant identification responses from up to four listeners, and provides feedback concerning the correct response. Trial information, including the listeners' responses, is then transmitted in a special code to the Teletype punch.

B. Syllable Sets

Although the complete stimulus set contained 24 consonants in CV form and 19 in VC form, it was necessary for instrumental reasons to restrict the number of response alternatives on a given trial to 16. We therefore developed four syllable sets, which are shown in Table I. The use of four syllable sets also permitted us to check for context effects in the results of the feature analysis.

Each set consisted of 48 syllables (16 consonants combined with each of three vowels). The CV-1 set differed from that used by Miller and Nicely[1] only in the

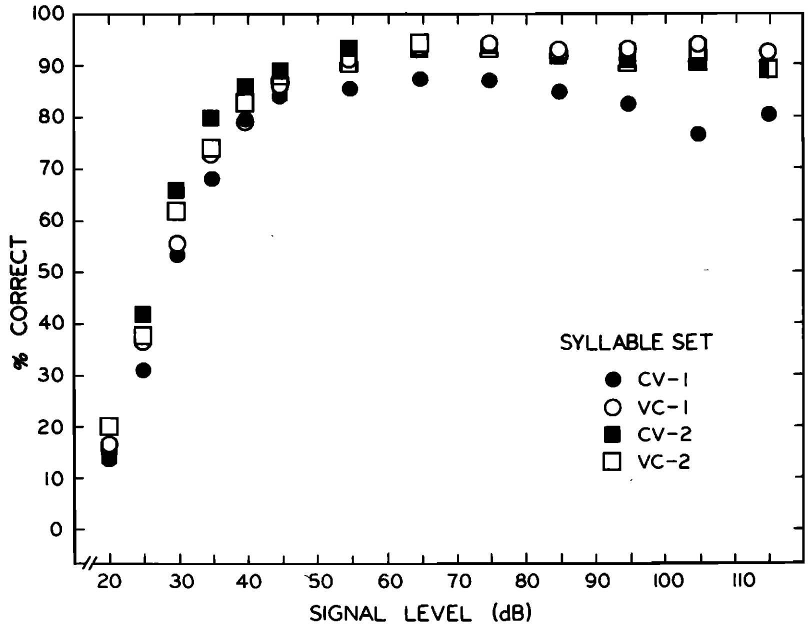

FIG. 1. Mean performance as a function of signal level for the four syllable sets presented without masking noise. Each point represents the average of 12 blocks of trials or approximately 1152 observations.

substitution of /tʃ/ and /dʒ/ for /m/ and /n/. Except for consonant position, the CV-1 and VC-1 sets were identical, and the effects of this variable on confusion patterns could be studied directly. The CV-2 set contained all initial consonants not included in the CV-1 set, plus additional consonants chosen so as to maximize the differences in feature composition among the stimuli. The VC-2 set contained the remaining final consonants /m/, /n/, and /ŋ/ in place of /d/, /t/, and /k/.

C. Noise Experiment

Confusion matrices were obtained at six speech-to-noise (S/N) ratios ranging from −10 dB to +15 dB in equal steps. These levels were based on pilot data indicating that across syllable sets, near chance and near perfect performance would be obtained using this range. At each S/N ratio, data were collected at noise levels of 50, 65, 80, and 95 dB SPL. Variation in absolute noise level was incorporated so that the results from normal listeners could later be compared with masking data obtained at comparable noise levels from hearing-impaired listeners.

Since the output of the Speechmaker was found to be highly stable over time, speech levels were calibrated only twice during the experiment. Partial control over the level of individual syllables was available, and the syllables were matched as nearly as possible in terms of the peak reading on a VU meter, which typically occurred during the vowel portion of the syllable. The level of the entire system was then defined as the SPL of a 1-kHz tone matched to the average peak VU reading obtained on the recorded syllables. The speech signal was gated by an electronic switch (GS-1287) with a rise time of 1 msec, amplified, attenuated, and fed to a passive four-way splitter.

Four independent noise generators (GS-455C) were used to produce a broad-band white noise masker. The output of each noise source was attenuated and then mixed with the output of one channel of the four-way splitter. The combined signal was attenuated further, if necessary, and delivered to one of four sound-treated booths (IAC 401A). Speech and noise were presented monoaurally to the right ear over a TDH-49 earphone mounted in an MX-41/AR cushion; an identical dummy phone was used for the left ear.

Each booth contained a response console with a 4×4 array of response buttons, each labeled with a different consonant sound in conventional orthography. To the left of this array, a list of monosyllabic cue words was available for reference throughout the experiment. In addition to the response buttons, there were three coincidence indicators labeled "Warning," "Observe," and "Answer."

Sixteen paid volunteer subjects were assigned to four listening groups, one for each syllable set. The listeners, six males and 10 females, ranged in age from 17 to 24 years, and with one exception, had no previous listening experience.

Blocks of trials under a single experimental condition were 96 items long. Within a block each syllable occurred twice, and the order of syllables was completely random.

On each trial there was a 500-msec warning interval and a 511-msec observation interval during which the syllable was presented against a background of continuous noise. The subject responded by pressing one of the 16 response buttons. The answer interval was terminated when the last of the four listeners responded. A 200-msec feedback interval followed, during which a green light in the upper portion of the correct response button was lit.

The experiment consisted of three replications of the 24 experimental conditions. Within each replication the order of conditions was randomized by selecting a S/N ratio and holding it constant while noise level was varied across four blocks of trials.

On the first day, subjects listened to nine blocks of trials presented without masking noise. On this practice day, modified gain functions were obtained by varying the level of the speech signal from 25 to 110 dB SPL. Daily listening sessions lasted approximately $2\frac{1}{2}$ hours, including a 20-min midsession break.

D. Control Experiment

In order to determine the overall intelligibility of the speech materials used here and to obtain control data on consonant confusions without masking noise, a preliminary experiment was conducted in which the effects of signal level on performance were studied. Six subjects, divided into two listening groups, participated. For each syllable set, confusion matrices were obtained from all six subjects at 13 signal levels ranging from 20 to 45 dB SPL in 5-dB steps, and from 55 to 115 dB SPL in 10-dB steps. Two 96-item blocks of trials were presented at each level. The order of conditions was randomized, and the replications and listening sessions were organized as in the masking experiment.

II. RESULTS

A. Intelligibility

Gain functions obtained from the control experiment are plotted in Fig. 1. After all percent-correct scores had been converted to arcsins,[22] analysis of variance revealed the expected main effect due to signal level, and also a significant difference among syllable sets, primarily attributable to the difficulty of the CV-1 set. Three interactions were also significant: syllable set × vowel; syllable set × level; and vowel × level. The interactions with level simply reflect the fact that differences among vowels and among syllable sets were greater at low signal levels where the gain function is steep than at moderate and high levels where performance was nearly perfect. The interaction between vowel and syllable set is more interesting than the other two interactions and is plotted in Fig. 2. In all syllable sets, consonants were better identified when the accompanying vowel was /u/ rather than /ɑ/. The effects of the vowel /i/ on consonant intelligibility, however, depended on consonant position. In CV syllable sets, consonants followed by /i/ were the most difficult to identify, whereas in the VC syllable sets, consonants preceded by /i/ were the most easily identified. Singh and Black[2] also found that consonants followed by /i/ were less well discriminated by English-speaking listeners than consonants followed by /ɑ/.

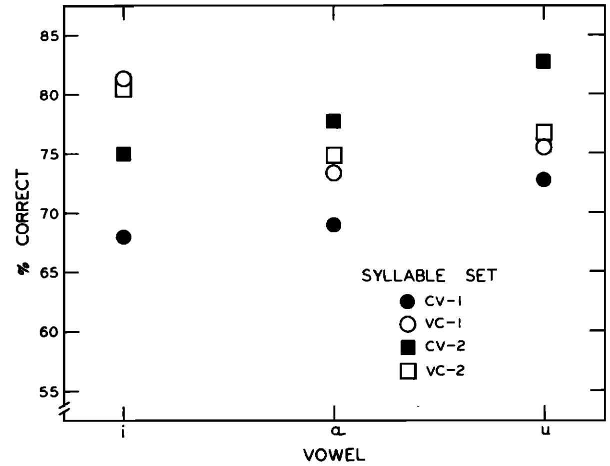

FIG. 2. Interaction between vowels and syllable sets for control experiment, plotted in mean percent correct.

Figure 3 presents the masking functions for the four syllable sets. An analysis of variance of arcsin-transformed scores revealed significant main effects for all four factors: syllable set, S/N, noise level, and vowel. As can be seen in Fig. 3, the CV-2 syllable set, in which feature differences among the consonants had been maximized, was the most intelligible in noise. Although the CV-1 syllable set had been least intelligible in the control experiment, it was not different from the VC-1 and VC-2 sets in noise. However, since the one subject with previous listening experience was in the CV-1 group, it is likely that his superior performance inflated the mean for that group.

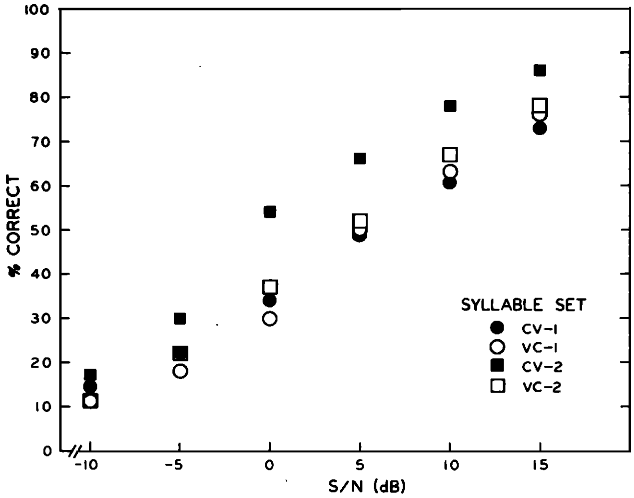

FIG. 3. Mean performance as a function of S/N for the four syllable sets presented in noise. Each point represents the average of 48 blocks of trials or approximately 4608 observations.

