NRL Report 8676

Controlled Test Procedures for Using Intervocalic Consonants to Assess Speech Intelligibility: A Feasibility Study

ASTRID SCHMIDT-NIELSEN

Communications Systems Engineering Branch
Information Technology Division

April 27, 1983

NAVAL RESEARCH LABORATORY
Washington, D.C.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER NRL Report 8676	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CONTROLLED TEST PROCEDURES FOR USING INTERVOCALIC CONSONANTS TO ASSESS SPEECH INTELLIGIBILITY: A FEASIBILITY STUDY		5. TYPE OF REPORT & PERIOD COVERED Interim report on a continuing NRL problem. 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Astrid Schmidt-Nielsen		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Naval Research Laboratory Washington, DC 20375		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 75-0129-0-1; 75-1596-0-2; PE61153N RR021-05-42
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Washington, DC 22222		12. REPORT DATE April 27, 1983 13. NUMBER OF PAGES 22
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Consonant identification
Speech intelligibility
Digital voice processor intelligibility

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

A preliminary version of an intelligibility test based on intervocalic consonant excised from connected speech was developed and compared with the Diagnostic Rhyme Test (DRT), which is based on initial consonants of syllables spoken in isolation. The intervocalic consonants showed different patterns of losses than the DRT when the speech was processed through four digital voice processors. Intelligibility losses under conditions of noise and low-pass filtering were also investigated.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

CONTENTS

CONTROLLED TEST PROCEDURES FOR USING INTERVOCALIC CONSONANTS TO ASSESS SPEECH INTELLIGIBILITY: A FEASIBILITY STUDY

INTRODUCTION

Phoneme intelligibility tests such as the Diagnostic Rhyme Test (DRT) [1] or the Modified Rhyme Test (MRT) [2] are highly reliable ways of measuring the intelligibility of voice communication systems. With tape-recorded test materials and carefully controlled test procedures, it is possible to obtain scores that are repeatable to within one or two points. Total scores on the DRT and the MRT correlate highly with one another, and both of these tests also correlate highly with other tests such as the Phonetically Balanced (P-B) words [3], although scores on the P-B test tend to be more variable than on the DRT and MRT. The DRT has a number of advantages over other existing intelligibility tests and serves as the model for the development of the test to be described in this report. The DRT is known to yield highly repeatable scores. The DRT has been widely used within DoD for tests of digital voice equipment [4], and a large data base exists for a wide variety of voice systems and conditions. A standard set of tape recordings based on up to 18 speakers and including a variety of background noises of military interest is also in existence. Another significant advantage of the DRT is that it also provides diagnostic subscores based on six distinctive phonemic features: voicing, nasality, sustention, sibilation, graveness, and compactness.

A disadvantage of the DRT is that it tests only initial consonants of carefully pronounced words spoken in isolation. There is good reason to believe that the cues used to recognize speech sounds in running speech are not the same as those used for carefully pronounced isolated words, and the cues for consonants at the beginning of a word are not the same as for medial and final consonants. Voiers [5] has shown that although tests using word-final consonants yield somewhat lower scores than initial consonants, overall scores are very highly correlated. However, one would expect to find that feature scores would be quite different for initial and noninitial consonants even though average scores are highly correlated.

Since total scores on a different intelligibility measures are so highly correlated with one another, for comparison purposes the test that is most reliable is generally to be preferred. However, a major use of voice system testing is to evaluate specific weaknesses in order to improve performance. The diagnostic feature scores on the DRT offer more to meet this need than other tests. Since voice communications usually involve connected speech as well as isolated words, it would be extremely useful to have a reliable test that is based on the voice cues that are used in connected speech. Such a test would uncover different patterns of equipment weaknesses than the DRT and would be a good supplement to standard DRT scores.

Pols [6] collected samples of informal conversational speech and removed vowel-consonant-vowel (VCV) segments using computer-controlled waveform editing techniques. Subjects were asked to identify the central consonants in the excerpts under a variety of conditions including various forms of degradation. Since phonemes are not as carefully articulated in conversational speech as in isolated word units, identification was less than perfect to begin with and deteriorated markedly under noise degradation.

Manuscript approved November 23, 1982.

The present research explored the feasibility of using segments excised from connected speech in a controlled intelligibility test. The test procedures were based on those of the DRT in order to develop a test that could be used under the same repeatable, carefully controlled conditions as the DRT.

VOICE MATERIALS

The Diagnostic Rhyme Test is based on contrasting word pairs that differ only in the initial consonants, which are separated by a single phonemic feature (e.g., Moon-Boon, which differs in the presence or absence of the feature nasality). Since everyday conversations such as those used by Pols are not likely to carry pairs of consonants separated by a single feature and surrounded by the same two vowels, paragraphs and sentences carrying the contrasting consonants were developed for reading.

Forty-six phoneme contrast pairs were selected for testing. These are shown in Table 1 and are grouped according to the feature classification used on the DRT. All but a few of these pairs can be described as differing by a single phonemic feature. A few contrasts that were of interest are not as simply characterized in the binary feature system of Jakobson, Fant, and Halle [7] that formed the basis for the DRT. Where feature comparisons are of interest, these have been classified in the most appropriate category when they differed by more than one distinctive feature.

A set of matched paragraph and sentence materials was developed in which each member of a phoneme pair could occur in the same surrounding context. For example, the following two sentences occurred in different versions of the text:

> The company <u>b</u>oasted a profit gro<u>ss</u> of three <u>m</u>illion dollars last year.
>
> The company <u>p</u>osted a profit grow<u>th</u> of three <u>b</u>illion dollars last year.

Three contrasts are included here: /b/-/p/, /s/-/θ/, and /m/-/b/. The sentences were constructed so that the excised VCV would include normal cues to consonant identity (coarticulation, vowel duration, etc.) while all possible precautions were taken to eliminate accidental cues that might differentiate the pairs on irrelevant grounds: each contrast pair occurred in the same place in the sentence so the prosody would be the same. Identical vowels surrounded the two consonants in the pair, and the same consonants flanked the VCV sequence. In five cases, the two sets of materials had a difference in the third phoneme from the test consonant (e.g., pre<u>ss</u>ure-trea<u>s</u>ure for the /ʃ/-/ʒ/ contrast), but with only the VCV portion excised from the sentence, it is doubtful that there were any usable cues carried over from the discarded portion. The complete texts for the two versions are given in the appendix.

Tape recordings were made of two readers, one male and one female, reading the texts. The readers were instructed to read "naturally, as though you were reading to someone" and not to try to articulate extra precisely. Each reader read one version of the entire text followed by the second version, and then read both versions a second time. This procedure avoided any special emphasis or effort on the test consonants, which were not marked in the text in any way.

