THE EFFECTS OF CONTROLLED SPEECH LEVEL INPUT ON THE INTELLIGIBILITY TESTING OF SPEECH COMPRESSION ALGORITHMS

James T. Sims and John D. Tardelli

ARCON Corporation
Waltham, MA

ABSTRACT

Speech level measurements were made for a collection of Diagnostic Rhyme Test (DRT) word lists as used for intelligibility testing of voice communications systems. These measurements were used in the generation of a subsequent set of DRT source tapes for which all DRT list segments on the tapes were equalized with respect to speech level. The effects of this equalization on intelligibility scores was evaluated for unprocessed and processed speech (both parametric and waveform coding algorithms). The manipulation of the speech levels for the DRT lists did not result in significant, predictable, and replicable effects on either system mean scores or on the DRT score variability across lists. The absence of effects was noted for all speech processing conditions and for both male and female speakers. Practical advantages in the use of the equalized word lists are discussed.

INTRODUCTION

In support of the DoD Digital Voice Processing Consortium's (DVPC) testing of speech compression algorithms the authors have been responsible for the preparation of diagnostic intelligibility test material. The measurement method used is the Diagnostic Rhyme Test [1]. The DRT utilizes a set of 96 rhyming word pairs to test the intelligibility of the present and absent states of six features of speech (i.e. voicing, nasality, sustention, sibilation, graveness and compactness). Various scramblings and speakers for the word set are available in the form of audio recordings. These tapes are used as the input test material for the voice processor or communications system being evaluated. The system's audio output is presented to a crew of listeners whose tasks are to indicate which of the rhyming words from each pair is perceived to have been presented. Final scores are corrected for guessing and averaged over the listener crew. These scores are presented as the percent correct listener response and are available for Speaker, feature/state and total system intelligibility.

Audible differences in speech level on the input tapes from one DRT list to the next have been noted by experienced DRT researchers with objective validation of these differences provided by VU meter readings during the playback of the lists. The level differences occur across lists read by different speakers as well as across lists read by the same speaker (i.e., different word scramblings) [2]. Based upon the assumption that there exists a functional relationship between the speech level of an input word list and corresponding DRT score, it is hypothesized that a reduction in speech level variability across these lists will result in a decrease in the variability of the DRT scores across these same lists. This relationship is not necessarily independent of type of processing performed on the DRT word list or the voice qualities of the speaker. In fact, various processors may not be optimized for the dynamic range required by some of the 3-Speaker input tapes. Experimental support for the contention that speech-level variability, as presented across DRT lists, is a source of DRT score variability has important ramifications for experimental designs established for making comparisons of intelligibility across different speech processing systems. Such designs involve multiple list presentations for each system with the variability across the lists presented for a given system being a component in the 'error variance' term used to test for significant systems effect (i.e., to test the hypothesis that the mean DRT scores for all systems under test are equal). The experiment's sensitivity to detect actual significant differences between processors is decreased as the uncontrolled error variance increases.

The current study was performed in order to test the hypothesis that a reduction in the variability of speech level for the various word lists presented to obtain an overall mean DRT score for a particular system will result in a reduction in the DRT score variability obtained from these separate lists. This hypothesis will be tested for both male and female speakers, and for both unprocessed and processed speech. Support for this hypothesis will provide validation for the use of speech level equalized lists as a means of increasing the resolution of the intelligibility testing methodology used to compare speech processing systems.

EXPERIMENTAL DESIGN

The experimental design used for this study consisted of the manipulation of three independent variables:

1. Speech-level equalized vs. unequalized DRT list.
2. Sex of Speaker.
3. Type of speech processing done to DRT word list.

This work has been supported by RADC/EEV under Contract No. F19628-83-C-0024.

CH2118-8/85/0000-0430 $1.00 © 1985 IEEE

The unequalized speech-level condition represents the use of recorded DRT lists taken from the DVPC's library which were known to have differences in speech level as quantified by an Equivalent Peak Level measurement algorithm [3]. The equalized speech-level condition, in contrast, represents the use of the same set of DRT lists which were re-recorded with record levels set such that the Equivalent Peak Level for all lists were made equivalent (within some tolerance range). The 'sex of speaker' and the 'type of processing' independent variables are introduced into the design to explore the possibility that the hypothesized relationship between speech level and intelligibility score variability is dependent upon the speech processing algorithm placed between the reading of the word list and the listener, and/or is dependent upon the speech characteristics of the speaker with the sex of the speaker being a classification variable used to make gross discriminations of these speech attributes. A full matrix of experimental conditions was obtained by crossing all three independent variables with one another. Each cell in the matrix represents the presentation of 9 separate DRT lists, each read by a different speaker.

The dependent variables are the total DRT score and the DRT speech feature scores provided by each of 8 listeners to whom the DRT list was presented.