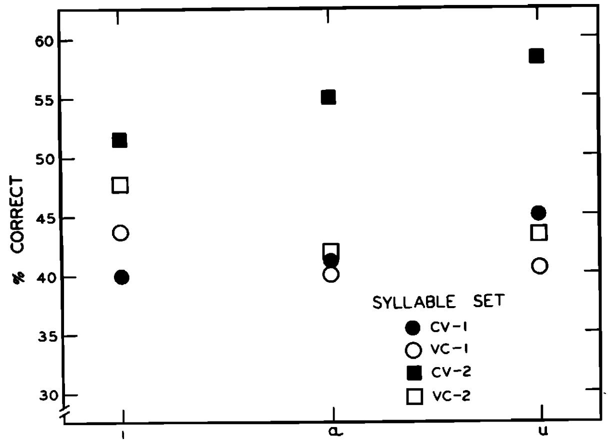

FIG. 4. Interaction between vowels and syllable sets for masking experiment plotted in mean percent correct.

Absolute level of the masking noise produced systematic differences in performance, with the two higher noise levels, 80 and 95 dB, producing slightly lower discrimination scores. These differences were most marked at moderate S/N ratios, producing a significant interaction between noise level and S/N. These effects of noise level are similar to those obtained by other researchers using monosyllabic stimuli.[23,24] Noise level also interacted with vowels, in that there was no difference among vowels at the highest noise level.

Across syllable sets, consonants paired with /ɑ/ were most difficult and consonants paired with /u/ were least difficult to discriminate. However, the interaction between vowel and syllable set was also significant, and was very similar to the comparable interaction obtained in the control experiment. As can be seen in Fig. 4, VC syllables containing /i/ were easier than other VC syllables, while CV syllables with /i/ were more difficult. Differences between /ɑ/ and /u/ were consistent across syllable sets.

Additional significant interactions were obtained for vowel×S/N, syllable set×S/N, and vowel×syllable set×S/N. None of these interactions was systematic or readily interpretable, but their significance suggests caution in the interpretation of lower-order effects involving these factors.

B. Perceptual Confusions

Confusion matrices for the four syllable sets are presented in Tables II–V for the masking experiment, and in Tables VI–IX for the control experiment. Although equal numbers of presentations were programmed for all stimuli, occasional equipment malfunctions caused some trials to be lost. This accounts for the slight inequality of row totals in Tables II–IX.

In Table X, 19 features are defined for the 25 consonants studied. The first 12 features are those proposed by Chomsky and Halle[25] which are distinctive for English consonants. These binary phonological features are abstract classificatory devices, which represent independently controllable aspects of articulation. The remaining features, which resemble more traditional articulatory descriptors, were taken from the feature analyses of perceptual data reported by other researchers. Frication, duration, and place (MN) were taken from Miller and Nicely.[1] Singh and Black[2] extended place to four categories [place (SB)], and Wickelgren[11] differentiated five places of articulation [place (W)]. We have not attempted to extend place (MN) and place (SB) to consonants in the CV-2 syllable set, but we have compared the three place features in other syllable sets. Sibilant, a feature which sets /tʃ, dʒ, s, z, ʃ, ʒ/ apart from other consonants, has been shown by Singh and his colleagues to be a strong perceptual feature both for English and for Hindi.[15–18] The last feature, open, was suggested by Wickelgren[11] and describes the general degree of openness of the vocal tract during articulation.

TABLE II. Consonant confusions for CV-1 syllables addes across all speech-to-noise ratios.

	Response															
Stimulus	p	t	k	b	d	g	f	θ	s	ʃ	v	ð	z	ʒ	tʃ	dʒ
p	933	210	191	16	16	47	98	39	30	17	11	10	26	16	14	24
t	245	843	213	35	67	31	68	30	46	25	22	11	12	12	29	24
k	324	247	565	30	35	53	89	27	39	40	24	8	15	7	97	102
b	136	60	68	486	60	25	191	126	65	5	316	104	33	6	5	9
d	33	56	33	178	819	133	53	34	46	11	94	78	62	17	14	42
g	20	22	20	82	148	938	15	12	20	12	132	29	57	37	7	162
f	198	88	69	91	40	9	765	168	113	19	48	42	28	6	4	11
θ	107	67	69	107	25	13	667	275	204	29	58	32	29	4	12	14
s	44	38	34	51	38	25	150	70	905	57	70	29	157	14	13	15
ʃ	26	37	51	17	31	38	26	13	84	870	12	8	34	35	288	130
v	34	20	37	204	43	52	84	65	43	13	705	193	155	23	4	27
ð	37	30	18	239	82	64	49	67	56	10	534	270	192	12	9	28
z	20	22	23	72	58	115	31	32	41	12	231	114	787	68	3	84
ʒ	19	20	24	32	66	286	19	22	16	16	77	38	137	420	6	502
tʃ	67	149	92	15	20	14	46	24	144	221	13	8	16	15	829	37
dʒ	20	21	12	30	73	152	8	6	18	39	34	8	106	148	11	1016

TABLE III. Consonant confusions for VC-1 syllables, added across all speech-to-noise ratios.

	Response															
Stimulus	p	t	k	b	d	g	f	θ	s	ʃ	v	ð	z	ʒ	tʃ	dʒ
p	911	163	316	38	10	23	61	29	44	26	15	17	7	14	37	24
t	182	922	169	24	40	27	46	62	70	46	9	15	12	13	65	19
k	252	200	813	33	19	52	75	44	64	51	10	15	12	13	51	17
b	70	34	70	845	77	254	50	38	29	24	92	43	35	18	10	27
d	27	35	33	178	821	121	21	41	22	29	63	115	78	40	14	83
g	48	49	126	106	105	896	30	30	39	42	49	43	33	27	20	72
f	403	110	242	70	27	96	437	83	68	41	45	26	12	9	28	25
θ	64	200	98	92	93	88	171	437	135	64	54	103	37	18	21	32
s	85	183	91	23	12	20	44	54	1040	49	10	22	21	10	44	16
ʃ	47	168	116	35	23	57	42	58	114	764	27	28	27	34	119	62
v	17	27	24	182	95	121	43	57	28	15	723	190	111	39	8	44
ð	11	14	9	137	289	95	21	58	16	20	327	410	213	49	10	39
z	17	47	44	42	83	78	41	70	52	50	242	212	550	111	19	64
ʒ	14	23	21	43	63	78	24	33	25	29	154	115	186	743	20	146
tʃ	53	286	95	34	28	34	54	40	56	103	28	26	17	23	593	249
dʒ	54	143	96	53	307	86	41	43	53	90	23	27	32	29	146	482

It should be noted that those features in Table X which are not binary have been interpreted in some studies as ordered categories[11,16]; that is, the degree of feature difference between the first and the third place categories would be considered equivalent to two binary feature differences. However, since the information analysis is insensitive to the order in which categories are specified, the digits used in Table X should only be considered arbitrary labels for categories that are not necessarily ordered.

The 19 features in Table X are highly redundant, both with respect to the complete set of 25 consonants, and with respect to the 16-consonant syllable sets used in the experiments. Stimulus information, in bits, for each feature in each syllable set is given in Table XI. The CV-1 and VC-1 sets are identical. In them stimulus information for the features vocalic, consonantal, nasal, and round is zero because all consonants in these sets belong to the same category on each of these features. Moreover, there are two pairs of features which are perfectly correlated. All consonants which are high are also non-anterior; all consonants which are continuant are also in the same category of the feature open. If all of the specified features were perfectly independent, there would be a total of 15.81 bits of stimulus information. Since for 16 stimuli the maximum possible stimulus information is four bits, this feature system contains 11.81 bits of internal redundancy.

All features defined for the CV-2 syllable set are distinctive, and none of the features are perfectly correlated. The total feature information, and hence redundancy, is slightly higher in this set than in the preceding sets. The pattern of values for the VC-2 set is quite similar to that for the CV-1 and VC-1 sets, except that an additional feature, nasal, is distinctive. Here again, high and non-anterior, and continuant and open are perfectly correlated and thus indistinguishable.

Tables XII–XV contain performance measures for each feature in each syllable set. In this and subsequent analyses, feature performance is assessed relative to stimulus feature information, i.e., it is the percentage of information transmitted which constitutes the feature

TABLE IV. Consonant confusions for CV-2 syllables, added across all speech-to-noise ratios.

	Response															
Stimulus	p	b	tʃ	dʒ	l	r	f	s	v	z	m	n	h	h^w	w	j
p	963	12	51	29	20	12	76	45	16	6	19	17	371	37	28	25
b	140	549	14	16	51	16	237	68	220	30	119	44	144	38	28	10
tʃ	75	9	1159	31	13	12	93	154	13	11	8	12	64	26	23	21
dʒ	7	23	47	1030	35	36	12	44	31	135	23	12	25	40	40	175
l	11	27	3	16	1274	53	22	14	30	34	81	35	19	20	57	31
r	8	14	10	28	256	1089	5	13	22	44	38	23	4	29	66	79
f	150	135	35	16	24	16	804	130	52	41	48	12	223	23	13	2
s	44	62	54	23	32	15	168	1036	54	103	25	14	52	27	6	9
v	23	128	7	15	132	53	132	30	562	254	181	56	18	30	82	21
z	20	55	14	56	58	75	36	69	192	871	43	31	24	38	82	67
m	45	68	6	11	45	18	44	30	64	23	1189	90	30	20	21	15
n	25	15	6	14	39	22	25	27	24	27	416	978	38	20	31	20
h	470	24	70	18	17	16	93	49	18	10	21	10	822	56	21	11
h^w	28	21	31	43	43	51	41	37	23	22	19	14	88	769	447	43
w	25	14	16	56	41	73	20	20	21	68	26	13	25	351	842	113
j	18	10	15	52	147	45	16	25	12	28	35	24	14	17	42	1220

Table V. Consonant confusions for VC-2 syllables, added across all speech-to-noise ratios.