The tape recordings were digitized at 12,000 bits per second (bps) and computer edited using the Interactive Laboratory System (ILS) software developed by Signal Technology, Inc. The editing of each contrast pair proceeded separately for each reader as follows: First four excerpts, including some surrounding text were assembled (from both readings for each of the two consonants). The four waveforms were displayed simultaneously on the CRT and two observers listened to each stimulus while viewing the displayed waveforms. The two contrasting consonants that were embedded in the most similar surrounding contexts were selected, and beginning and end points for the VCV segments were chosen. Each of the two selected excerpts was then displayed in turn, and the segment between the selected endpoints was stored and labeled. The two excised VCVs were checked both by visual

Table 1 — Intervocalic Phoneme Pairs and DRT Pairs Arranged by Feature Contrast (whole words are given for intervocalic pairs, but only the VCV portion was heard by the listeners)

Phoneme Contrast	VCV Pairs	DRT Pairs
<u>Voicing</u>		
/b/ - /p/	COMPANY BOASTED – COMPANY POSTED	BOND – POND
		BEAN – PEEN
/d/ - /t/	MEAD AS – MEET AS	DAUNT – TAUNT
		DINT – TINT
		DENSE – TENSE
		DUNE – TUNE
/g/ - /k/	WEEKLY GOLD – WEEKLY COLD	GOAT – COAT
		GAFF – CALF
/v/ - /f/	BELIEVE I – BELIEF I	VOLE – FOAL
		VAST – FAST
		VAULT – FAULT
		VEAL – FEEL
/ð/ - /θ/	EITHER – ETHER	
/z/ - /s/	RAISING – RACING	ZED – SAID
		ZOO – SUE
/ʒ/ - /ʃ/	TREASURE – PRESSURE	
/d/ - /tʃ/	MIDGE AND – MITCH AND	GIN – CHIN
		JOCK – CHOCK
<u>Nasality</u>		
/m/ - /b/	THREE MILLION – THREE BILLION	MOOT – BOOT
		MOAN – BONE
		MAD – BAD
		MOSS – BOSS
		MEND – BEND
		MITT – BIT
		MOM – BOMB
		MEAT – BEAT
/n/ - /d/	RON AND – ROD AND	GNAW – DAW
		NECK – DECK
		NIP – DIP
		KNOCK – DOCK
		NEED – DEED
		NEWS – DUES
		NOTE – DOTE
		NAB – DAB
/ŋ/ - /g/	LONG INTERVALS – LOG INTERVALS	
<u>Sustention</u>		
/v/ - /b/	MY VOTE – MY BOAT	VEE – BEE
		VILL – BILL
		VON – BON
		VOX – BOX
/ð/ - /d/	BREATHING – BREEDING	THOSE – DOZE
		THOUGH – DOUGH
		THEN – DEN
		THAN – DAN
/f/ - /p/	BY FERRY – BY PERRY	FOO – POO
		FENCE – PENCE
/θ/ - /t/	BETHIE – BETTY	THICK – TICK
		THONG – TONG
/ʒ/ - /dʒ/	EROSION – THE TROJAN	
/ʃ/ - /tʃ/	WASHING – WATCHING	SHEET – CHEAT
		SHOES – CHOOSE
		SHAW – CHAW
		SHAD – CHAD

Table 1 (Continued) — Intervocalic Phoneme Pairs and DRT Pairs Arranged by Feature Contrast (whole words are given for intervocalic pairs, but only the VCV portion was heard by the listeners)

Phoneme Contrast	VCV Pairs	DRT Pairs
Sibilation		
/z/ - /ð/	CLOSING — CLOTHING	ZEE — THEE
/s/ - /θ/	GROSS OF — GROWTH OF	SING — THING
		SOLE — THOLE
		SAW — THAW
		SANK — THANK
/dʒ/ - /g/	SLUDGE IN — SLUG IN	JUICE — GOOSE
		JILT — GILT
		JOE — GO
		JEST — GUEST
		JAWS — GAUZE
		JAB — GAB
		JOT — GOT
/dʒ/ - /d/	AGENDA — ADDENDA	
/tʃ/ - /k/	LEACHING — LEAKING	CHEEP — KEEP
		CHOO — COO
		CHAIR — CARE
		CHOP — COP
/tʃ/ - /t/	H-INSTANT — EIGHT-INSTANT	
/z/ - /d/	LAZY — LADY	
/s/ - /t/	STUDY SEALS — STUDY TEALS	
Graveness		
/v/ - /z/	HAVE EXTRA — HAS EXTRA	
/f/ - /s/	RELIEF UNIT — RELEASE UNIT	
/m/ - /n/	JIMMY — GINNY	MOON — NOON
		MET — NET
/b/ - /d/	RUBY — RUDY	BID — DID
		BOWL — DOLE
		BONG — DONG
		BANK — DANK
/p/ - /t/	REPORT — RETORT	PEAK — TEAK
		POOL — TOOL
		PENT — TENT
		POT — TOT
/f/ - /θ/	WE FOUGHT — WE THOUGHT	FIN — THIN
		FORE — THOR
		FOUGHT — THOUGHT
		FAD — THAD
/w/ - /r/	STOW ALL — STORE ALL	WEED — REED
		WAD — ROD
/w/ - /l/	TOO WEAK — TWO LEAK	
/l/ - /r/	LEVEL IS — LEVER IS	
/v/ - /ð/	MOVING — SMOOTHING	
Compactness		
/ŋ/ - /m/	HANGERS — HAMMERS	
/ŋ/ - /n/	RANG ALL — RAN ALL	
/g/ - /b/	EXTRA GASKETS — EXTRA BASKETS	GHOST — BOAST
		GAT — BAT
/g/ - /d/	SEE GAIL — SEE DALE	GILL — DILL
		GOT — DOT
/k/ - /p/	SOAKING — SOAPING	COOP — POOP
		KEG — PEG
/k/ - /t/	THE CAN — THE TAN	KEY — TEA
		CAUGHT — TAUGHT
/h/ - /p/	YOU HOLLY — YOU POLLY	HOP — POP
/h/ - /f/	YOU HOLD — YOU FOLD	HIT — FIT
/j/ - /w/	FREDDY YU — FREDDY WU	YIELD — WIELD
		YAWL — WALL
/j/ - /r/		YOU — RUE
		YEN — WREN
/ʃ/ - /s/	PRESSURE — PRESSER	SHOW — SO
		SHAG — SAG
/ʒ/ - /z/	ROUGE ON — BRUISE ON	

inspection of the waveforms and by listening, and if necessary the editing process was repeated until both observers were satisfied. All editing took place at zero-crossings in order to avoid extraneous clicks and pops. The beginning and end points were always selected with reference to the consonants before and after the VCV segment so that both temporal and coarticulatory information in the vowels was preserved. In the /d/-/t/ contrast, for example, "meat as" and "mead as" were cut at the end of the nasalized /m/ portion and just before the frication for the /s/ began. Owing to the effects of coarticulation, the identity of the surrounding consonants was recognizable for some pairs, but since both members of a pair were cut at nearly identical points on the waveform, both VCVs were alike in this respect and differed only in their center consonants. For a few phoneme pairs, the original readings did not yield two closely matched tokens, either because of level differences or changes in rate or emphasis. These were rerecorded by the readers and edited as above.

Analog stimulus tapes were generated by a program that randomly assigned one member of each pair to the first sublist and the other member to the second. A full list consisted of four sublists, two for each reader, so that one full list included every consonant for both speakers. The lists were assembled in the order male-first sublist, female-first sublist, male-second sublist, female-second sublist. This made it extremely unlikely that a listener would realize that the second sublist contained the items not on the first, or would be able to remember the first half even if the list construction were known. The lists were output to magnetic tape and recorded on an Ampex tape recorder. When digitizing and when converting back to analog, the signal was passed through a 6000 Hz low-pass filter to avoid quantization noise and aliasing. The average duration of the VCV excerpts was 0.3 s, and there was 1.1 s of silence between stimuli, so that the rate of one item every 1.4 s was the same as the rate for the DRT.

