Title: **Intelligibility Testing - Diagnostic Rhyme Test (DRT) Report**

Authors: **Alan D. Sharpley and Ira L. Panzer (Dynastat Inc.)**

Date: **26-Mar-2003**

Version: **3.0**

Revision history:
Version 1.0: *Creation, March 20, 2003*
Version 2.0: Dynastat additions
Version 3.0: Dynastat corrections

1. Scope

This document presents the results of the Diagnostic Rhyme Tests (DRT) performed by Dynastat Inc. on the processed speech material provided by IBM and Motorola.

2. Diagnostic Rhyme Test

The DRT is an ANSI standard (S3.2-1989 available at www.ansi.org) for measuring the effects of communication systems on speech intelligibility. In one version or another, it has been in use for 35 years and has been extensively used by the communications engineering community for evaluating digital voice communication systems and techniques.

The DRT and its various derivatives are based on the principle that the intelligibility-relevant information in speech is carried by a small number of *distinctive features*, such that intelligibility depends most immediately on how well a communication link or device has preserved the acoustical correlates of these features. The DRT was designed, specifically, to measure how well information as to the states of six binary distinctive features, *Voicing, Nasality, Sustention, Sibilation, Graveness, and Compactness* (See Appendix A for additional information) have been preserved by the link. For this purpose, it uses a suite of 96 rhyming word pairs, in which the initial consonants of the two words of each pair differ only with respect to one of the six distinctive features. (See Appendix B for details) The listener's task with each item is to judge which of the two rhyming words (e.g., **Z**oo vs. **S**ue) has been spoken. Incorrect judgments indicate that the system has failed to preserve information contained in the feature involved. Like most other intelligibility tests in use today, the DRT and its derivatives test only for the discriminability of consonant phonemes, which carry the bulk of the useful information in speech and are generally more sensitive than vowels to speech degradation.

The DRT yields a total score, which, under properly controlled conditions, is highly correlated with scores yielded by all other intelligibility tests in use today. The DRT also yields a diversity of diagnostic scores that can be useful in pinpointing specific deficiencies or defects in the system or device under test.

Minimal listener training is required by the DRT. With a carefully-selected and monitored panel of eight listeners, the DRT has extremely high resolving power and test-retest reliability. It can resolve differences of less than 1 dB in speech-to-noise ratio.

Appendix C presents information on Dynastat's history of involvement with the development and implementation of speech intelligibility, speech quality, and speaker recognizability evaluation methods.

3. Test Methodology

Materials

Dynastat provided six talkers (three male, three female) DRT source material to IBM and Motorola for purposes of these evaluations. The DRT speech materials were processed for 34 conditions and recorded as 16kHz, 16 bit linear PCM files on CD-ROM.

Equipment

The processed speech materials were presented to listeners seated at separate, visually screened listening stations contained within a Tracoustics soundproof room with an ambient noise level less than 30dBA. Speech materials were presented over TDH-39 headphones (same signal both ears). Headphones were driven by a distribution amplifier set to deliver an overall sound level of the speech stimuli to the listeners at an active level of -15 dBPa (79 dB SPL). Headphone calibration was accomplished using a B&K 4153 Artificial Ear with circumaural headphone adapter, 4134 Microphone element, and 2609 Measurement Amplifier. The processed speech files were channeled through a Townshend Computer Tools DAT-Link+ and recorded on Digital Audio Tape (DAT) for presentation to the listening panels.

Listeners

DRT materials were presented to Dynastat's trained panel of listeners. (See Appendix D for additional information on the listener selection process) The DRT results presented in this report are based on a crew of eight listeners.

4. Results

Table I shows the mean DRT scores for each of the 34 conditions averaged across combined (6 talkers), three male, and three female talkers. In addition the standard error (across listeners) and the lower and upper 95% confidence intervals (CI) are presented. Table II shows the diagnostic scores for the six features measured by the DRT.