	Response															
Stimulus	p	b	g	ŋ	m	n	f	θ	s	ʃ	v	ð	z	ʒ	tʃ	dʒ
p	1044	66	53	10	24	19	109	79	100	42	20	24	17	15	61	37
b	90	733	194	22	130	66	78	81	56	27	69	45	34	29	19	37
g	48	173	824	34	30	61	51	62	51	37	57	40	32	51	34	136
ŋ	18	24	45	991	148	284	26	26	24	17	33	22	25	12	15	16
m	24	24	26	185	974	192	29	41	19	14	55	34	25	20	12	41
n	27	21	21	243	154	915	13	28	18	20	29	28	65	72	15	49
f	362	93	87	21	29	34	577	195	112	32	38	30	24	18	36	35
θ	66	101	67	14	18	40	195	678	201	59	68	67	37	32	42	37
s	103	23	26	15	8	19	75	130	1055	71	32	10	26	15	74	22
ʃ	102	41	45	26	29	41	70	104	195	632	38	18	25	38	203	113
v	26	151	114	46	41	88	40	86	45	30	626	223	98	52	10	47
ð	27	95	109	32	31	107	23	72	28	15	340	543	160	68	12	57
z	28	30	55	40	45	139	45	111	85	45	172	226	479	143	19	54
ʒ	28	36	65	30	22	82	29	50	36	32	112	168	223	641	15	146
tʃ	133	79	31	23	26	43	65	64	93	76	24	13	24	28	777	226
dʒ	86	94	97	28	29	27	77	82	107	61	34	26	25	60	158	721

performance measure. Percent information transmitted is shown in Tables XII–XV for each S/N ratio and for all S/N ratios combined. Also included are the results for the control experiment.

Two general points can be made about these results. First, the relative importance of the features changes as a function of the listening conditions, i.e., noise versus quiet. Voicing and nasality are well perceived in the presence of masking noise, but their intelligibility drops relative to that of other features in quiet. Secondly, the relative importance of the features is not invariant across syllable sets. A comparison of the CV-1 and VC-1 sets, which contained the same consonant phonemes, demonstrates this. In the CV-1 set, the combined feature high–anterior is the most intelligible of all features in quiet and is second only to voicing in the presence of masking noise. In the VC-1 set, however, the importance of high–anterior is somewhat diminished in quiet and is considerably lessened in noise.

Many comparisons of this sort are possible but the lack of independence of the features makes them risky. For example, consider the two features sibilant and duration. The latter specifies /s, z, ʃ, ʒ/ as long, whereas the former specifies these four sounds plus /tʃ, dʒ/ as sibilant. Clearly, any situation in which duration is well perceived will result in good performance on the sibilant feature, and vice versa. Inspection of Tables XII–XV verifies that both features tend to be well perceived. To what extent can performance on duration be accounted for by the intelligibility of sibilant, and vice versa? Are both features necessary in a perceptual feature system, or are they functionally equivalent? These are the issues we attempt to deal with in the next section.

C. Sequential Information Analysis

The technique for partialing out the effects of one variable on another in information analysis is well known. Consider two categorical variables X and Y. The information (uncertainty) of the joint variable (X,Y) is given by

$$U(X,Y)=U(X)+U_X(Y)=U(Y)+U_Y(X), \qquad (1)$$

Table VI. Consonant confusions for CV-1 syllables, added across all signal levels.

	Response															
Stimulus	p	t	k	b	d	g	f	θ	s	ʃ	v	ð	z	ʒ	tʃ	dʒ
p	773	38	35	9	14	6	16	8	9	2	2	2	4	7	1	4
t	33	783	27	7	19	6	0	3	8	3	3	4	11	8	6	9
k	71	89	585	11	14	18	14	2	4	15	9	2	6	9	37	47
b	21	11	8	587	28	17	33	28	8	1	133	47	5	3	0	3
d	9	10	4	32	771	24	1	9	12	3	18	8	12	5	3	9
g	4	4	8	27	41	764	3	3	5	1	21	3	8	14	2	19
f	43	7	13	44	4	4	681	63	17	3	20	11	13	4	2	4
θ	22	15	13	64	11	2	321	355	34	2	27	38	14	0	1	2
s	10	4	4	23	7	3	31	26	732	7	13	5	60	4	0	1
ʃ	5	4	7	3	5	6	0	1	7	745	2	0	9	19	93	21
v	7	10	9	128	11	2	23	31	6	3	567	38	50	6	4	6
ð	11	1	5	88	24	14	21	57	18	7	407	204	58	9	2	7
z	7	8	9	15	9	2	10	9	20	5	28	70	756	15	2	7
ʒ	6	9	8	10	16	48	4	4	5	21	7	34	17	612	8	149
tʃ	3	17	4	7	5	3	9	6	13	37	3	3	4	11	794	14
dʒ	3	6	4	2	8	40	3	4	8	12	5	0	25	76	3	734

TABLE VII. Consonant confusions for VC-1 syllables, added across all signal levels.

	Response															
Stimulus	p	t	k	b	d	g	f	θ	s	ʃ	v	ð	z	ʒ	tʃ	dʒ
p	763	23	80	9	3	10	14	6	6	4	4	2	3	5	1	4
t	25	795	31	1	9	6	4	4	12	3	0	2	2	2	22	9
k	66	36	753	2	7	18	7	5	4	10	3	1	3	5	7	6
b	55	12	28	608	19	87	46	9	9	2	43	6	3	3	2	4
d	9	16	7	10	763	13	8	13	12	2	15	33	14	9	3	9
g	16	10	44	22	7	762	13	5	3	4	20	4	3	4	2	12
f	56	0	11	22	9	13	712	61	10	3	25	8	0	1	0	2
θ	11	14	9	9	42	10	73	656	28	7	23	36	13	1	1	3
s	7	15	12	1	3	4	15	23	828	14	1	1	11	0	1	0
ʃ	9	17	10	7	9	6	6	1	18	788	4	2	4	18	19	18
v	5	5	8	38	22	19	48	7	11	4	637	104	12	7	1	5
ð	8	7	5	38	77	4	10	37	6	7	136	556	34	8	1	2
z	4	8	15	4	12	10	8	22	42	8	54	35	687	17	4	3
ʒ	9	3	9	1	13	7	4	7	15	30	24	13	22	752	4	20
tʃ	10	17	14	5	2	7	2	1	5	14	2	1	0	5	727	121
dʒ	16	20	31	2	25	17	4	1	11	10	3	2	8	7	86	689

that is, the joint uncertainty reflects the uncertainty of one variable plus the residual, conditional uncertainty in the second when the first is held constant.

The term $U_X(Y)$ is obtained by computing $U(Y)$ separately for each category of X, using conditional probabilities, and taking a weighted average across all of the X categories:

$$U_X(Y) = -\sum_{i=1}^{r} p(X_i) \sum_{j=1}^{s} p_{X_i}(Y_j) \log_2 p_{X_i}(Y_j). \quad (2)$$

The conditional probabilities, $p_{X_i}(Y_j)$, sum to 1.00 within each category X_i, and the weights, $p(X_i)$, sum to 1.00 across all categories of X.

It is possible to partial out the effects of two variables on a third in a similar manner:

$$U_{XY}(Z) = -\sum_{i=1}^{r} \sum_{j=1}^{s} p(X_i,Y_j) \sum_{k=1}^{t} p_{X_i,Y_j}(Z_k) \times \log_2 p_{X_i,Y_j}(Z_k). \quad (3)$$

Again, the conditional probabilities, $p_{X_i,Y_j}(Z_k)$, sum to 1.00 within each joint category (X_i,Y_j), and the joint probabilities, $p(X_i,Y_j)$, sum to 1.00 across the rs joint categories of X and Y.

Although the notation becomes increasingly cumbersome, calculation of other conditional uncertainties follows the same pattern. For example, a conditional (or partial) contingent uncertainty, $U_Z(X:Y)$, may be obtained by computing $U(X:Y)$ separately for each category of Z, using conditional probabilities, and taking a weighted average across the categories of Z.

In applying these techniques to uncertainty analysis in terms of features, we begin by specifying a set of features, say A, B, and C, which describe a set of stimuli, S, such that $U(A,B,C) = U(S)$. This condition implies only that each stimulus has a unique description in terms of the features. In general, more than three features would be necessary to accomplish this, but in order to simplify the illustration we will use only three. Similarly, we classify a set of response categories, R, in terms of the same features, using a lower case notation

TABLE VIII. Consonant confusions for CV-2 syllables, added across all signal levels.

	Response															
Stimulus	p	b	tʃ	dʒ	l	r	f	s	v	z	m	n	h	h^w	w	j
p	769	7	3	4	4	3	21	10	7	6	7	7	67	2	4	2
b	17	611	2	3	12	4	37	12	143	14	29	7	17	6	4	4
tʃ	1	4	820	26	4	6	8	26	1	9	5	5	6	1	3	2
dʒ	0	2	22	802	10	4	1	15	7	26	4	6	0	3	2	14
l	3	14	5	4	776	10	4	11	15	15	16	4	8	16	17	3
r	6	7	2	5	27	809	6	6	12	9	7	2	7	6	8	3
f	28	43	5	5	8	2	726	33	21	12	9	5	26	3	3	1
s	5	23	4	2	9	6	43	762	10	49	6	2	3	2	1	0
v	4	129	2	3	21	14	23	8	593	52	26	12	3	11	7	15
z	7	14	5	10	5	9	12	43	29	766	2	5	5	4	5	13
m	9	50	2	2	9	6	14	4	26	10	763	20	5	1	4	2
n	2	6	6	0	7	3	6	23	6	12	80	753	7	4	6	5
h	88	13	5	5	7	3	33	20	13	7	4	5	698	9	7	7
h^w	3	10	3	8	14	8	2	9	7	6	6	4	15	630	186	7
w	5	16	2	3	9	19	3	6	5	6	8	2	13	322	502	8
j	4	11	6	19	17	8	4	13	4	16	4	8	11	5	2	801

TABLE IX. Consonant confusions for VC-2 syllables, added across all signal levels.