Different randomized tapes were processed through four digital voice processors: linear predictive coding (LPC) at 2.4 kilobits per second (kbps), adaptive predictive coding (APC) at 9.6 kbps, continuously variable slope delta modulation (CVSD) at 16 and 32 kpbs [8]. The output was recorded on an Ampex tape recorder, and these four tapes and an unprocessed recording of the stimuli constituted the test tapes. Additional randomizations were used for practice lists. One practice list was recorded with 2.0 s instead of 1.1 s of silence between stimuli in order to give the listeners extra time the first time they heard the lists.

Three experiments were carried out using these tape recordings. The first experiment evaluated the usefulness of the technique and explored the confusions made on the individual VCV pairs. The second experiment compared VCV results with DRT scores on the same voice systems, and the third experiment explored the effects of noise and bandpass limiting on VCV intelligibility. The tests for the first experiment were conducted using naive listeners, and the tests for the second and third experiments were carried out by Dynastat, Inc. using their test crews, who are highly trained on the DRT.

EXPERIMENT I

The processed and unprocessed tape recordings were evaluated by using a set of inexperienced listeners to determine how well they could recognize the consonants in the excised VCV segments. These tests were intended to evaluate possible shortcomings of the test procedure and also to determine which phoneme contrasts were more readily confused than others.

Method

Subjects were 25 University of Maryland students who volunteered to participate for extra course credit in psychology courses. Non-native speakers of English were excluded from the data. Subjects were tested in groups of one to five and heard one of five counterbalanced orders of the five test lists.

The subjects were told that they were going to hear word fragments such as "eebo" or "eepo" and that the fragments had been taken from naturally spoken sentences. Before testing began, the experimenter went over the answer form illustrated in Fig. 1. Each phoneme contrast was explained, and the words from which the sounds had been taken were read to the subjects. For each fragment they heard, the subjects were to listen for the consonant sound and mark the word the fragment sounded like.

VCV ANSWER SHEET

K - T	THE CAN - THE TAN	T - S	STUDY TEALS - STUDY SEALS
F - P	BY FERRY - BY PERRY	D - DH	BREEDING - BREATHING
W - R	STOW ALL - STORE ALL	Z - S	RAISING - RACING
NG - M	HANGERS - HAMMERS	D - Z	LADY - LAZY
S - F	RELEASE UNIT - RELIEF UNIT	TH - T	BETHIE - BETTY
G - J	SLUG IN - SLUDGE IN	N - M	GINNY - JIMMY
P - K	SOAPING - SOAKING	NG - G	LONG INTERVALS - LOG INTERVALS
K - CH	LEAKING - LEACHING	H - F	YOU HOLD - YOU FOLD
SH - S	PRESSURE - PRESSER	D - B	RUDY - RUBY
DH - V	SMOOTHING - MOVING	Z - ZH	BRUISE ON - ROUGE ON
W - L	TOO WEAK - TWO LEAK	TH - DH	ETHER - EITHER
R - L	LEVER IS - LEVEL IS	T - D	MEAT AS - MEAD AS
SH - CH	WASHING - WATCHING	T - CH	EIGHT INSTANT - H INSTANT
Y - W	FREDDY YU - FREDDY WU	J - ZH	TROJAN - EROSION
P - T	REPORT - RETORT	N - NG	RAN ALL - RANG ALL
J - D	AGENDA - ADDENDA	J - CH	MIDGE AND - MITCH AND
F - V	BELIEF I - BELIEVE I	N - D	RON AND - ROD AND
P - B	COMPANY POSTED - COMPANY BOASTED	DH - Z	CLOTHING - CLOSING
TH - S	GROWTH OF - GROSS OF	B - V	MY BOAT - MY VOTE
B - M	THREE BILLION - THREE MILLION	K - G	WEEKLY COLD - WEEKLY GOLD
F - TH	WE FOUGHT - WE THOUGHT	Z - V	HAS EXTRA - HAVE EXTRA
SH - ZH	PRESSURE - TREASURE	D - G	SEE DALE - SEE GAIL
P - H	YOU POLLY - YOU HOLLY	G - B	EXTRA GASKETS - EXTRA BASKETS
		SH - S	THE SHEETS - THE SEATS

Date ____________________

Initials ________________

Test Number ______________

Fig. 1 — Sample answer form

The tapes were played on a Nagra IVS tape recorder, and subjects listened using KOSS PRO 4AA headphones. To familiarize the subjects with the task and to eliminate initial learning effects, there were three practice lists before the five test lists. The first practice list was at a slower rate (one item approximately every 2.3 s), and the remaining two lists were at the normal test rate (one item approximately every 1.4 s).

Results

Scores were computed in terms of percent correct responses with the correction for guessing: % correct = (Right − Wrong)/Total × 100. After a few practice trials, the subjects did not find the task difficult in spite of the sometimes odd-sounding fragments. There was a steady improvement over the three practice trials as shown in Table 2. The second half of the table shows the average performance on the last 5 trials. These scores are lower than for the first trials since they include scores for the four processed tapes, but since processors were balanced across trials for different groups of subjects, the processor effects are the same across trials and only additional learning effects influence the average scores. Analysis of variance showed a significant learning effect for the first three trials, $F(2, 50) = 6.93$, $p < 0.01$, and a nonsignificant effect for the last five trials, $F(4, 100) < 1$.

Table 2 — The Effect of Learning on VCV Performance; Average Scores Over Trials (Trials 5 to 8 include processed tapes and consequently have lower average scores)

Trial	% Correct	Trial	% Correct
1	84.3	4	81.0
2	86.3	5	82.2
3	88.5	6	83.9
		7	82.8
		8	83.6

The relative scores for the different voice processors and the unprocessed speech showed the pattern one might expect—decreasing scores with decreasing data rate. Stated in percent correct: Unprocessed, 92.1; CVSD 32, 90.0; CVSD 16, 87.5; APC 9.6, 82.4; LPC 2.4, 65.4. The differences were statistically significant based on an analysis of variance, $F(4, 100) = 185.9$, $p < 0.001$. This overall result was similar to what might be expected from knowledge of the voice processors. It is the detailed analysis of the confusions that is of more interest, and comparisons with DRT results are made in the discussion of the second experiment.

Confusion matrices (total errors out of 50 possible), for the different voice systems are shown in Figs. 2 through 6. It can be seen that even for the unprocessed speech some pairs were more difficult than others. There are many possible reasons for these differences: some sounds are inherently more confusable than others, a particular sound may have been less carefully articulated than the rest, the position of the phoneme in the word is important, the removal of cues form the sentence context may affect some sounds more than others. The ten phoneme pairs with the most errors for each processor are given in Table 3. To the extent that the pattern of errors differs among processors or between processed and unprocessed speech, specific weaknesses of individual processors are indicated. The two CVSD processors had similar patterns of confusion whereas quite different pairs gave problems with the two other processors. Three of the most difficult pairs for the unprocessed speech involved the phoneme /ʒ/, which is infrequent in English and consequently would be less familiar to the subjects. The /d/-/n/ contrast occurred in the words ROD AND-RON AND, and the male speaker tended to pronounce these Rod 'n' or Ron 'n'. The nasalized sound in what remained of the second vowel caused subjects to identify the intended /d/ as /n/ in an unusually large proportion of the cases. Three of the difficult pairs for the LPC processor involved the sustention contrast, which is notoriously difficult to preserve with this processor [9].