Table I. Summary DRT Scores

Condition	Total DRT Score	Standard Error	L-95% CI	U-95% CI	Total Male	Total Female
1A1	95.7	0.41	94.9	96.5	96.6	94.8
1A2	95.5	0.41	94.7	96.3	96.0	95.0
1A3	92.4	0.62	91.2	93.6	94.4	90.5
1A4	93.8	0.35	93.1	94.5	95.1	92.5
1B1	93.0	0.71	91.6	94.4	92.7	93.3
1B2	88.8	0.56	87.7	89.9	90.5	87.2
1B3	85.0	0.63	83.8	86.2	88.0	82.0
1B4	87.1	0.57	86.0	88.2	88.7	85.5
1C1	92.8	0.46	91.9	93.7	93.1	92.5
1C2	88.9	0.47	88.0	89.8	90.3	87.5
1C3	87.5	0.64	86.2	88.8	90.0	84.9
1C4	87.9	0.73	86.5	89.3	89.7	86.1
1D1	86.9	0.75	85.4	88.4	88.8	85.0
1D2	81.3	0.37	80.6	82.0	81.4	81.1
1D3	81.2	0.72	79.8	82.6	82.3	80.0
1D4	81.2	0.89	79.5	82.9	82.3	80.0
1E1	91.6	0.45	90.7	92.5	93.1	90.1
1E2	86.8	0.30	86.2	87.4	88.0	85.6
1E3	85.0	0.65	83.7	86.3	88.1	82.0
1E4	85.3	0.65	84.0	86.6	88.1	82.5
2A1	91.7	0.78	90.2	93.2	93.1	90.3
2A3	93.3	0.61	92.1	94.5	94.3	94.3
2B1	92.1	0.47	91.2	93.0	92.5	91.7
2B3	92.8	0.48	91.9	93.7	93.6	92.1
3A2	92.6	0.53	91.6	93.6	93.3	91.9
3A3	92.1	0.68	90.8	93.4	92.9	91.4
3A4	83.4	0.78	81.9	84.9	84.4	82.3
3B2	92.6	0.50	91.6	93.6	93.7	91.4
3B3	92.0	0.48	91.1	92.9	92.8	91.2
3B4	83.4	0.46	82.5	84.3	84.8	82.0
4A2	92.9	0.35	92.2	93.6	92.1	93.8
4A3	94.2	0.35	93.5	94.9	93.4	94.9
4B2	93.5	0.43	92.7	94.3	94.9	92.1
4B3	92.1	0.55	91.0	93.2	92.9	91.4

Table II. Summary DRT Feature Scores

Condition	Total Feature Score (3m, 3f)					
	Voicing	Nasality	Sustention	Sibilation	Graveness	Compact
1A1	95.6	98.6	97.3	93.4	90.9	98.7
1A2	95.7	99.1	94.7	94.0	90.9	98.7
1A3	94.7	97.3	87.6	88.5	88.4	98.2
1A4	94.8	97.8	92.2	89.8	90.0	98.4
1B1	95.7	98.6	89.6	90.4	85.8	97.9
1B2	88.4	97.7	77.2	87.5	85.2	97.0
1B3	84.1	95.6	73.0	85.2	79.2	93.0
1B4	88.8	96.2	72.5	88.2	83.6	93.4
1C1	93.1	98.3	91.9	90.2	87.6	95.4
1C2	85.4	95.2	84.8	89.7	81.8	96.6
1C3	85.5	96.0	78.0	86.5	85.4	93.5
1C4	86.2	96.9	77.9	89.5	83.5	93.5
1D1	89.8	94.0	81.9	88.7	73.0	93.9
1D2	81.6	86.7	72.0	90.2	68.6	88.4
1D3	82.6	91.0	67.8	82.4	74.3	88.9
1D4	81.6	91.5	68.2	86.7	71.0	88.0
1E1	93.5	97.1	91.4	89.1	82.4	96.0
1E2	83.2	96.2	76.0	90.5	82.8	92.2
1E3	85.5	95.2	72.8	85.9	77.2	93.5
1E4	87.4	96.1	72.7	85.4	78.0	92.1
2A1	93.0	98.0	87.9	89.3	84.9	96.9
2A3	94.3	98.8	92.2	91.0	85.8	97.8
2B1	92.7	98.3	93.0	89.2	84.1	95.3
2B3	94.1	97.9	91.4	89.5	86.2	97.9
3A2	94.4	98.8	89.1	89.6	86.8	97.0
3A3	94.3	96.5	90.9	87.8	88.3	95.2
3A4	87.1	89.7	77.6	84.0	72.7	89.2
3B2	94.5	98.6	91.5	87.9	85.2	97.7
3B3	93.4	98.6	89.1	87.5	85.8	97.7
3B4	87.6	92.7	74.9	79.9	74.3	91.0
4A2	93.9	96.5	90.5	96.7	83.2	96.6
4A3	95.3	97.9	93.6	97.7	84.5	96.0
4B2	94.5	98.7	91.7	90.8	87.6	97.7
4B3	92.3	98.6	89.2	93.2	84.1	95.4

References

[1] Voiers, W.D. (1977), "Diagnostic Evaluation of Speech Intelligibility." in Speech Intelligibility and Speaker Recognition, M.E. Hawley, ed., Dowden, Hutchinson & Ross, Stroudsburg, PA, 374-387.