	Response															
Stimulus	p	b	g	ŋ	m	n	f	θ	s	ʃ	v	ð	z	ʒ	tʃ	dʒ
p	813	11	24	5	6	3	23	3	14	2	5	2	4	3	8	1
b	75	592	92	4	27	5	57	6	11	1	45	1	2	3	4	5
g	13	35	759	5	2	11	11	6	7	7	15	5	8	6	12	25
ŋ	9	4	10	789	44	41	6	1	4	6	8	1	0	2	4	4
m	9	12	6	24	784	46	12	2	6	4	16	1	7	1	1	2
n	10	14	8	19	40	757	5	6	5	4	16	25	16	1	2	2
f	62	18	22	1	4	6	695	64	9	3	20	11	2	2	4	4
θ	14	13	14	2	2	9	74	653	27	5	22	63	9	5	2	13
s	13	5	1	0	2	2	22	25	822	5	1	1	14	2	7	2
ʃ	16	1	7	4	3	6	8	4	33	782	3	1	0	22	27	16
v	6	44	19	5	5	14	38	12	7	3	636	112	18	3	2	3
ð	12	53	6	2	3	57	10	27	10	2	147	565	33	7	3	2
z	16	14	10	6	4	10	13	22	68	4	53	29	661	14	4	5
ʒ	7	3	6	3	5	13	2	6	10	23	22	13	26	755	9	21
tʃ	20	2	4	4	4	3	4	2	11	15	3	2	3	6	742	111
dʒ	22	4	27	5	7	5	8	2	11	16	6	6	4	11	90	706

to differentiate the response categories from the stimulus categories, i.e., $U(R)=U(a,b,c)$. Given a stimulus–response confusion matrix, we can compute the contingent uncertainty $U(S:R)$ which is normally designated "transmitted information." In terms of the features,

$$U(S:R)=U(A,B,C:a,b,c), \tag{4}$$

that is, the transmitted information represents a contingency between the joint feature categories of the stimulus and the joint feature categories of the response.

There are many ways to partition $U(A,B,C:a,b,c)$, some of which are more interpretable than others. We have concentrated on the following expansion:

$$\begin{aligned} U(A,B,C:a,b,c) &= U(A:a)+U_A(B:a)+U_{AB}(C:a) \\ &\quad +U_a(A:b)+U_{aA}(B:b)+U_{aAB}(C:b) \\ &\quad +U_{ab}(A:c)+U_{abA}(B:c)+U_{abAB}(C:c). \end{aligned} \tag{5}$$

The terms lying along the main diagonal of this array are readily interpretable. The first, $U(A:a)$, is the transmitted information for feature A. The second term, $U_{aA}(B:b)$, is the conditional transmitted information for feature B, in which the effects of feature A have been partialed out of both stimuli and responses. The third term is the conditional transmitted information

TABLE X. Features specified for the sequential information analysis.

Phoneme	Voc	Cons	High	Low	Back	Cor	Ant	Voi	Nas	Cont	Str	Rnd	Fric	Dur	Pl[1]	Pl[2]	Pl[11]	Sib[2]	Op[11]
p	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
t	0	1	0	0	0	1	1	0	0	0	0	0	0	0	1	1	1	0	0
k	0	1	1	0	1	0	0	0	0	0	0	0	0	0	2	3	4	0	0
b	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
d	0	1	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0
g	0	1	1	0	1	0	0	1	0	0	0	0	0	0	2	3	4	0	0
f	0	1	0	0	0	0	1	0	0	1	1	0	1	0	0	0	0	0	1
θ	0	1	0	0	0	1	1	0	0	1	0	0	1	0	1	1	1	0	1
s	0	1	0	0	0	1	1	0	0	1	1	0	1	1	1	1	2	1	1
ʃ	0	1	1	0	0	1	0	0	0	1	1	0	1	1	2	2	3	1	1
v	0	1	0	0	0	0	1	1	0	1	1	0	1	0	0	0	0	0	1
ð	0	1	0	0	0	1	1	1	0	1	0	0	1	0	1	1	1	0	1
z	0	1	0	0	0	1	1	1	0	1	1	0	1	1	1	1	2	1	1
ʒ	0	1	1	0	0	1	0	1	0	1	1	0	1	1	2	2	3	1	1
tʃ	0	1	1	0	0	1	0	0	0	0	1	0	1	0	2	2	3	1	0
dʒ	0	1	1	0	0	1	0	1	0	0	1	0	1	0	2	2	3	1	0
m	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0
n	0	1	0	0	0	1	1	1	1	0	0	0	0	0	1	1	1	0	0
ŋ	0	1	1	0	1	0	0	1	1	0	0	0	0	0	2	3	4	0	0
r	1	1	0	0	0	1	0	1	0	1	0	0	0	0	—	2	1	0	2
l	1	1	0	0	0	1	1	1	0	1	0	0	0	0	—	1	2	0	2
w	0	0	1	0	1	0	0	1	0	1	0	1	0	0	—	—	0	0	2
h	0	0	0	1	0	0	0	0	0	1	0	0	1	0	—	—	4	0	2
h^w	0	0	0	1	0	0	0	0	0	1	0	1	1	0	—	—	4	0	2
j	0	0	1	0	0	0	0	1	0	1	0	0	0	0	—	—	3	0	2

for feature C, in which the effects of both A and B have been partialed out. Thus, $U(A:a)$, $U_{aA}(B:b)$, and $U_{abAB}(C:c)$ represent independent components of the total transmitted information corresponding to direct contingencies between stimulus and response features.

The remaining terms in Eq. 5 represent what Garner[26] has termed "cross-contingencies," i.e., contingencies between a stimulus feature and an inappropriate response feature. In the present context, however, they are partial cross-contingencies because certain stimulus and response features have been held constant. These terms, if significantly large, would indicate complicated perceptual interactions among stimulus and response features.[27]

In a real feature system, the designations A, B, C, are clearly arbitrary; that is, there are six different ways these labels could be assigned to three features and each would result in a somewhat different set of values for the nine terms in Eq. 5. The object of the sequential information analysis is to sequentially designate features as A, B, C, etc., using a performance criterion. Throughout the analysis relative frequencies are used in place of probabilities; the analysis proceeds in the following steps:

Step 1. Unconditional transmitted information is estimated for each feature in a proposed feature system. Since, as can be seen in Table XI, stimulus feature information may vary considerably from feature to feature, performance is assessed in terms of the percentage of information transmitted. This is given by the quantity $U(A:a)/U(A)$, which has the effect of normalizing the features for inequalities in stimulus feature information. Tables XII–XV, which have already been discussed, contain these relative performance measures for our data. The feature with the highest percentage of information transmitted is designated as "A."

Step 2. With feature A held constant, we calculate the conditional transmitted information for the remaining features, yielding a set of values of the form $U_{aA}(B:b)$. These terms represent transmitted information which is independent of the information transmitted on feature A. In order to assess performance relative to stimulus uncertainty, we also calculate a set of terms of the form $U_{aA}(B)$. Such terms represent the maximum possible value of conditional transmitted information for feature B when feature A is held constant, because they represent the stimulus information for feature B that is independent of stimulus feature A and responses to feature A. The ratio of these two terms, $U_{aA}(B:b)/U_{aA}(B)$, indicates what proportion of feature B stimulus information was transmitted, when feature A is held constant. The feature with the highest ratio is designated as feature B, provided $U_{aA}(B:b)$ is not less than 1% of the total transmitted information. This additional criterion is necessary because in later iterations, when many features are held constant, the conditional stimulus information for some features approaches zero, and the performance ratio becomes unstable.

TABLE XI. Feature information in bits per stimulus for each syllable set.

Feature	Syllable set			
	CV-1	VC-1	CV-2	VC-2
Vocalic	0.00	0.00	0.54	0.00
Consonantal	0.00	0.00	0.81	0.00
High	0.95	0.95	0.81	0.95
Low	0.00	0.00	0.54	0.00
Back	0.54	0.54	0.34	0.54
Coronal	0.95	0.95	0.99	0.99
Anterior	0.95	0.95	0.99	0.95
Vocie	1.00	1.00	0.95	0.95
Nasal	0.00	0.00	0.54	0.70
Continuant	1.00	1.00	0.95	1.00
Strident	1.00	1.00	0.95	1.00
Round	0.00	0.00	0.54	0.00
Frication	0.95	0.95	0.99	0.95
Duration	0.81	0.81	0.54	0.81
Place (MN)	1.56	1.56	—	1.58
Place (SB)	1.90	1.90	—	1.92
Place (W)	2.25	2.25	2.19	2.23
Sibilant	0.95	0.95	0.81	0.95
Open	1.00	1.00	1.56	1.00
Total	15.81	15.81	16.04	16.52
Redundancy	11.81	11.81	12.04	12.52

TABLE XII. Percent information transmitted for each feature in the CV-1 syllable set under various listening conditions. The figures in the last two columns are based on Tables II and VI.

Feature	Speech-to-noise ratio							
	−10dB	−5dB	0dB	+5dB	+10dB	+15dB	All	Quiet
Voice	8.7	24.5	38.0	51.3	63.5	78.1	38.4	57.1
Frication	2.2	7.0	18.3	30.1	39.3	52.9	19.9	50.5
Sibilant	3.0	10.0	23.6	38.2	54.7	77.9	27.1	60.1
Strident	1.7	5.0	11.4	19.8	26.2	37.6	13.7	37.1
Duration	1.5	3.9	10.8	19.5	34.3	59.6	15.6	49.6
Place (MN)	1.7	6.7	17.2	32.2	46.4	61.0	20.7	50.6
Place (SB)	2.0	7.6	18.6	35.9	49.8	62.9	22.5	52.6
Place (W)	2.2	7.8	19.2	34.5	48.8	66.3	22.8	53.6
High, anterior	2.6	9.7	23.9	45.4	61.4	79.5	28.6	63.4
Back	0.4	2.4	7.0	20.2	30.8	55.5	13.5	48.0
Coronal	0.4	2.2	7.8	16.8	28.9	41.3	11.5	36.3
Continuant, open	1.6	5.4	14.6	24.8	33.4	46.6	16.7	46.1

Table XIII. Percent information transmitted for each feature in the VC-1 syllable set under various listening conditions. The figures in the last two columns are based on Tables III and VII.