The cues that are used for the consonant identification differ with the position of the consonant in the word. For example voice-onset-time is an important cue for the voiced-unvoiced contrast in the word initial position, and duration of the preceding vowel becomes an important cue in word medial and final positions. All of the fragments in this study were intervocalic consonants, but the fragments could cross word boundaries, and there were 14 consonants in word initial position, 18 medial, and 14 final. Table 4 shows scores as a function of position in the word. For comparison, DRT scores for the same voice systems are included as well.

Phonene Spoken

Phoneme Heard	m	n	ŋ	b	d	g	p	t	k	v	ð	z	ʒ	f	θ	s	ʃ	h	dʒ	tʃ	w	r	j	l
m		2	0	1																				
n	2		0		14																			
ŋ	0	5				1																		
b	0				1	0	0			0														
d		6		0		0		6			0	0							0					
g			2	1	0				0										3					
p				3				1	4					0				0						
t					5		0		0						0	0				2				
k						0	1	0												0				
v				2							4	2		8										
ð					5					0		0			4									
z					1					4	0		5			1								
ʒ												8					3		6					
f							0			5					3	0		0						
θ								0			1			1		0								
s								0				0		0	0		1							
ʃ													6			8				0				
h							1							0										
dʒ					2	5							11							0				
tʃ								1	0								1		0					
w																						0	3	0
r																					1			5
j																					0			
l																					0	5		

Fig. 2 — Confusion matrix for unprocessed VCV pairs. Only the cells with entries are possible confusions.

Phoneme Spoken

Phoneme Heard	m	n	ŋ	b	d	g	p	t	k	v	ð	z	ʒ	f	θ	s	ʃ	h	dʒ	tʃ	w	r	j	l
m		1	0	0																				
n	1		1		19																			
ŋ	0	2				3																		
b	0				0	0	0			1														
d		5		0		0		7			0	0							0					
g			1	0	0				0										2					
p				0				0	2					3				0						
t					3		0		1						0	0				1				
k						0	0	15												0				
v				5							8	2		9										
ð					5					3		6			2									
z					0					1	4		1			1								
ʒ												3					2		4					
f							0			3					3	0		1						
θ								0			3			6		0								
s								1				1		5	1		0							
ʃ													6			14				0				
h							0							0										
dʒ					0	3							10							0				
tʃ								5	0								3		16					
w																						0	6	1
r																					2			7
j																					3			
l																					0	7		

Fig. 3 — Confusion matrix for VCV pairs processed through CVSD voice processor at 32 kbps. Only the cells with entries are possible confusions.

Phoneme Spoken

Phoneme Heard	m	n	ŋ	b	d	g	p	t	k	v	ð	z	ʒ	f	θ	s	ʃ	h	dʒ	tʃ	w	r	j	l
m		8	0	0																				
n	3		3		14																			
ŋ	0	5				7																		
b	0				0	0	1			2														
d		6		0		0		6			0	0							0					
g			1	2	0				1										3					
p				1				1	5					0				0						
t					6		3		1						1	1				9				
k						4	4	20												0				
v				4							8	7		15										
ð					7					0		9			3									
z					2					6	3		6			1								
ʒ												8					2		5					
f							0			4					2	0		0						
θ								0			1			2		1								
s								0				0		2	0		1							
ʃ													6			11				0				
h							0							1										
dʒ					2	4							9							1				
tʃ								1	0								0		12					
w																						0	5	2
r																					2			9
j																					1			
l																					2	2		

Fig. 4 — Confusion matrix for VCV pairs processed through CVSD voice processor at 16 kbps. Only the cells with entries are possible confusions.

Phoneme Spoken

Phoneme Heard	m	n	ŋ	b	d	g	p	t	k	v	ð	z	ʒ	f	θ	s	ʃ	h	dʒ	tʃ	w	r	j	l
m		16	0	0																				
n	7		1		9																			
ŋ	1	0				2																		
b	0				2	1	0			3														
d		7		2		1		5			2	0							1					
g			1	0	2				0										2					
p				3				2	11					2				0						
t					6		0		0						0	1				0				
k						1	8	19												0				
v				23							25	7		13										
ð					9					4		9			2									
z					1					8	7		9			0								
ʒ												2					1		5					
f							0			7					2	0		0						
θ								2			2			7		2								
s								0				4		7	1		3							
ʃ													20			21				0				
h							0							0										
dʒ					0	0							10							0				
tʃ								6	0								0		14					
w																						0	7	17
r																					2			13
j																					9			
l																					2	3		

Fig. 5 — Confusion matrix for VCV pairs processed through APC voice processor at 9.6 kbps. Only the cells with entries are possible confusions.

Phoneme Spoken (columns) / Phoneme Heard (rows)

	m	n	ŋ	b	d	g	p	t	k	v	ð	z	ʒ	f	θ	s	ʃ	h	dʒ	tʃ	w	r	j	l
m		19	0	2																				
n	6		1		14																			
ŋ	0	5				6																		
b	2				13	1	1			4														
d		27		2		6		11			1	4							1					
g			7	12	26				3										5					
p				10				8	13					3				5						
t					19		3		1						0	0				2				
k						2	2	30												0				
v				29							13	6		8										
ð					24					1		4			14									
z					7					1	5		5			9								
ʒ												4					3		13					
f							6			5					24	1		5						
θ								4			10			5		2								
s								2				1		9	16		2							
ʃ													17			29				4				
h							9							1										
dʒ					28	8							36							3				
tʃ								32	24								3		21					
w																						0	6	13
r																					4			12
j																					12			
l																					12	6		

Fig. 6 — Confusion matrix for VCV pairs processed through LPC voice processor at 2.4 kbps. Only the cells with entries are possible confusions.

Table 3 — The Ten Pairs with the Greatest Number of Errors for Each Voice System (Percent correct is given in parentheses)

Unprocessed	Voice System			
	CVSD 32	CVSD 16	APC 9.6	LPC 2.4
n-d (60)	n-d (52)	k-t (58)	v-ð (42)	dʒ-ʒ (2)
dʒ-ʒ (66)	k-t (68)	n-d (60)	v-b (48)	n-d (18)
v-f (74)	dʒ-tʃ (68)	v-f (62)	s-ʃ (52)	t-tʃ (32)
z-ʒ (74)	s-ʃ (72)	z-ʒ (72)	m-n (54)	v-b (34)
d-t (78)	r-l (72)	dʒ-tʃ (74)	ʒ-ʃ (58)	g-d (36)
r-l (80)	dʒ-ʒ (72)	v-z (74)	v-f (60)	k-t (38)
s-ʃ (82)	v-f (76)	s-ʃ (76)	w-l (62)	s-ʃ (38)
dʒ-tʃ (82)	v-ð (78)	z-ð (76)	k-t (62)	t-d (40)
ʒ-ʃ (82)	z-ð (78)	r-l (78)	k-p (62)	dʒ-d (42)
g-dʒ (84)	d-t (80)	dʒ-ʒ (78)	w-j (68)	f-θ (42)

Table 4 — Percent Correct as a Function of Position of the Consonant in the Word (Experiment 1)

Consonant Position	Voice System				
	Unprocessed Speech	CVSD 32 kbps	CVSD 16 kbps	APC 9.6 kpbs	LPC 2.4 kbps
Initial	97.7	93.3	92.8	86.5	68.3
Medial	92.8	91.6	88.3	81.3	64.1
Final	85.4	84.4	81.4	80.4	65.7
VCV total score	92.1	90.0	87.5	82.4	65.4
DRT score (Initial)	97.6	95.2	92.3	91.0	87.4

For the unprocessed speech, word-initial VCV phonemes were recognized as well as DRT words. However initial VCV scores decreased more for the digital voice processors than did DRT scores. Word medial and final VCV scores were lower than word initial scores for all systems except the LPC processor. The difference between scores for unprocessed and processed speech was actually smaller for medial and final position than for initial. It seems that while medial and final position phonemes were originally harder to discriminate, they did not lose as much when processed through the digital voice processors. Word medial and final consonants may be less distinctively articulated than initial consonants, and in listening to normal conversational speech, the listener uses contextual cues from the sentence and the rest of the word as well as expectations based on knowledge of the world to recognize these sounds. When the phonemes are taken out of context, they are harder to identify. On the other hand the vowel preceding the consonant carries more information about consonant identity for medial and final consonants, and it seems that the coarticulatory and durational information carried by the vowel is useful in preserving consonant identifiability under the degradations caused by digital analysis and resynthesis processing.