PROCEDURE

A set of 18 DRT lists were selected from the library of analog tapes made available by the DVPC--9 lists read by males, and 9 read by females. These lists were for the quiet noise condition using an Altec 659A dynamic microphone without a puff screen. Speech level measurements were made for each list using the Equivalent Peak Level algorithm described below. On the basis of these measurements, second recordings of these same 18 lists were made in which all speech levels were equalized to a set reference level of 9 dBm. Both the 18 unequalized and the 18 equalized lists were used as speech input signals for 3 speech compression algorithms (CVSD, APC/SQ, LPC-10) while the analog output signals from these processings were recorded.

The DRT listener sessions were conducted at the RADC/EEV in-house speech evaluation facility [4]. During a given DRT test session, 9 unequalized lists produced by 9 speakers all of the same sex and receiving the same type of processing were presented to a group of female listeners with the number of listeners per session ranging from 9 to 12. Also presented during the same session were the matching 9 lists for which speech level had been equalized. The DRT data accumulated during one session was reduced such that only the scores for 8 listeners were retained for analysis. A single iteration of this data culling process eliminates the data for a listener who demonstrates the greatest amount of cumulative variation from the group's means for the individual lists in the complete set of 18 lists presented during the session. Repeated iterations of the process are made until the data for 8 listeners remains.

Measurement of Speech Level

The Equivalent Peak Level (EPL) is the quantity used to express the level of speech for the DRT lists used in this study. The EPL is a statistically estimated voltage level that instantaneous peaks of the speech signal under analysis often approach but rarely exceed. The estimation of this parameter is based upon a statistical model of instantaneous speech waveform magnitudes in which speech samples above a given threshold level can be modelled as a random variable having a log-uniform distribution between the threshold and a peak value. A functional relationship among the estimated peak value, the threshold level, and the Vrms value observed for only the portion of the signal exceeding the threshhold has been derived based upon this model. A known threshold value and an observed Vrms-above-threshold quantity can be applied to a piecewise-linear approximation to this functional equation with the result being the EPL estimate for the input signal. An EPL estimate was made for each DRT list used in the study by applying this algorithm to the first 40 seconds of the list.

Digital Speech Processors

All speech compression algorithms tested were available in real time on a special purpose DoD multiple rate processor. This unit is a 16 bit, fixed point, bit slice, parallel processor architecture with a fast multiply capability. The frequency characteristics are defined by a 100 Hz high pass filter and a 3600 Hz low pass filter. All of the algorithms run in full duplex, use the same audio processing front end and 12 bit A/D - D/A system. The 16 kbps Continuously Variable Slope Delta Modulation (CVSD) algorithm represents waveform coder algorithms. The 2.4 kbps Linear Predictive Coder (LPC) algorithm tested [5,6] is an early version of the DoD standard. It is a block form tenth order predictor utilizing the Choleski inversion, an AMDF pitch and voicing detector with a direct form recursive filter synthesis. This algorithm represents the parametric speech processors. The 9.6 kbps Adaptive Coder Segmented Quantizer (APC/SQ) algorithm [6] has a fourth order linear predictor, AMDF pitch detector and a residual error estimation algorithm. This type of algorithm is a combination of parametric and waveform coding.

RESULTS

Descriptive statistics for the EPL measurements made on the DRT word lists for male and for female speakers in both the speech-level equalized and unequalized conditions are presented at the top of Table 1. The significant reduction in the variability of the EPL measures in the equalized condition validates the experimental manipulation of this independent variable.

Table 2 is a key to the various 9-Speaker DRT system numbers and the test conditions that they represent. Table 3 presents the DRT feature score results for each system averaged across speakers and feature state (i.e. present or absent). Table 4 gives the total DRT results for individual speakers averaged across features and state. The systems in Tables 3

and 4 are grouped by sex, equalized or unequalized and process. All Standard Errors for Table 3 and for the total scores in Table 4 are calculated across listeners and speakers.

The F-ratios used to test for significant overall DRT score differences between the equalized and the unequalized lists (see 'Delta' values in Table 3) failed to reveal significant differences in any of the experimental conditions ($p < 0.05$). The differences between equalized and unequalized overall scores for CVSD processed male and female speech approach significance with the equalized mean DRT score being higher than the unequalized mean. A T-test analysis of these DRT means which makes no assumption regarding equality of population variances for the two means also failed to reveal significant mean differences.

An unbiased estimate of the DRT total score variability across a set of 9 lists for each of the experimental conditions is presented in Table 1. Graphical illustrations of this variability are presented in Figures 1A and 1B which provide plots of the overall mean for each of the systems evaluated, as well as the individual means for each list within a particular system. These figures also list the specific DRT source tapes, wordlists and EPL adjustments for each speaker. The F-ratio statistic used to test the null hypothesis that the DRT score variability across unequalized lists is not different from the variability across equalized versions of the same lists is also presented in Table 1. The rejection of this null hypothesis is indicated in the conditions involving male speakers with unprocessed speech and with LPC processing. Table 1 includes a linear correlation coefficient relating the change in DRT scores observed in going from an unequalized list to an equalized list, and the corresponding change in EPL.