Feature	Speech-to-noise ratio							Quiet
	−10dB	−5dB	0dB	+5dB	+10dB	+15dB	All	
Voice	3.4	21.4	32.8	51.2	64.1	74.0	34.2	54.5
Frication	0.7	3.1	11.4	27.2	41.6	57.6	16.7	57.3
Sibilant	0.2	1.4	9.3	30.3	52.2	70.8	17.9	63.6
Strident	0.2	1.4	6.5	20.6	35.3	50.3	12.6	51.0
Duration	0.0	1.0	8.1	31.4	55.7	74.2	18.1	62.8
Place (MN)	0.4	2.1	8.7	25.2	39.5	55.7	15.2	56.5
Place (SB)	0.5	2.6	10.4	31.2	47.2	64.8	18.3	60.1
Place (W)	0.5	2.3	10.2	30.6	48.4	65.5	18.3	60.6
High, anterior	0.2	1.9	5.9	21.9	40.0	59.7	14.1	59.9
Back	0.4	1.8	8.3	26.8	38.2	58.1	15.7	55.5
Coronal	0.6	2.4	12.6	35.9	47.7	60.8	18.6	55.5
Continuant, open	0.9	3.8	11.9	28.7	42.8	60.5	17.6	60.0

Table XIV. Percent information transmitted for each feature in the CV-2 syllable set under various listening conditions. The figures in the last two columns are based on Tables IV and VIII.

Feature	Speech-to-noise ratio							Quiet
	−10dB	−5dB	0dB	+5dB	+10dB	+15dB	All	
Vocalic	3.3	19.3	52.9	75.2	87.9	91.0	44.3	67.3
Consonantal	1.7	9.1	29.2	44.5	62.4	73.9	28.5	62.6
High	3.7	11.9	32.2	50.8	66.4	70.6	32.3	54.2
Low	2.1	6.2	16.4	22.7	32.2	42.2	17.2	41.6
Back	0.2	4.9	23.4	33.8	40.2	46.9	20.8	30.8
Coronal	1.1	6.6	29.8	44.5	66.3	80.2	27.9	61.7
Anterior	3.2	11.3	28.8	47.2	65.7	76.0	30.7	63.4
Place (W)	2.9	9.8	32.9	48.9	67.2	76.9	31.6	61.9
Voice	8.0	22.9	43.0	50.1	56.7	65.4	36.6	51.9
Nasal	4.9	22.9	56.9	71.4	85.5	92.3	46.4	67.2
Frication	1.5	5.7	18.8	28.1	41.7	54.5	19.6	43.1
Duration	1.8	7.1	22.9	33.3	60.3	80.1	26.6	60.7
Strident	2.1	9.2	32.7	45.5	59.8	71.0	29.0	56.9
Sibilant	2.9	9.8	34.9	46.8	70.6	84.7	32.7	66.6
Continuant	1.2	6.0	23.4	38.3	54.1	66.9	23.9	54.0
Open	2.2	10.0	30.1	45.5	61.1	72.8	29.4	59.4
Round	1.7	11.2	43.6	70.7	86.4	89.0	39.4	71.8

Table XV. Percent information transmitted for each feature in the VC-2 syllable set under various listening conditions. The figures in the last two columns are based on Tables V and IX.

Feature	Speech-to-noise ratio							Quiet
	−10dB	−5dB	0dB	+5dB	+10dB	+15dB	All	
Voice	2.4	14.9	31.9	47.5	62.1	72.2	31.0	56.0
Nasal	2.8	27.3	49.8	68.7	78.0	85.2	42.5	70.7
Frication	0.5	7.3	21.5	42.2	54.7	68.9	23.8	60.3
Sibilant	0.4	5.1	13.5	31.7	50.8	71.1	20.3	65.7
Strident	0.5	3.1	9.9	23.5	35.6	50.7	14.8	51.4
Duration	0.1	2.5	7.8	23.2	48.0	68.0	16.4	62.9
Place (MN)	0.1	2.5	12.7	25.0	43.2	56.1	16.4	58.1
Place (SB)	0.5	4.3	14.8	28.7	47.5	62.5	19.3	61.3
Place (W)	0.6	4.4	14.2	28.1	47.0	62.4	19.2	61.7
High, anterior	0.2	2.8	12.3	23.2	43.7	62.1	16.6	63.8
Back	1.3	6.1	15.8	27.3	44.6	57.4	20.5	59.1
Coronal	0.0	3.3	13.2	29.4	43.6	53.7	16.6	54.3
Continuant, open	0.2	5.8	17.8	34.0	53.2	66.6	20.8	59.9

TABLE XVI. Sequential information analysis of consonant confusions for CV-1 syllable set in quiet.

Iteration No. 1

Feature	Feat. inf.	Trans. inf.	% Trans. inf.
Voice	1.000	0.571	0.571
Fric	0.954	0.481	0.505
Sibil	0.954	0.573	0.601
Strid	1.000	0.371	0.371
Dur	0.811	0.403	0.496
Pl (MN)	1.561	0.790	0.506
Pl (SB)	1.905	1.002	0.526
Pl (W)	2.250	1.206	0.536
Hi/Ant	0.954	0.605	0.634
Back	0.543	0.261	0.480
Coron	0.954	0.346	0.363
Cont/open	1.000	0.461	0.461

Iteration No. 2; Constant: high/anterior

Feature	Cond. Feat. inf.	Cond. Trans. inf.	% Cond. Trans. inf.
Voice	0.999	0.595	0.595
Fric	0.942	0.495	0.526
Sibil	0.788	0.461	0.585
Strid	0.943	0.372	0.395
Dur	0.793	0.412	0.519
Pl (MN)	0.607	0.183	0.302
Pl (SB)	0.943	0.383	0.406
Pl (W)	1.287	0.586	0.455
Back	0.336	0.198	0.589
Coron	0.943	0.383	0.406
Cont/open	0.947	0.446	0.471

Iteration No. 3; Constant: high/anterior, voice

Feature	Cond. Feat. inf.	Cond. Trans. inf.	% Cond. Trans. inf.
Fric	0.930	0.546	0.587
Sibil	0.781	0.488	0.625
Strid	0.936	0.403	0.430
Dur	0.788	0.432	0.549
Pl (MN)	0.604	0.200	0.330
Pl (SB)	0.936	0.413	0.441
Pl (W)	1.279	0.627	0.490
Back	0.331	0.211	0.638
Coron	0.936	0.413	0.441
Cont/open	0.937	0.490	0.522

Iteration No. 4; Constant: high/anterior, voice back

Feature	Cond. Feat. inf.	Cond. Trans. inf.	% Cond. Trans. inf.
Fric	0.598	0.330	0.551
Sibil	0.448	0.271	0.605
Strid	0.603	0.187	0.310
Dur	0.697	0.385	0.552
Pl (MN)	0.604	0.200	0.331
Pl (SB)	0.604	0.200	0.331
Pl (W)	0.946	0.409	0.432
Coron	0.604	0.200	0.331
Cont/open	0.846	0.442	0.523

Iteration No. 5; Constant: high/anterior, voice, back, sibilant

Feature	Cond. Feat. inf.	Cond. Trans. inf.	% Cond. Trans. inf.
Fric	0.487	0.250	0.514
Strid	0.403	0.096	0.239
Dur	0.247	0.112	0.453
Pl (MN)	0.497	0.136	0.275
Pl (SB)	0.497	0.136	0.275
Pl (W)	0.497	0.136	0.275
Coron	0.497	0.136	0.275
Cont/open	0.735	0.363	0.494

Iteration No. 6; Constant: high/anterior, voice, back, sibilant, frication

Feature	Cond. Feat. inf.	Cond. Trans. inf.	% Cond. Trans. inf.
Strid	0.246	0.022	0.089
Dur	0.247	0.112	0.453
Pl (MN)	0.485	0.175	0.361
Pl (SB)	0.485	0.175	0.361
Pl (W)	0.485	0.175	0.361
Coron	0.485	0.175	0.361
Cont/open	0.247	0.112	0.453

Iteration No. 7; Constant: high/anterior, voice, back, sibilant, frication, duration

Feature	Cond. Feat. inf.	Cond. Trans. inf.	% Cond. Trans. inf.
Strid	0.246	0.022	0.089
Place (MN)	0.484	0.175	0.361
Place (SB)	0.484	0.175	0.361
Place (W)	0.484	0.175	0.361
Coron	0.484	0.175	0.361
Cont/open	0.000	0.000	···

Step 3. Calculations analogous to those in Step 2 are carried out for the remaining features, with both feature A and feature B held constant. Feature C is identified as the feature with the highest ratio of the form $U_{abAB}(C:c)/U_{abAB}(C)$, given that $U_{abAB}(C:c)$ is not less than 1% of the total transmitted information.

For feature systems with more than three features, iterations continue in this manner until one of the following criteria is met:

(i) The contributions of all features have been accounted for.

TABLE XVII. Sequential information analyses for CV-1 syllable set under various listening conditions.