EXPERIMENT 2

Inexperienced listeners tend to score lower on intelligibility tests than do practiced listeners whose scores have stabilized. For a more direct comparison of VCV scores with the DRT, the VCV tapes were scored by the trained listening crews of Dynastat, Inc. Dynastat maintains screened and trained crews of listeners and conducts DRT tests and other voice quality tests for a wide variety of customers. Copies of the tapes that had been tested with the naive listeners as well as sample answer forms were sent to Dynastat. They trained their experienced listeners on the VCV test format and then had them score the processed and unprocessed VCV tapes. Subscores for the feature contrasts used on the DRT were computed for the comparable VCV feature contrasts. Table 1 shows that number of pairs and the phonemes contrasted were not identical to those used in the DRT. Some of the VCV comparisons cannot occur in word initial position, and others were of special interest for this exploratory study.

A comparison by distinctive features of VCV scores from the present experiment with DRT scores obtained by the DoD Digital Voice Processor Consortium is shown in Table 5. The top half of the table gives a direct comparison of the scores, and the bottom half shows the difference between unprocessed and processed speech and is an indication of which features are the most vulnerable to loss in intelligibility under the various forms of digital processing. VCV scores and DRT scores differ

Table 5 — Comparison of VCV and DRT Feature Scores for
Four Digital Voice Processors (Experiment 2)

	Voice System									
	Unprocessed		CVSD 32 kbps		CVSD 16 kbps		APC 9.6 kbps		LPC 2.4 kbps	
	VCV	DRT	VCV	DRT	VCV	DRT	VCV	DRT	VCV	DRT
Feature scores										
Voicing	82.4	97.5	77.7	91.3	78.7	91.1	78.4	92.3	72.5	89.7
Nasality	84.8	99.3	83.9	98.7	83.3	98.5	78.9	97.7	70.7	94.6
Sustention	91.5	97.6	90.6	89.4	89.8	82.6	83.9	85.1	78.2	80.1
Sibilation	96.0	98.5	96.7	93.0	95.6	85.2	97.3	93.7	81.9	89.8
Graveness	93.6	91.9	92.7	85.7	90.8	80.6	84.4	82.9	74.0	78.0
Compactness	93.4	99.2	91.4	99.1	89.7	96.8	86.0	96.4	78.6	90.9
Difference (Unprocessed minus processed)										
Voicing			4.7	6.2	3.7	6.4	4.1	5.2	10.0	7.8
Nasality			0.9	0.6	1.5	0.8	5.9	1.6	14.1	4.7
Sustention			0.9	8.2	1.8	15.0	7.6	12.5	13.3	17.5
Sibilation			−0.7	5.5	0.4	13.3	−1.3	4.8	14.1	8.7
Graveness			0.9	6.2	2.7	11.3	9.2	9.0	19.6	13.9
Compactness			2.0	0.1	3.7	2.4	7.5	2.8	14.9	8.3

markedly both in individual feature scores and in which features show the greatest loss for the different voice processors. Within each test, the two CVSD processors show quite similar losses, which suggests that the tests themselves are fairly stable.

Even though the voicing feature was relatively weak intervocalically in unprocessed speech, it was relatively robust under LPC processing. This is probably because the duration of the preceding vowel is one of the cues to voicing in word medial and word final position, and vowel duration information was retained in the VCV excerpts. Therefore even when information about voice onset time was degraded, the presence of vowel duration information permitted a higher rate of correct identifications.

Sustention, which is one of the weakest DRT features under LPC processing, also fared somewhat better intervocalically, which may indicate that this problem is not as serious in conversational speech as it is with isolated words. On the other hand, the "place" features—graveness and compactness—suffered the most intervocalically under LPC processing. The information for place of articulation is carried primarily by the formant transitions, and this information tends to be less distinct in continuous speech than in isolated words. Although performance was good with unprocessed speech, the effects of LPC processing (where information is averaged over a 22.5 ms frame) seem to be particularly damaging to this kind of information.

Since each of the feature scores is based on a subset of phoneme pairs, they are not as stable as total DRT or VCV scores, and occasional reversals may occur as a result of normal variability. Thus the sibilation feature on the VCV shows essentially no loss for any but the LPC processor, and two processor scores were insignificantly higher on this feature than the unprocessed speech. Likewise, on the DRT, the 9.6 kbps processor scored higher than the 16 kbps processor on four of the six features.

Standard errors for VCV scores (usually in the range of 2 to 4 points) were on the whole slightly larger than comparable standard errors for DRT scores (in the range of 1 to 3 points). This could have resulted from any of a number of factors: The test crews were not as experienced with the VCV, there were only two speakers instead of the customary six or more for the DRT, VCV scores were lower than comparable DRT scores and low scores on the DRT have larger standard errors than high scores, the use of fragments instead of whole words could have been confusing and caused some erratic responding. Even though the standard errors on this preliminary version were slightly larger than might be desirable, there was a very high correlation for feature scores on retests of the same processors as shown in Table 6. This table also shows retest correlations for DRT scores where they could be obtained as well as DRT-VCV correlations. The correlation data suggest that while each test provides reasonably stable feature scores, the two tests are measuring different aspects of intelligibility loss due to digital voice processing. It should be possible to reduce the standard error of the VCV with further test development.

Table 6 — Correlations (Pearson's r)
Between Feature Scores
for Selected Test Conditions

Processor Tests	r
LPC-2.4 kbps	
VCV-VCV	0.854*
DRT-DRT	0.936*
VCV-DRT	−0.223
APC-9.6 kbps	
VCV-VCV	0.976*
DRT-DRT	0.954*
VCV-DRT	−0.144
CVSD-16 kbps	
VCV-VCV	0.963*
VCV-DRT	−0.627
CVSD-32 kbps	
VCV-VCV	0.967*
VCV-DRT	−0.156
CVSD 16-CVSD 32	
VCV	0.996*
DRT	0.924*

*Statistically significant at $p \leq 0.05$.

Since there was only one VCV pair per talker for each phoneme contrast, it is not possible at this stage to determine the extent to which these differences are caused by word position effects or by different speech contexts—isolated words vs running speech. Further research with multiple tokens of the various contrasts in different word positions will be needed to clarify this issue. At this stage, it can be said that if feature scores are to be used to evaluate processor weaknesses, it is important to use samples from continuous speech in addition to standard DRT scores to obtain a balanced diagnostic evaluation.