[2] Voiers, W.D. (1983), "Evaluating Processed Speech Using the Diagnostic Rhyme Test." Speech Technology, 30-39.

[3] ANSI S3.2-1989, Method for Measuring the Intelligibility of Speech over Communications Systems, Standards Committee S3, Bioacoustics, Standards Secretariat, New York, NY.

APPENDIX A

INFORMATION ON DISTINCTIVE FEATURES

The articulatory bases of the six distinctive features are well understood. All voiced phonemes involve free vibration of the vocal cords; unvoiced phonemes do not. Nasals are produced by lowering of the velum, allowing air to escape through the nasal passages; nonnasals by closing the nasal passages. Sustained phonemes are produced by incomplete constriction of the vocal tract; interrupted phonemes by complete constriction of the tract at some point. Sibilants involve extreme constriction of the vocal tract that produces turbulence and high-frequency noise. Grave phonemes are produced by constriction toward the anterior of the vocal tract; acute by constriction in the middle of the tract. Compact phonemes are produced by constriction toward the rear of the vocal tract; diffuse phonemes by constriction near the middle.

Each of the six perceptual distinctive features has multiple acoustical correlates, where the relative saliency of each depends on phonemic environment and the states of one or more noncritical features. However, some generalizations are possible.

Voiced phonemes are distinguished from their unvoiced counterparts, or *cognates*, by the presence of periodicity, and, in particular, by the time of onset in periodicity. In voiced consonants, preceding vowels tend to be of greater duration than in the case of unvoiced consonants.

Nasal phonemes are distinguished by relatively pronounced resonances at *circa* 200, 800, and 2200 Hz, and by the presence of nulls throughout the frequency range.

Sustained phonemes are distinguished by their gradual onset and by the presence of mid-frequency noise; interrupted by their abrupt onset. Sustained phonemes have characteristic durational and high-frequency cues that distinguish them from their interrupted counterparts.

Sibilant consonants are characterized by higher-frequency noise and greater duration than their nonsibilant counterparts.

Grave phonemes are distinguished among other things by the origin and direction of second-formant transitions. Grave consonants always involve relatively steep upward transitions of the second formant. Acute consonants usually involve downward second-formant transitions, depending on vowel environment and the phoneme involved. In general, grave phonemes are characterized by greater concentration of low-frequency spectral energy than are acute phonemes.

Compact phonemes are characterized by the concentration of spectral energy in the mid-frequency range; diffuse phonemes by the distribution of energy over more-widely separated spectral peaks.

APPENDIX B

DRT-IV: ITEMS BY GROUPED BY CRITICAL DISTINCTIVE FEATURE

VOICING (Vc)

F	VQ	NR	PRESENT	ABSENT	MODE	NR	PRESENT	ABSENT	MODE
Vc	HF	15	DINT	TINT	N	71	GIN	CHIN	F
Vc	HF	43	BEAN	PEEN	N	99	VEAL	FEEL	F
Vc	HB	8	ZED	SAID	F	64	DENSE	TENSE	N
Vc	HB	36	VAST	FAST	F	92	GAFF	CALF	N
Vc	LF	22	VOLE	FOAL	F	78	GOAT	COAT	N
Vc	LF	50	ZOO	SUE	F	106	DUNE	TUNE	N
Vc	LB	1	DAUNT	TAUNT	N	57	VAULT	FAULT	F
Vc	LB	29	BOND	POND	N	85	JOCK	CHOCK	F

NASALITY (Ns)

F	VQ	NR	PRESENT	ABSENT	MODE	NR	PRESENT	ABSENT	MODE
Ns	HF	23	NIP	DIP	A	79	MITT	BIT	G
Ns	HF	51	NEED	DEED	A	107	MEAT	BEAT	G
Ns	HB	16	NECK	DECK	A	72	MEND	BEND	G
Ns	HB	44	MAD	BAD	G	100	NAB	DAB	A
Ns	LF	2	MOOT	BOOT	G	58	NEWS	DUES	A
Ns	LF	30	MOAN	BONE	G	86	NOTE	DOTE	A
Ns	LB	9	GNAW	DAW	A	65	MOSS	BOSS	G
Ns	LB	37	KNOCK	DOCK	A	93	MOM	BOMB	G