	Speech-to-noise ratio							
	−10dB	−5dB	0dB	+5dB	+10dB	+15dB	All	Quiet
Feature $U(S:R)$	0.220	0.595	1.145	1.733	2.292	2.871	1.179	2.323
Voice	(1)0.087	(1)0.245	(1)0.380	(1)0.513	(1)0.635	(2)0.789	(1)0.384	(2)0.595
Frication	(4)0.008	(4)0.020	(4)0.083	(4)0.211	(4)0.273	(4)0.254	(4)0.099	(5)0.250
Sibilant	(2)0.030	(2)0.099	(2)0.233	(5)0.065	(5)0.136	(3)0.604	(3)0.196	(4)0.271
Strident								
Duration					(7)*0.060*	(5)*0.105*		(6)*0.112*
Place (MN)								
Place (SB)				(6)*0.062*	(6)*0.127*			
Place (W)	(5)*0.006*	(5)*0.006*	(5)*0.024*			(6)*0.177*	(6)*0.046*	(7)*0.175*
High, anterior	(3)0.019	(3)0.065	(3)0.161	(2)0.416	(2)0.571	(1)0.759	(2)0.277	(1)0.605
Back				(3)0.120	(3)0.164			(3)0.211
Coronal								
Continuant, open				(7)*0.027*			(5)*0.025*	
Total	0.150	0.435	0.881	1.414	1.966	2.688	1.027	2.219
Proportion of $U(S:R)$	0.682	0.731	0.769	0.816	0.858	0.936	0.871	0.955

(ii) The sum of the partial contingent uncertainties, $U(A:a)$, $U_{aA}(B:b)$, $U_{abAB}(C:c)$, etc., is greater than 99% of the total transmitted information.

(iii) No remaining feature accounts for an additional 1% of the total transmitted information.

The analysis is presently programmed for up to 11 iterations, although this maximum is rarely needed due to the high degree of redundancy in the feature systems studied. When the analysis terminates, the transmitted information accounted for by the features is given by the sum of the (direct) partial contingent uncertainties. The off-diagonal terms in Eq. 5, which represent partial cross-contingent uncertainties, and which are not computed in the analysis, make up the remainder. In the analyses which we have thus far performed, these terms appear to be negligible.

It should be pointed out that this form of information analysis is similar in many respects to Klatt's sequential analysis of features using a similarity metric.[21] Klatt pointed out that the redundancy of features should be taken into account and that feature analysis should provide some indication of the completeness of a feature system. He also noted that there should be an algorithm for determining whether even better features than those specified exist. The sequential information analysis described here clearly meets the first two criteria and could, in principle, be modified to search for optimum features also.

An illustration of the Sequential Information Analysis (SINFA) is presented in Table XVI, and is based on the data in Table VI from the control experiment. In the first iteration stimulus feature information is computed for each uniquely defined feature in the system. These values appear in the first column and are identical to those in the first column of Table XI. The second column gives the transmitted information in bits per stimulus for each feature. The third column gives the performance measure: percent information transmitted. Since performance is highest on the feature high–anterior, this feature is identified and held constant in all successive iterations.

In the second iteration, conditional stimulus information is calculated for the remaining features with high–anterior held constant. It should be noted that the effect of holding a feature constant is to reduce the stimulus information for each remaining feature or to leave it unchanged; stimulus information cannot be increased by holding another feature constant. The second column gives the conditional transmitted information in bits per stimulus for the features, and the third column gives the percentage of conditional transmitted information. Since in this iteration it is voice which has the highest percentage of information transmitted, it is identified and held constant in all successive iterations.

Entries for remaining iterations are analogous to those for the second iteration. Two characteristics of the analysis are worth noting. First, when any particular feature is held constant, the effect on remaining features is not uniform. For example, in this illustration, the effect of holding high–anterior constant is to substantially reduce the performance level associated with the feature place (MN) (see column 3, Table XVI). This implies that much of the information apparently transmitted by the feature place (MN) was actually redundant information associated with the feature high–anterior which had a higher overall level of performance associated with it. Other features, notably voice, sibilant, back, and frication, are unaffected in that performance is approximately the same for these features in both iterations. It is these "robust" features which are eventually identified as perceptually important in subsequent iterations.

A second consequence of holding features constant is that some of the remaining features in the system may become equivalent, i.e., perfectly redundant with respect to one another. If two features make exactly the same distinctions within each joint category of the

TABLE XVIII. Sequential information analyses for VC-1 syllable set under various listening conditions.

	Speech-to-noise ratio							
	−10dB	−5dB	0dB	+5dB	+10dB	+15dB	All	Quiet
Feature $U(S:R)$	0.140	0.450	0.827	1.616	2.226	2.820	1.024	2.512
Voice	(1)0.034	(1)0.214	(1)0.328	(1)0.512	(1)0.641	(2)0.736	(1)0.342	(4)0.598
Frication								
Sibilant						(3)0.320		(1)0.607
Strident					(4)0.074			
Duration				(5)*0.036*	(2)0.467	(1)0.602		(2)0.251
Place (MN)								
Place (SB)				(4)*0.229*			(4)*0.164*	
Place (W)	(3)0.033	(3)0.074	(5)0.059		(3)*0.769*	(5)*0.230*	(5)*0.031*	(6)*0.266*
High, anterior						(4)0.457		(3)0.428
Back			(4)0.030					
Coronal			(3)0.113	(2)0.348			(2)0.191	
Continuant, open	(2)0.012	(2)0.045	(2)0.136	(3)0.272	(5)*0.051*	(6)*0.210*	(3)0.189	(5)*0.292*
Total	0.079	0.333	0.666	1.397	2.002	2.555	0.917	2.442
Proportion of $U(S:R)$	0.564	0.740	0.805	0.864	0.899	0.906	0.896	0.972

feature(s) held constant, then conditional stimulus feature information, conditional transmitted information, and the ratio of these two quantities will be identical for the two features. In this example, as early as the second iteration, with high–anterior held constant, place (SB) and coronal have become indistinguishable. Other examples may be seen in later iterations. When this happens, and in addition, the features are the highest in percentage of information transmitted, selection of a feature label becomes arbitrary. For example, in the sixth iteration duration and continuant–open are equivalent and highest in terms of the performance criterion. We have arbitrarily selected duration to represent the feature because in most preceding iterations performance was higher for duration than for continuant–open. It should be apparent that this *post hoc* labeling of the features simply represents an expedient for the purposes of summarizing a set of such analyses and in no way affects the analyses themselves. However, the problems of interpreting features identified in later iterations of the analysis cannot be overlooked. Features which are consistently identified only in the later iterations, if at all, are probably of marginal perceptual significance.

Initially, SINFA was performed on confusion matrices for individual subjects (Ss). For each S, matrices were compiled for six S/N ratios by summing across noise levels and replications, and for four noise levels by summing across S/N ratios and replications. Despite the fact that these matrices contain relatively few observations (approximately 1152 and 1728, respectively), highly similar results were obtained for different Ss under the same conditions. Moreover, the pattern of results changed as a function of S/N for all Ss and did not change significantly as a function of overall noise level. We therefore based subsequent analyses on the pooled data in Tables II–IX.

Table XVII presents SINFA summaries for the CV-1 syllable set for various listening conditions. Features identified as high in terms of the performance criterion are indicated in the table together with the iteration in which they were identified. The entry in the table gives the value of (conditional) transmitted information, in bits per stimulus, associated with the feature in that iteration. The rightmost column of this table summarizes the analysis presented in Table XVI. Entries for features which were labeled *post hoc* are italicized to emphasize the arbitrariness of the feature label. Features with no entry in a particular column are those which were not identified by the analysis as making significant[28] independent contributions to the total transmitted information.

SINFA for the CV-1 syllable set reveals that four features consistently contribute to discrimination performance under all listening conditions: voice, sibilant, high–anterior, and frication. A fifth feature, back, is identified as important only at moderately high S/N ratios and in quiet. Interestingly, the order in which the features are identified is quite similar under the various masking conditions, with voice losing its primacy only at +15 dB S/N and in quiet.

With the effects of these five features held constant, place (MN), place (SB), place (W), and coronal become equivalent, as do the features duration and continuant–open. Additional small increments in conditional transmitted information are provided by these features. The only feature in this system which is never identified in the analysis is strident.

Because the entries in Table XVII represent conditional transmitted information, their sum, for a given column, can never exceed the actual transmitted information for the corresponding confusion matrix, $U(S:R)$. When unconditional values of transmitted information are used to describe performance on individual features (cf. Miller and Nicely[1]), this sum can easily exceed the total information transmitted because feature redundancy has not been taken into account. The bottom row of Table XVII expresses the sum of the conditional

TABLE XIX. Sequential information analyses for CV-2 syllable set under various listening conditions.

		Speech-to-noise ratio							
		−10dB	−5dB	0dB	+5dB	+10dB	+15dB	All	Quiet
Feature	$U(S:R)$	0.252	0.809	1.798	2.429	2.974	3.292	1.593	2.653
Vocalic			(3)0.054	(2)0.282	(1)0.409	(1)0.479	(2)0.470	(2)0.237	(2)0.361
Consonantal									
High		(3)0.024	(5)0.071	(5)0.191	(4)0.389	(4)0.517	(7)0.252	(4)0.251	
Low									
Back									
Coronal						(7)0.066	(6)0.086	(7)0.032	(7)0.074
Anterior				(7)0.050	(7)0.088				(6)0.293
Place (W)		(5)0.014	(8)0.013	(8)*0.067*	(9)*0.050*	(8)*0.129*	(9)*0.172*	(8)*0.063*	
Voice		(1)0.076	(1)0.218	(4)0.255	(5)0.308	(6)0.327	(8)0.410	(5)0.233	(8)0.432
Nasal		(2)0.022	(2)0.099	(1)0.309	(3)0.340	(3)0.413	(1)0.501	(1)0.252	(3)0.332
Frication					(8)0.031				
Duration							(5)0.233		(5)0.187
Strident			(6)0.055	(6)0.169	(6)0.233				
Sibilant			(7)*0.011*			(5)0.353	(4)0.497	(6)0.159	(4)0.404
Continuant									
Open		(4)0.034				(9)*0.086*	(10)*0.130*	(9)*0.047*	(9)*0.114*
Round			(4)0.060	(3)0.210	(2)0.371	(2)0.453	(3)0.437	(3)0.200	(1)0.390
Total		0.170	0.581	1.533	2.219	2.827	3.188	1.474	2.587
Proportion of $U(S:R)$		0.675	0.718	0.853	0.914	0.951	0.968	0.925	0.975

terms as the proportion of the total transmitted information accounted for by the features identified in the table. As the amount of transmitted information increases, from one listening condition to another, the proportion of transmitted information explained by the features also increases. At −10 dB S/N, where there is only 0.220 bits of information transmitted (out of a maximum of 4.0 bits), 68.2% of the transmitted information can be accounted for in terms of features. Moreover, the features are the same as those which account for information transmitted under more favorable listening conditions. At +15 dB S/N, where transmitted information is 2.871 bits, approximately 94% of the information transmitted is accounted for by six features.