EXPERIMENT 3

Several different randomizations of unprocessed tapes were evaluated by the same Dynastat crew that evaluated the tapes for the second experiment. The tapes were scored under seven noise conditions and nine conditions of low-pass filtering. Table 7 gives the signal-to-noise ratios and the filter cutoff values and the total VCV score for each of the conditions. The scores fall off as would be expected under these conditions, although they drop more than comparable DRT scores.

Table 7 — Effect of Noise and Lowpass Filtering on VCV Scores

Condition	Intervocalic Score
−12 dB S/N	22.8
−6 dB S/N	43.6
0 dB S/N	64.6
+6 dB S/N	77.4
+12 dB S/N	88.3
+18 dB S/N	93.3
+24 dB S/N	94.2
LP 200 Hz	14.7
LP 464 Hz	50.7
LP 728 Hz	61.3
LP 992 Hz	65.5
LP 1290 Hz	74.6
LP 1650 Hz	86.1
LP 2090 Hz	86.8
LP 2620 Hz	91.6
LP 3250 Hz	93.1

Individual feature scores are plotted in Fig. 7 for the low-pass filtered conditions and in Fig. 8 for the noise conditions. The effect of noise on DRT features is given in Ref. 2, and Miller and Nicely [10] tested consonant confusions under a variety of noise and filtering conditions. The DRT and VCV are alike in test methodology and differ in speech materials. The Miller and Nicely data are based on phonemes in syllable initial position spoken in isolation, but the test methodology is quite different from the DRT. Randomized lists of 16 consonants (all followed by the vowel /a/) were read by the talkers, and the listeners recorded their responses from the entire range of possible alternatives. For comparison with DRT and VCV data we used the Miller-Nicely confusion matrices to calculate scores that would be comparable to the DRT features. This was done by counting as errors for each feature all of those responses which differed from the spoken phoneme on the feature in question. Thus for the voicing feature, if /b/ were spoken, all responses that were unvoiced phonemes would be errors and all that were voiced would be correct. For the sustention feature for the same stimulus, stops would be classified as correct and continuants as errors. The derived feature scores are plotted in Figs. 9 and 10 for low-pass filtering and noise conditions. In comparing the filtering data, note that there is no comparable condition in the Miller-Nicely data to the lowest cutoffs in the VCV data. At high cutoff frequencies voicing and nasality fared relatively poorly while the remaining features all had higher

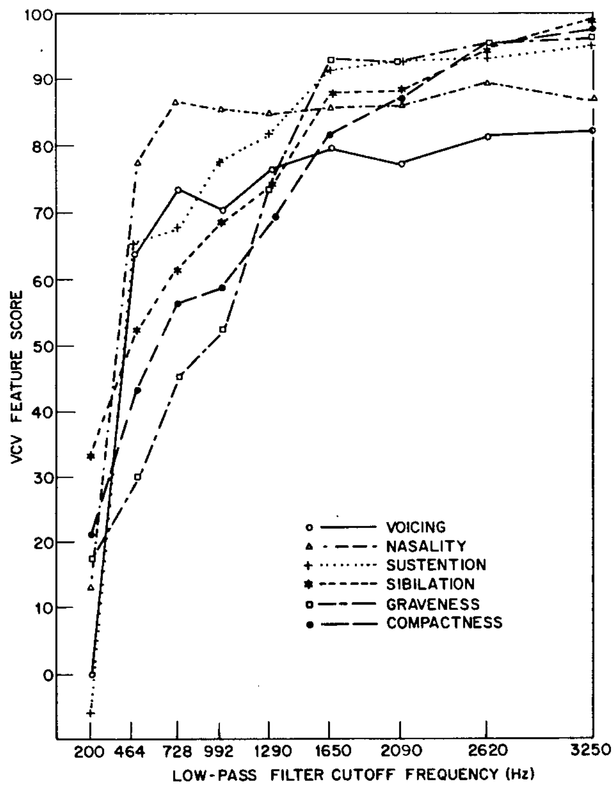

Fig. 7 — The effect of low-pass filtering on feature scores

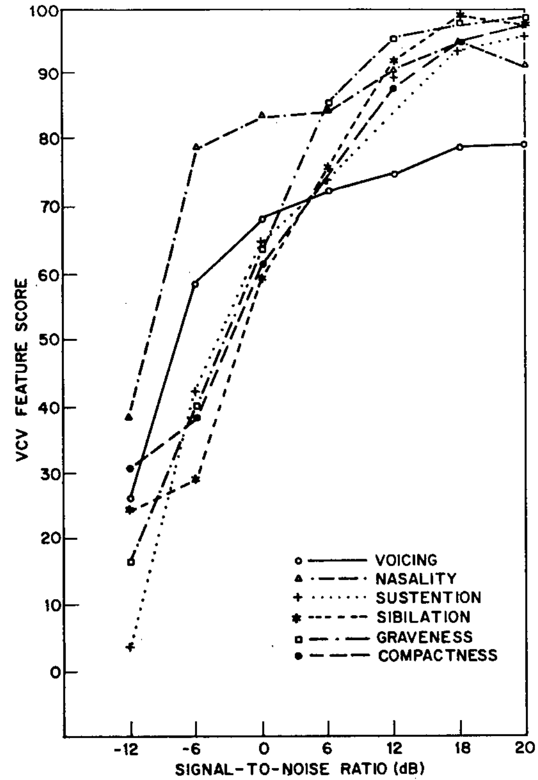

Fig. 8 — The effect of noise on feature scores

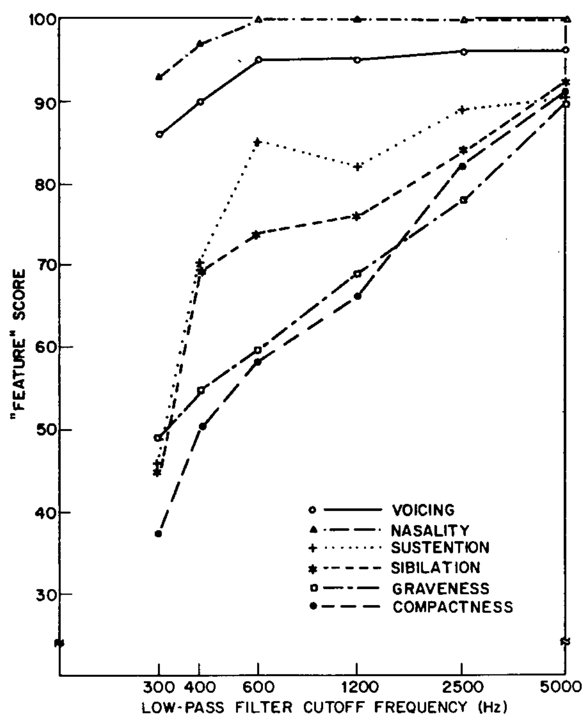

Fig. 9 — Derived feature scores for initial consonants based on Miller and Nicely's 1967 confusion data for filtering conditions

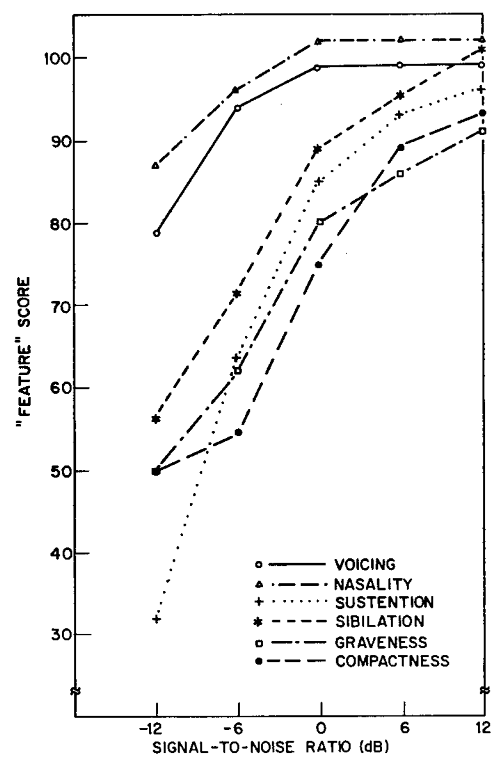

Fig. 10 — Derived feature scores for initial consonants based on Miller and Nicely's 1967 confusion data for noise conditions

scores for the VCV data, and the reverse was true for the Miller-Nicely data. Down to cutoff frequencies of about 400 Hz, the two data sets showed very similar types of losses in that voicing and nasality scores remained relatively near their original level and the place features—graveness and compactness—showed steep losses with lower cutoff frequencies.