SUSTENTION (St)

F	VQ	NR	PRESENT	ABSENT	MODE	NR	PRESENT	ABSENT	MODE
St	HF	3	SHEET	CHEAT	U	59	VEE	BEE	V
St	HF	31	VILL	BILL	V	87	THICK	TICK	U
St	HB	24	FENCE	PENCE	U	80	THEN	DEN	V
St	HB	52	THAN	DAN	V	108	SHAD	CHAD	U
St	LF	10	SHOES	CHOOSE	U	66	FOO	POOH	U
St	LF	38	THOSE	DOZE	V	94	THOUGH	DOUGH	V
St	LB	17	THONG	TONG	U	73	SHAW	CHAW	U
St	LB	45	VOX	BOX	V	101	VON	BON	V

SIBILATION (Sb)

F	VQ	NR	PRESENT	ABSENT	MODE	NR	PRESENT	ABSENT	MODE
Sb	HF	11	CHEEP	KEEP	U	67	ZEE	THEE	V
Sb	HF	39	SING	THING	U	95	JILT	GUILT	V
Sb	HB	4	JAB	GAB	V	60	SANK	THANK	U
Sb	HB	32	JEST	GUEST	V	88	CHAIR	CARE	U
Sb	LF	18	CHEW	COO	U	74	JUICE	GOOSE	V
Sb	LF	46	JOE	GO	V	102	SOLE	THOLE	U
Sb	LB	25	SAW	THAW	U	81	JAWS	GAUZE	V
Sb	LB	53	CHOP	COP	U	109	JOT	GOT	V

GRAVENESS (Gv)

F	VQ	NR	PRESENT	ABSENT	MODE	NR	PRESENT	ABSENT	MODE
Gv	HF	19	WEED	REED	VL	75	PEAK	TEAK	UP
Gv	HF	47	BID	DID	VP	103	FIN	THIN	UL
Gv	HB	12	BANK	DANK	VP	68	FAD	THAD	UL
Gv	HB	40	MET	NET	VL	96	PENT	TENT	UP
Gv	LF	26	POOL	TOOL	UP	82	MOON	NOON	VL
Gv	LF	54	FORE	THOR	UL	110	BOWL	DOLE	VP
Gv	LB	5	POT	TOT	UP	61	WAD	ROD	VL
Gv	LB	33	FOUGHT	THOUGHT	UL	89	BONG	DONG	VP

COMPACTNESS (Cm)

F	VQ	NR	PRESENT	ABSENT	MODE	NR	PRESENT	ABSENT	MODE
Cm	HF	27	YIELD	WIELD	VSa	83	KEY	TEA	UIg
Cm	HF	55	HIT	FIT	USa	111	GILL	DILL	VIg
Cm	HB	20	SHAG	SAG	USg	76	GAT	BAT	VIa
Cm	HB	48	YEN	WREN	VSg	104	KEG	PEG	UIa
Cm	LF	6	GHOST	BOAST	VIa	62	SHOW	SO	USg
Cm	LF	34	COOP	POOP	UIa	90	YOU	RUE	VSg
Cm	LB	13	GOT	DOT	VIg	69	HOP	FOP	USa
Cm	LB	41	CAUGHT	TAUGHT	UIg	97	YAWL	WALL	VSa

EXPERIMENTAL ITEMS (Ex)

F	VQ	NR			NR		
Ex	HF	7	NILL	RILL	63	NIP	RIP
Ex	HF	35	NEAP	REAP	91	NEATH	WREATH
Ex	HB	28	GNAT	RAT	84	NAP	RAP
Ex	HB	56	NEST	REST	112	NED	RED
Ex	LF	14	NOSE	ROSE	70	NODE	RODE
Ex	LF	42	NUDE	RUDE	98	NOOSE	ROOSE
Ex	LB	21	KNOB	ROB	77	NOT	ROT
Ex	LB	49	NOUGHT	ROUGHT	105	GNAW	RAW

NOTES:

F : feature
VQ: vowel quadrant
- HF high front
- HB high back
- LF low front
- LB low back

NR: item number

"Mode" refers to the state(s) of "noncritical" features in both Critical Phonemes (CP's). V=voiced, U=unvoiced; F=frictional, N=nonfrictional; S=sustained, P=plosive, L=non-plosive, I=interrupted; G=grave, A=acute; g=back/front, a=back/middle

APPENDIX C

BACKGROUND AND HISTORICAL HIGHLIGHTS

Dynastat, Inc., Austin TX, was formed in 1974 by Dr. William Voiers and his colleagues, Alan Sharpley and Ira Panzer. Its purpose was to provide the voice communications community with valid, reliable, practical, and economical means of predicting the effects of voice communication systems and devices on speech intelligibility, speech acceptability ("quality"), speaker-recognizability and "communicability." To this end it has developed the capability for providing test and evaluation services with all the recognized tests of speech intelligibility, speech acceptability and communicability in use today, including several developed by Dynastat. Dynastat has a continuing program of research to develop new methods of evaluation and improve existing methods. Some of the more important highlights of Dynastat's history of accomplishment in the voice communications field are shown below.