It is interesting to compare the results at +15 dB S/N with those from the control experiment in quiet. Data for the latter are based on 13 signal levels ranging from 20 to 115 dB. Despite the fact that overall performance based on observations pooled across all levels is lower, the proportion of transmitted information explained is higher. Since this is true for the remaining syllable sets as well, it suggests that discrimination errors made in quiet are slightly more systematic and slightly better explained in terms of features than errors made in the presence of noise distortion.

SINFA summaries for the VC-1 syllable set are presented in Table XVIII. Although the same consonants are involved in both syllable sets, the pattern of results differs considerably from that in Table XVII. Although voice retains its perceptual prominence in the VC-1 set, frication, sibilant, and high–anterior, which were significant for the CV-1 syllables, are rarely identified in this analysis. Instead, continuant–open and place (W) take on perceptual importance for these syllables. It should be noted that although the proportion of information accounted for is similar in both syllable sets, in the VC-1 set there is much less consistency across conditions in terms of which features are important.

Results for the CV-2 syllable set are given in Table XIX. It should be recalled that this set has the highest redundancy and that 17 distinctive features are uniquely defined. As with the CV-1 set, there is considerable consistency in the results across listening conditions. Five features are consistently identified by SINFA as important: vocalic, nasal, round, high, and voice. The first three features were not represented in the CV-1 set and therefore could not be assessed there. In the CV-1 set the confounded features high–anterior were identified as important; in the CV-2 set, where these two features can be distinguished, it is clear that high is the better perceived feature, and that anterior is of marginal significance.

Three additional features, coronal, sibilant, and open, contribute to transmitted information in the CV-2 set, but only at high S/N ratios and in quiet. Frication, duration, and strident, on the other hand, are only occasionally identified by the analysis, and then only in the later iterations.

Those features which were never identified by SINFA in the analysis of CV-2 data include consonantal, low, and back. Although there are no entries in the table for continuant, this is due to the arbitrary choice of open indicated in the table. Finally, it is clear that information concerning place of articulation contributes significantly to transmitted information, although it is difficult to identify the optimal definition for this feature.

Results for the fourth syllable set, VC-2, are presented in Table XX. Here four features are consistently identified by SINFA: voice, nasal, continuant–open,

TABLE XX. Sequential information analyses for VC-2 syllable set under various listening conditions.

	Speech-to-noise ratio							Quiet
	−10dB	−5dB	0dB	+5dB	+10dB	+15dB	All	
Feature $U(S:R)$	0.096	0.524	1.086	1.693	2.340	2.838	1.097	2.553
Voice	(2)0.012	(2)0.079	(2)0.209	(2)0.333	(2)0.461	(4)0.574	(2)0.220	(5)0.421
Nasal	(1)0.020	(1)0.190	(1)0.348	(1)0.478	(1)0.541	(1)0.593	(1)0.296	(1)0.493
Frication		(4)0.017	(4)0.055	(3)0.233	(3)0.320	(6)*0.190*	(3)0.146	(6)0.196
Sibilant			(7)*0.019*		(6)*0.137*	(2)0.556	(7)*0.054*	
Strident								
Duration						(3)0.248		(4)0.423
Place (MN)								
Place (SB)					(7)*0.107*	(7)*0.136*	(8)*0.049*	
Place (W)	(3)0.022	(5)0.041		(5)0.296				(7)*0.167*
High, anterior					(5)0.320	(5)0.366	(6)0.067	(2)0.622
Back			(5)0.037				(5)0.052	(3)0.157
Coronal			(6)0.049					
Continuant, open	(4)0.004	(3)0.053	(3)0.154	(4)0.143	(4)0.236		(4)0.117	
Total	0.058	0.380	0.871	1.483	2.122	2.663	1.001	2.479
Proportion of $U(S:R)$	0.604	0.725	0.802	0.876	0.907	0.938	0.912	0.971

and frication. Although voice and continuant-open were also important for the VC-1 syllable set, frication, which clearly contributes to transmitted information here, was never identified in the VC-1 set. Another similarity between the VC-2 and VC-1 sets is that there is little consistency among syllable sets and conditions concerning which definition of place is optimal.

From the analyses presented above, it is clear that the particular features which result in high levels of performance vary significantly from one syllable set to another and in some cases vary within syllable sets as a function of listening conditions. It was argued in the Introduction that if natural perceptual features exist, this should not occur, i.e., there should be a consistent tendency for certain features to account for transmitted information better than other, closely related, features. Two factors in the present experiments, however, probably contribute to and exaggerate the observed inconsistencies. The first factor involves the syllables themselves. Since individual syllables were recorded rather than spoken live, the type-token ratio for all syllables was 1.00. Although there were three syllables for each consonant (one for each vowel), it is possible that during the course of the experiment Ss learned to take advantage of the idiosyncrasies of individual syllables in order to improve their performance. The second factor involves syllable set size. In Sec. I-B it was noted that the use of different consonant subsets would permit context effects to emerge in the feature analyses. Although the presence of context effects in itself argues against a strong form of the natural features hypothesis, the use of consonant subsets may have encouraged Ss to adopt specific listening strategies for each syllable set, whereas a full syllable set would produce listening strategies more closely approximating those normally used by listeners to discriminate speech sounds.

One way to explore these possibilities is to use SINFA to analyze the perceptual confusion matrices reported by other researchers. Especially relevant are the data from Millet and Nicely,[1] Singh and Black,[2] and Graham and House.[6] These analyses were performed and are summarized in Table XXI.

The consonants studied by Miller and Nicely[1] were quite similar to those in our CV-1 syllable set; the only difference is that in the CV-1 set /tʃ/ and /dʒ/ are substituted for /m/ and /n/. The principal difference between the two experiments is that Miller and Nicely used only one vowel, /ɑ/, and that their syllables were presented by live voice. The results in Table XXI are based on the pooled data for six S/N ratios at the widest bandpass condition, 200–6500 Hz. The results are in very close agreement with those for the CV-1 set (see Table XVII). Although nasal is not distinctive in the CV-1 set, the fact that it is the best perceived feature in Miller and Nicely's experiment is in agreement with comparable data from our experiment (see Tables XIX and XX). Features which are identified as important in both analyses are voice, sibilant, frication, and a nonunique place feature which is identified only in the last iteration. One difference between the two sets of results is that high–anterior is not identified as important for the data of Miller and Nicely. The similarity between the two sets of data suggests, however, that where the composition of two syllable sets is similar, the resulting pattern of confusions will be similar, and that the effects of a high type-token ratio on perceptual confusions are minimal.

In their study of perceptual confusions among intervocalic consonants, Singh and Black[2] used callers and listeners from four language groups. They also used both English and non-English phonemes as stimuli. For purposes of comparison, we have selected their data for English callers and listeners, and we have deleted entries for non-English phonemes from their confusion matrix. The resulting confusion matrix is based on data for 21 consonants in combination with two vowels, /i/ and /ɑ/, and most closely resembles our CV-2 syllable set in the present experiment.

TABLE XXI. Sequential information analyses for confusion matrices of Miller and Nicely,[1] Singh and Black,[2] and Graham and House.[6]

Miller and Nicely		Singh and Black		Graham and House	
Feature	$U(S:R) = 1.559$	Feature	$U(S:R) = 3.644$	Feature	$U(S:R) = 3.138$
High, anterior		Vocalic	(2)0.405	Vocalic	
Back		Consonantal		Consonantal	(7)*0.103*
Coronal		High		High	
Place (MN)		Back	(4)0.246	Back	
Place (SB)	(5)*0.284*	Coronal		Coronal	
Place (W)		Anterior		Anterior	(5)0.566
Voice	(2)0.477	Place (SB)	(6)0.934	Place (W)	(10)*0.281*
Nasal	(1)0.299	Place (W)	(8)*0.059*	Voice	(4)0.554
Frication, continuant, open	(4)0.197	Voice	(5)0.627	Nasal	(8)0.194
		Nasal	(1)0.434	Frication	(3)0.380
Duration, sibilant	(3)0.262	Frication	(3)0.610	Duration	(2)0.159
Strident		Duration		Strident	
Total	1.519	Strident		Sibilant	(1)0.646
Proportion of $U(S:R)$	0.974	Sibilant		Continuant	
		Continuant		Open	(9)*0.103*
		Open	(7)*0.089*	Round	(6)*0.135*
		Total	3.404	Total	3.121
		Proportion of $U(S:R)$	0.934	Proportion of $U(S:R)$	0.995

Comparison of results for our CV-2 set with those for the Singh and Black data reveal both similarities and differences. Nasal, vocalic, and voice are important in both. Also, a place feature is identified in both in later iterations, as is a confounded continuant–open feature. The two features, frication and sibilant, on the other hand, which are very closely related, are each identified as important in one analysis and not the other. Similarly, back is important in Singh and Black's data, whereas high (and in quiet, anterior) is important in our CV-2 set.

The third analysis is based on perceptual data taken from Graham and House.[6] These authors obtained same–different judgments from young children for pairs of nonsense syllables, and generated a triangular matrix of error rates for each sound pair. From this matrix we have generated a quasi-confusion matrix by converting entries on the main diagonal to rates of correct response, and by using off-diagonal error rates to estimate frequencies of perceptual confusion.