The same pattern of both similarities and differences can be seen also in the noise data. The DRT data given by Voiers [2] are very similar to the Miller-Nicely data shown in Fig. 10. With the exception of the compactness feature which had somewhat higher scores on the DRT, the other five features were ranked the same for DRT and Miller-Nicely data at +12 dB and −12 dB, and the pattern of losses was very similar. Nasality and voicing were the most robust features in noise, and graveness and sustension showed the greatest losses. Again the VCV data showed both similarities and differences: nasality was the most robust feature and sustention the weakest, but the remaining features differed somewhat from the other two sets of data. There seem on the whole to be more differences between intervocalic and initial consonants than between the two sets of initial consonant data even with very different testing methods. These results should be viewed as tentative since they may depend very strongly on the particular speech samples used in this study. A broader study with more talkers and a variety of word contrasts for each phoneme pair would be needed for more definite conclusions.

Figures 11 and 12 show the effects of noise and filtering on different word positions. These results are similar to the results for the digital voice systems in that initial consonants scored higher than medial and final consonants in most of the conditions.

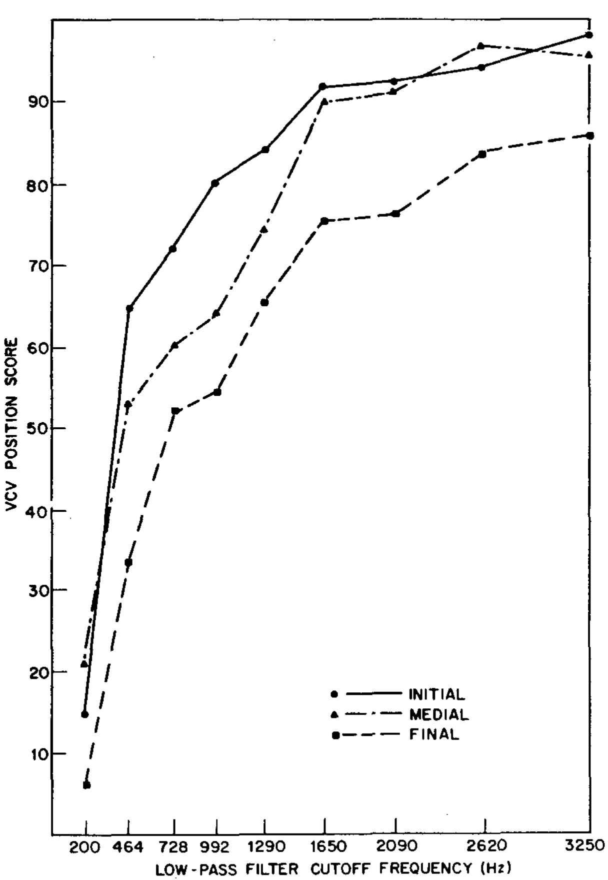

Fig. 11 — Word position scores as a function of filter cutoff frequency

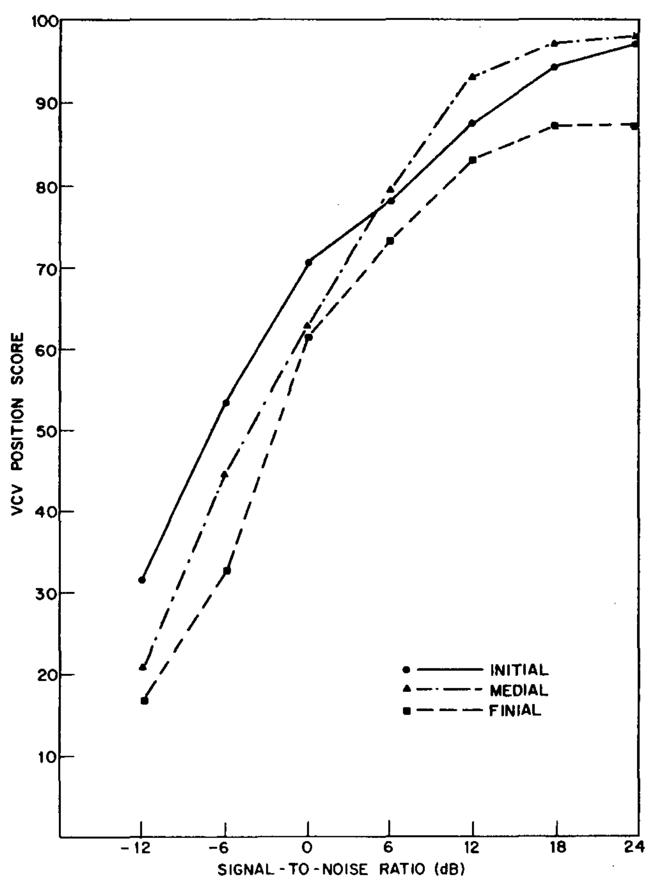

Fig. 12 — Word position scores as a function of signal-to-noise ratio

CONCLUSIONS AND RECOMMENDATIONS

The results of preliminary research indicate that the intelligibility of vowel-consonant-vowel (VCV) segments excised from running speech differs in various ways from scores on the Diagnostic Rhyme Test (DRT), which uses syllable-initial consonants spoken in isolation. The feature scores that show the greatest losses in intelligibility when the speech is processed through digital voice processors are not the same for the two tests. This is because different voice cues are important in connected speech than in carefully pronounced isolated words. There were also intelligibility differences depending on whether the consonant came from the beginning, middle, or end of a word. There were not enough different examples of phoneme contrasts in the preliminary tests to compare feature scores for different word positions. The evaluation of individual feature contrasts when they occur in different word positions could be highly informative for developing improved digital voice processing techniques.

In the long run, each of the six distinctive features represented on the DRT should be investigated using the VCV technique. The first step would be to select one or two features that are especially interesting, for example, those that show the greatest intelligibility losses for the standard DoD LPC 2.4 kbps processor. A set of sentences and paragraphs containing appropriate phoneme contrasts in various word positions could be developed, and these would be read by several different male and female talkers. The use of more word pairs and a larger sample of talkers will help ensure that the results are not specific to the way one person articulates a particular word. There are several ways of increasing the number of phoneme pairs, and this would be a fruitful area for further research. The contrasts should probably include all of the word positions that are possible for a given phoneme pair, i.e., /ŋ/ and /ʒ/ do not occur in word initial position in English and /h/ does not occur in word final position. The test could also be improved by excising longer segments so that the context for the phoneme contrast would be whole words instead of fragments. This would have the additional advantage of making the answer sheet easier to use.