1965 The first version of Dynastat's flagship test, the *Diagnostic Rhyme Test* (DRT), was developed by Dr. Voiers, at the Sperry-Rand Research Center in Sudbury, MA. The DRT is a context-free intelligibility test, sensitive only to the properties of the link or device being tested. It evaluates the effect of a speech communication device or link on speech intelligibility by measuring the fidelity with which the acoustical correlates of six, critical, information-bearing *distinctive features* are preserved.

1969 Improved version of the DRT (DRT-IV) was developed by Dr. Voiers and his colleague, Alan Sharpley, at TRACOR, Inc., Austin, TX.

1974 **Dynastat Inc. is formed.**

1975 The *Paired Acceptability-Rating Method* (PARM) was developed for the Defense Communications Agency. PARM was designed as an interim method for evaluating the acceptability of state-of-the-art digital voice coders.

1976 A representative sample of state-of-the-art coding algorithms was evaluated from the standpoints of intelligibility and acceptability (speech "quality") for the Defense Communications Agency.

1977 The first version of the *Diagnostic Acceptability Measure* (DAM) was developed. The DAM yields multiple direct (isometric) and indirect (metametric and parametric) estimates of the effects of a link on overall speech acceptability. Scores for the detectability of various perceived qualities of the signal and the background provide valuable diagnostic information as well as supplementary estimates of overall acceptability.

1978	The *Diagnostic Communicability Test* (DCT) was developed under Navy sponsorship. The DCT was designed primarily to evaluate communications effectiveness in conferencing networks.
1979	Dynastat conducted a joint project with Georgia Institute of Technology to validate the DAM and to test various objective methods of predicting speech acceptability.
1981	Dynastat developed the first version of the *Diagnostic Alliteration Test* (DALT), a test for evaluating the preservation of distinctive feature information in final, postvocalic consonants.
1982	Dynastat developed the first version of the *Diagnostic Discrimination Test* (DDT), which was designed for, among other things, to distinguish between coding and decoding defects or deficiencies in voice coding systems.
1985	Development of capability for administering the *Spelling Alphabet Test* (SpAT). Using the spelling alphabet of the International Congress of Aviation Organizations, the SpAT, developed by Dr. Astrid Schmidt-Nielsen, of NRL, provides a realistic test of speech intelligibility in aviation and other situations in which use of the ICAO alphabet is prescribed.
1987	Comprehensive evaluation (DRT and DAM) of state-of-the-art 16 and 2.4KB voice coders for the DOD Digital Voice Processing Consortium.
1988	Comprehensive evaluation (DRT and DAM) of state-of-the-art 4.8K voice coders for the DOD Digital Voice Processing Consortium.
1988-89	Refinement and validation of a new version of the DAM (DAM-IIB) under DOD sponsorship.
1991-93	Further refinement of the DAM; development of the current version (DAM-IIC)
	Comprehensive evaluation (DRT and DAM) of state-of-the-art 600-1200 bps coders for the DOD Digital Voice Processing Consortium.
1993	Development of the *Diagnostic Medial-Consonant Test* (DMCT). Based on the same principles as the DRT and DALT, the DMCT measures the effects of a communication link on the preservation of distinctive feature information in medial, intervocalic consonants

Development of the *Modified Modified Rhyme Test* (MMRT). The MMRT uses the same input-speech materials as the MRT, but employs a special algorithm to select the listener's response options to each test word. This gives the MMRT a statistically-sounder basis than the MRT, and enables the experimenter to systematically vary the difficulty of individual items and the test as a whole.

Development begun on *Operational Communicability Test* (OCT) for predicting intelligibility and user-acceptance in specific communications environments.

1994 Development of a new version (DRT-V) of the DRT, which has additional scoring options and increased resolving power.

Development begun on a practical test of the effects of communication devices and links on speaker-recognizability.