The SINFA summary for this matrix is presented in the third column of Table XXI. Three features, sibilant, duration, and frication are identified as most important in the first three iterations. Although sibilant and frication have been identified in other analyses, this result is surprising because these three features are very closely related, as can be seen in Table X. Another difference between these results and others is the failure of nasal to be important. In contrast, anterior, which did not appear to be of significance in other analyses, is identified here in the fifth iteration. The rather atypical results for the Graham and House data may be due to the fact that the Ss were children or they may be due to our manipulation of the same–different judgments in producing the quasi-confusion matrix.

III. DISCUSSION AND CONCLUSIONS

Two main points emerge from the analyses presented in Sec. II. First, it is possible to account for a large proportion of transmitted information in terms of articulatory and phonological features. In instances where the rate of transmitted information is high (better than 50% or 2 bits), it is not uncommon for the features to account for more than 90% of the information transmitted. These high values are due in part, however, to the leniency of the SINFA decision criterion. Features are identified and iterations continued as long as the conditional transmitted information for the best feature is at least 1% of the total transmitted information. A more conservative criterion, say 5%, would result in an earlier termination of the analysis and the identification of fewer features. It would also help to avoid the problems in interpreting features identified in later iterations.

The second major finding is that if natural perceptual features do exist, then there is little evidence in confusion-matrix data to support their existence. When the results of all SINFA summaries are considered together, the features defined in Table X tend to fall into several categories.

The first category contains the features nasal, voice, and round. Nasal and voice are features which have appeared in all feature systems suggested in the literature, and there are no alternative formulations of the distinctions carried by these two features. Because of this, and because both features are well perceived both in noise and in quiet, they are identified as perceptually important in every syllable set where they are distinctive. Similarly, the feature round is not closely related to other features and is well perceived.

The second category includes two features which do not appear to be perceptually important. Strident is rarely identified by SINFA in analyses of our data and it is likewise passed over in the analyses presented in Table XXI. Although low is distinctive in only one of the syllable sets studied, CV-2, it was never identified by SINFA. At least for perception of consonants, there is little evidence to support the relevance of strident and low as perceptual features.

All of the remaining features fall into an intermediate category, since they appear to be relevant in some cases and not others. High–anterior, for example, is a strong perceptual feature in our CV-1 syllable set but not elsewhere. Frication is important in the CV-1 syllable set and in the VC-2 syllable set, and in the analyses in Table XXI, but not in the VC-1 and CV-2 syllable sets. In fact, most of the variables we manipulated, syllable set, S/N, and noise versus quiet, produced some change in feature identification.

Particularly difficult to evaluate are those features which are not unique in the iteration during which they are identified by SINFA. In most analyses where one of the place features is identified as important, its identification typically occurs in a late iteration. When this happens, place (MN), place (SB), place (W), and coronal are usually equivalent. Therefore it is a matter of indifference which of these four feature labels is used to represent the feature. A similar confounding is common for duration and continuant–open. Thus, unless a feature is sufficiently strong perceptually to be identified before it becomes confounded with other closely related features, its perceptual status remains equivocal. Of the features we have studied, place (MN) is the only one which is never unequivocally identified by SINFA.

A consequence of these findings is that for a given confusion matrix, several different feature systems can be used to account for the information transmitted and all will work equally well. Consider the data from Singh and Black in Table XXI. Retaining the features identified in the first six iterations, we find that in the seventh iteration continuant and open are equivalent, and hence a feature system including either one would produce the same result in this iteration. In the eighth iteration duration, strident, sibilant, and place (W) are equivalent and any one of these features would produce the same result in this iteration. There are, therefore, eight different pair combinations of these features which could be substituted in an eight-feature system without altering the SINFA results. We performed an analysis using only the features defined by Singh and Black: voice, nasal, frication, place (SB), duration, and vocalic. Since these features were also identified by SINFA, it is not surprising that this system accounted for 93.8% of the information transmitted. Also, we analyzed the Singh and Black matrix using five features derived from an INDSCAL analysis of perceptual data by Singh, Woods, and Becker,[16] but found that these features accounted for only 77.7% of the information transmitted.

Since it can be shown that several feature systems will account equally well for the transmitted information in a given matrix, and since there is no tendency for particular features or feature systems to do better than others across conditions and syllable sets, the hypothesis that natural perceptual features exist is not supported. This conclusion is consistent with the fact that multidimensional scaling analyses of perceptual confusions and judgments of similarity also do not converge on a particular set of perceptual features which are independent of context, experimental conditions, and experimental task. Voice, nasal, and possibly round, appear to be the only exceptions to this rule. Thus, although articulatory and phonological features are extremely useful for describing patterns of perceptual data and for indicating which acoustic cues are most important in a given context, the notion that these features represent hypothetical perceptual constructs is open to serious question.

ACKNOWLEDGMENTS

This research was supported in part by Grant No. 14-P-55126/3 from the Social and Rehabilitative Service of HEW. Many of the facilities used in this research were the results of gifts from the Richard King Mellon Charitable Trusts, Inc. and the Edwards–Lazear Foundation and Grant No. NS 07790 from NINDS. The authors want to express their appreciation to Linet J. Feigel, Walt Jesteadt, Charlotte M. Reed, and Richard V. Wolf for their contributions to the collection and analysis of these data.

---

*Present address: Department of Psychology, University of Maryland, Baltimore County, 5401 Wilkens Avenue, Baltimore, Maryland 21228.

†Please send reprint requests to Marilyn D. Wang.

[1] G. A. Miller and P. E. Nicely, "An Analysis of Perceptual Confusions Among Some English Consonants," J. Acoust. Soc. Am. 27, 338–352 (1955).

[2] S. Singh and J. W. Black, "Study of Twenty-Six Intervocalic Consonants as Spoken and Recognized by Four Language Groups," J. Acoust. Soc. Am. 39, 372–387 (1966).

[3] S. Singh, "Crosslanguage Study of Perceptual Confusion of Plosive Phonemes in Two Conditions of Distortion," J. Acoust. Soc. Am. 40, 635–656 (1966).

[4] R. Ahmed and S. S. Agrawal, "Significant Features in the Perception of (Hindi) Consonants," J. Acoust. Soc. Am. 45, 758–763 (1969).

[5] J. P. Gupta, S. S. Agrawal, and R. Ahmed, "Perception of (Hindi) Consonants in Clipped Speech," J. Acoust. Soc. Am. 45, 770–773 (1969).

[6] L. W. Graham and A. S. House, "Phonological Oppositions in Children: A Perceptual Study," J. Acoust. Soc. Am. 49, 559–566 (1971).

[7] R. S. Tikofsky and J. R. McInish, "Consonant Discrimination by Seven Year Olds," Psychon. Sci. 10, 61–62 (1968).

[8] J. R. McInish and R. S. Tikofsky, "Distinctive Features and Response Latency: A Pilot Study," Percept. Psychophys. 6, 267–268 (1969).

[9] S. Singh, "A Distinctive-Feature Analysis of Responses to a

Multiple-Choice Intelligibility Test," Int. Rev. Appl. Ling. VI/I, 37–53 (1968).

[10] R. A. Cole and B. Scott, "Distinctive Feature Control of Decision Time: Same-Different Judgments of Simultaneously Heard Phonomes," Percept. Psychophys. **12**, 91–94 (1972).

[11] W. Wickelgren, "Distinctive Features and Errors in Short-Term Memory for English Consonants," J. Acoust. Soc. Am. **39**, 388–398 (1966).

[12] Although the data which Wickelgren reported were confusions in short-term memory rather than perceptual confusions, his technique for analyzing the matrix is equally applicable to perceptual data and is therefore included here.

[13] K. V. Wilson, "Multidimensional Analyses of Confusions of English Consonants," Am. J. Psychol. **76**, 89–95 (1963).

[14] S. Singh and D. R. Woods, "Perceptual Structure of 12 American English Vowels," J. Acoust. Soc. Am. **49**, 1861–1866 (1971).

[15] S. Singh, D. R. Woods, and A. Tishman, "An Alternative MD-SCAL Analysis of the Graham and House Data," J. Acoust. Soc. Am. **51**, 666–668(L) (1972).

[16] S. Singh, D. R. Woods, and G. M. Becker, "Perceptual Structure of 22 Prevocalic English Consonants," J. Acoust. Soc. Am. **52**, 1698–1713 (1973).

[17] F. F. Weiner and S. Singh, "Perceptual Features of Nine English Consonants Determined by Choice Reaction Time," J. Acoust. Soc. Amer. (to be published).

[18] S. Singh and K. Singh, "A Search for the Perceptual Features of Hindi Consonants," J. Acoust. Soc. Amer. (to be published).

[19] M. Wish, "An INDSCAL Analysis of the Miller-Nicely Consonant Confusion Data," J. Acoust. Soc. Am. **49**, 85(A) (1971).

[20] R. W. Peters, "Dimensions of Perception for Consonants," J. Acoust. Soc. Am. **35**, 1985–1989 (1963).

[21] D. H. Klatt, "Structure of Confusions in Short-Term Memory Between English Consonants," J. Acoust. Soc. Am. **44**, 401–407 (1968).

[22] H. M. Walker and J. Lev, *Statistical Inference* (Henry Holt and Company, New York, 1953).

[23] I. Pollack and J. M. Pickett, "Masking of Speech by Noise at High Sound Levels," J. Acoust. Soc. Am. **30**, 127–130 (1958).

[24] D. D. Dirks, R. H. Wilson, and D. R. Bower, "Effect of Pulsed Masking on Selected Speech Materials," J. Acoust. Soc. Am. **46**, 898–906 (1969).

[25] N. Chomsky and M. Halle, *The Sound Pattern of English* (Harper and Row, New York, 1968).

[26] W. R. Garner, *Uncertainty and Structure as Psychological Concepts* (Wiley, New York, 1962). This book is a general reference for uncertainty analysis and provides a comprehensive treatment of the topics discussed in this section.

[27] W. R. Garner and J. Morton, "Perceptual Independence: Definitions, Models, and Experimental Paradigms," Psychol. Bull. **72**, 233–259 (1969).

[28] The term "significant" is not used here in the statistical sense, since no statistical tests were incorporated in the information analysis.