A detailed series of experiments on the effect of word position on intelligibility and intelligibility loss under various forms of voice degradation would be a very promising research area regardless of whether an intervocalic consonant intelligibility test is actually developed. An analysis of the cues at different word positions in continuous speech that are most susceptible to loss for the DoD standard LPC algorithm or for other digital voice algorithms could lead to new ways of improving these algorithms.

Intervocalic consonants also showed some similarities and some differences when compared with isolated syllable-initial consonants under conditions of noise and low-pass filtering. With only two speakers, some of the specific effects may have been due to idiosyncratic aspects of the way a particular word was spoken by one of the speakers, but the consistent results across the three experiments suggests that there are many real differences between the two types of stimuli. A test using speech stimuli from connected speech would supplement the diagnostic information available from the DRT. The DRT also tests only distinctions that occur at the onset of speech, and the performance of many digital voice algorithms is more stable after the first few samples, so a test in which the discrimination is to be made further into the speech stream would be more indicative of actual performance. DRT scores are at present being written into performance specifications in a number of government contracts. This creates the possibility of intentionally or unintentionally "tuning" a system to obtain the highest possible DRT score at the possible expense of real losses in other aspects of performance. Although it is never possible to guard entirely against this possibility, the existence of a second intelligibility test that is based on cues in other parts of the voice signal would help researchers in deciding more realistically whether there is any actual improvement in performance.

This research has shown that the development of an intervocalic consonant test is feasible and could be informative about the performance of digital voice systems in ways that supplement the DRT tests presently used.

REFERENCES

1. A.S. House, C.E. Williams, M.H.L. Hecker, and K.D. Kryter, "Articulation-Testing Methods: Consonantal Differentiation with a Closed-Response Set," J. Acoust. Soc. Am. **37**, 158-166, Jan. 1965.

2. W.D. Voiers, "Diagnostic Evaluation of Speech Intelligibility," in *Speech Intelligibility and Recognition,* M.E. Hawley, ed., Stroudsburg, Pa: Dowden, Hutchinson and Ross, 1977.

3. J.P. Egan, "Articulation Testing Methods," *Laryngoscope,* **58**, 955-991 (1948).

4. T. Tremain, "The Government Standard Linear Predictive Coding Algorithm: LPC-10," *Speech Technology,* **1**(2), 40-49, Apr. 1982.

5. W.D. Voiers, Uses, Limitations, and Interrelations of Present-Day Intelligibility Tests, presented at the National Electronics Conference, Chicago, Oct. 1981.

6. L.C.W. Pols, Intelligibility of Intervocalic Consonants in Noise, in *Proc. IEEE Conf. on Acoustics Speech and Signal Processing,* Apr. 1979.

7. R. Jakobson, C.G.M. Fant, and M. Halle, *Preliminaries to Speech Analysis the Distinctive Features and their Correlates,* Cambridge, MA: MIT Press, 1963.

8. G.F. Sandy (ed.), Digital Voice Processor Consortium Interim Report—Appendix A MTR-81W00159-02, Mitre Corp. Jan. 1982.

9. G. Kang, personal communication, July 1982.

10. G.A. Miller and P.E. Nicely, "An Analysis of Perceptual Confusions Among Some English Consonants," J. Acoust. Soc. Am. **27**, 338-352, 1955.

Appendix
SENTENCES READ BY THE TWO TALKERS FOR THE VCV TEST MATERIALS

LIST A

Have you heard about all the accidents they've been having at the tanning factory out by Perryville Station, in the next text town? The first one was when someone went out to the shed where they stow all their extra supplies, and when he opened the door, a whole box of hammers fell on top of him. One time a relief unit had a slug in it so that the contents of the soaking vat were leaching onto the floor. One worker lost his balance and fell against the presser element and was severely burned. The man who was moving the vat covers says they are too weak, and the baskets should also be replaced. In addition, the input level is adjusted wrong, and some people are not watching their hands properly when they work with strong chemicals. There was an interview in today's paper with Mr. Freddy Wu, the company president. When he was questioned about the accidents, he reported that he had already sent out the safety addenda. He added "It's my belief I have done everything I can at this time." The company boasted a profit gross of 3 billion dollars last year, but they didn't spend any of it on safety. When they asked Mr. Wu about this, he said "We thought about it for a long time last year before making our decision."

The young pearl divers found that the treasure at the bottom of the ocean was greater than they had expected.

Did I tell you Holly is majoring in biology and wants to study seals and their breathing patterns?

They have been raising horses for many years now, and this is the first time they every had a lazy groom in their stables.

Betty and Jimmy are helping each other with their math, but the log intervals on the graph are still confusing.

Will you hold the seats for me while I go get some popcorn?

Ruby usually wore very little make-up and the bruise on here face was very obvious.

Bobby is doing very well in Spelling class. Today he got everything right except carnival and either.

This article on England in the Middle Ages is very interesting, but I didn't think they drank as much mead as it says they did.

Have you seen my new Super-8 Instant Movie camera? I got it for my birthday.

The Temple of Poseidon stood until erosion undermined the foundations so much that it fell into the sea.

I forgot to turn off the alarm clock when I left yesterday, and it rang all day long.

Hey Dale, Mitch and Rod and I are going to Anderson's closing sale. Would you like to come with us?

Where I take my boat this fall will depend on what happens with the cold weather.

My Grandfather has to have weekly gold treatments for his arthritis.

LIST B

Have you heard about all the accidents they've been having at the canning factory out by Ferryville Station, in the next text town? The first one was when someone went out to the shed where they store all their extra supplies, and when he opened the door, a whole box of hangers fell on top of him. One time a release unit had a sludge in it so that the contents of the soaping vat were leaking onto the floor. One worker lost his balance and fell against the pressure element and was severely burned. The man who was smoothing the vat covers says there are two leaks, and the gaskets should also be replaced. In addition, the input lever is adjusted wrong, and some people are not washing their hands properly when they work with strong chemicals. There was an interview in today's paper with Mr. Freddy Yu, the company president. When he was questioned about the accidents, he retorted that he had already sent out the safety agenda. He added "Now I believe I have done everything I can at this time." The company posted a profit growth of three million dollars last year, but they didn't spend any of it on safety. When they asked Mr. Yu about this, he said "We fought about it for a long time last year before making our decision."

The young pearl divers found that the pressure at the bottom of the ocean was greater than they had expected.

Did I tell you Polly is majoring in biology and wants to study teals and their breeding patterns?

He has been racing horses for many years now, and this is the first time he has ever had a lady groom in his stables.

Bethie and Ginny are helping each other with their math, but the long intervals on the graph are still confusion.

Will you fold the sheets for me while I go check the dinner?

Trudy usually wore very little make-up and the rouge on her face was very obvious.

Timmy is doing very well in Spelling class. Today he got everything right except cannibal and ether.

This article on England in the Middle Ages is very interesting, but I didn't think they ate as much meat as it says they did.

Have you seen my new Super-H Instand Movie camera? I got it for my birthday.

The Temple of Poseidon stood until the Trojans attacked the city and burned it to the ground.

I forgot to turn off the heater when I left yesterday, and it ran all day long.

Hey Gail, Midge and Ron and I are going to Anderson's clothing sale. Would you like to come with us?

Where I cast my vote this fall will depend on what happens with the gold market.

My grandfather has to have weekly cold treatments for his bad back.