1995 SBIR Phase I award titled " The Development of Auditory Icons for Representation of Virutal Objects in 3-D Space"

1996 Comprehensive evaluation of 2400 bps for new DoD standard for the Digital Voice Processor Consortium

SBIR Phase I award titled "The Development of a Method for Evaluating and Predicting Speaker Recognizability in Voice Communication Systems"

1997-2001 SBIR Phase II award to continue work on speaker recognizability

1999 Listening Laboratory for ITU 4K Qualification Phase

2000 Listening Laboratory for ITU 4K Selection Phase
Listening Laboratory for ITU 16K Qualification Phase
Listening Laboratory for ETSI-3GPP AMR-NB Characterization Phase
Listening Laboratory for ETSI-3GPP AMR-WB Selection Phase
Listening Laboratory for TIA-3GPP2 SMV Selection Phase

2001 Listening Laboratory for ITU 4K Selection Phase
Listening Laboratory for ITU 16K Selection Phase
Listening Laboratory for ETSI-3GPP AMR-WB Characterization Phase
Listening Laboratory for TIA-3GPP2 SMV Characterization Phase
Listening Laboratory for iBiquity Digital IBOC Radio Standardization

2002 Listening Laboratory for ITU 16K Characterization
Listening Laboratory for iBiquity Digital IBOC Radio Standardization

APPENDIX D

DYNASTAT'S LISTENING CREWS

BACKGROUND

The reliability and validity of results yielded by Diagnostic Rhyme Test (DRT), the Diagnostic Acceptability Measure (DAM), and other speech evaluation measures under development at Dynastat are large ensured by the intrinsic characteristics of the measures themselves. But, the level of precision ultimately achieved with these measures also depends on the manner in which they are used, most particularly on the manner in which listener factors are controlled.

Inter- and intraindividual variation in listener performance are potential sources of error in all speech evaluation measurements. Effective control of these factors is thus essential to precise measurement of speech intelligibility, overall acceptability, and speaker recognizability. Being fully cognizant of this principle, Dynastat employs extremely rigorous procedures for the selection, training, monitoring, and maintenance of its listening crews.

THE LISTENER SELECTION PROCESS

In view of Dynastat's substantial investment in each listening crew member, listener candidates are subjected to very careful scrutiny. A comprehensive biographical survey and personal interview are used to identify individuals whose personal characteristics and circumstances indicate them to be adaptable to the demands of the testing situation and likely to remain with Dynastat for an extended period. Individuals are further screened by means of a pure-tone audiometric screening test and various speech perception tests. Measures of the candidate's self consistency and conformity with established norms provide the criteria for crew membership on a trial basis. Approximately 30% of all bona fide applicants survive this level of screening.

THE LISTENER TRAINING PROCESS

Following admission to trial membership, listener candidates participate in both routine and special testing sessions for an extended period during which time their performance on the DRT, the DAM, and other measures are closely monitored. Either erratic performance or erratic attendance during this period may provide grounds for dismissal.

THE LISTENER MONITORING PROCESS

After admission to permanent status, new crew members participate as alternates in all research and service testing operations, but their data are normally analyzed only as part of the continuing monitoring process. Once stabilized and calibrated, they

participate as useful crew members in cases of absence of tardiness by senior crew members. As vacancies occur they ascend to the status of senior crew membership.

CREW MAINTENANCE

It has been Dynastat's experience that the listening crew member maintains the highest level and greatest stability of performance when he/she is a test participant on a regular basis with a carefully regulated work load. At Dynastat, all crew members normally participate in 3-4 hour testing sessions at least three times a week. Within the testing session there are regular rest breaks, such that the listener's "duty cycle" averages approximately 50%.

Listeners are not usually provided feedback regarding the accuracy of their responses to test materials. Although the DRT is uniquely immune to such effects, rewards for the correct responses can, with some speech discrimination tests, induce the listener to respond on the basis of factors other than his unbiased perception to the speech test materials (e.g., familiarity with the corpus of test materials or with specific randomizations of them). Where listener responses are influenced by such extra-signal or contextual factors, the test results necessarily provide a less valid reflection of the intrinsic characteristics of the system or device being evaluated.

CREW FACTORS AND EXPERIMENTAL DESIGN

Dynastat normally maintains two crews for purposes of routine testing and research. There are several advantages associated with this policy. In particular, the effects of absences by one or more crew members are minimized, since individuals of highly similar response characteristics re available as replacements. The availability of more than one crew also permits greater flexibility with regard to experimental design.