The unequalized scores presented in Table 3 are essentially the same as previously reported results [7] for the three speech compression algorithms. The individual speaker results of Table 4 demonstrate differences when compared to these previous results. Differences in tapes and wordlists may be the source of this discrepancy although specific tape or wordlist identifiers are not provided with these previous results.

CONCLUSION

The use of speech-level equalized DRT word lists provides overall mean DRT scores for individual systems which do not differ significantly from the means obtained when a set of lists are used which have significantly more speech-level variability. Thus the mean score used to represent a particular communication system's intelligibility for purposes of system comparisons is not dependent upon the use of the equalized or unequalized DRT word lists.

Initial evaluation of the data does indicate that the DRT scores for equalized lists demonstrate less variability for unprocessed speech and LPC speech using male speakers, as compared to the DRT score variability across unequalized lists. A similar, although nonsignificant, trend is seen for female speakers with unprocessed and LPC speech. No reduction in DRT score variability is indicated for the other coding algorithms--CVSD,APC.

Attributing the cause of the observed reduction of DRT variability to the manipulation of the lists' speech levels is not supported by the correlation coefficients in Table 1 which indicate that a change in DRT score for the two presentations of the same list (equalized and unequalized) is independent of the measured change in speech-level for these two list presentations. Furthermore, an attempt to replicate these observed reductions of DRT variability for equalized lists in the cases of unprocessed and LPC processed male speech failed. In summary, the data collected in this study provides no support for the contention that a reduction in the variability of speech level for the word lists presented to a particular system will result in a reduction in the DRT score variability obtained from these separate lists.

The results of this study do not advocate for the use of the equalized DRT lists as a means of increasing the resolution of the experimental designs used to compare the intelligibility of speech processing systems. A practical advantage in using the speech-level equalized lists in processor evaluation studies is that once the level of the signal for the first list presented to the system under test has been set such that it is always in the linear range of operation for the system (i.e., there is no clipping of the signal), all subsequent lists can be assumed to present signals which are also maintained in the linear range of operation.

REFERENCES

1. W.D. Voiers, "Diagnostic Evaluation of Speech Intelligibility," Chap. 32, Speech Intelligibility and Recognition, M.E. Hawley, Ed., Dowden, Hutchinson and Ross, Stroudsburg, PA, (1977).
2. C.P. Smith "Talker Variance and Phonetic Feature Variance in Diagnostic Intelligibility Scores", Proc. IEEE ICASSP 1979.
3. P.T. Brady, "Equivalent Peak Level: A Threshold-Independent Speech-Level Measure", Journal of the Acoustical Society of America, Vol. 44, No. 3, 1968.
4. S. Meister, "The Diagnostic Rhyme Test: An Air Force Implementation", RADC-TR-78-1299, May 1978.
5. T.E. Tramain, et al, "Implementation of Two Real Time Narrowband Speech Algorithms", IEEE EASCON, Sept. 1978.
6. T.E. Tramain, "The Government Standard Linear Predictive Coding Algorithm: LPC-10", Speech Technology, April 1982.
7. C.P. Smith, "Narrowband (LPC-10) Vocoder Performance Under Combined Effects of Random Bit Errors and Jet Aircraft Cabin Noise", RADC-TR-83-293, Dec. 1983.

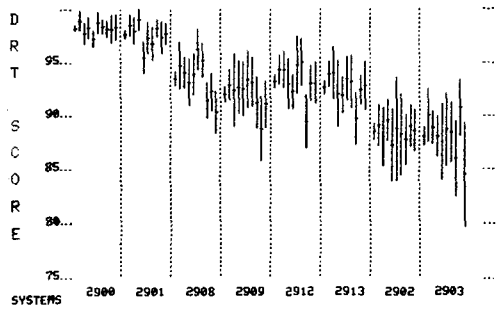

FIGURE 1A -MALES-
DRT RESULTS 02-JAN-85
TOTALS / ALL SPEAKERS
LISTENERS= 2
COMBINED FEATURE STATE

DATA ORDER	ITAPE	WLIST	DEPL
TOTAL SCORE			
SPEAKER LL	R1B	302B	-0.45
SPEAKER BV	R1B	303A	+1.44
SPEAKER PK	R1B	309A	+1.57
SPEAKER JE	R2A	305B	+1.43
SPEAKER CH	R2A	307A	+2.16
SPEAKER RH	R2A	310B	-0.66
SPEAKER AS	E1C	321A	-0.24
SPEAKER DP	E1C	324A	+1.33
SPEAKER IP	E1C	319A	+1.98

95% CI ON TOTALS FOR SE FROM L&S

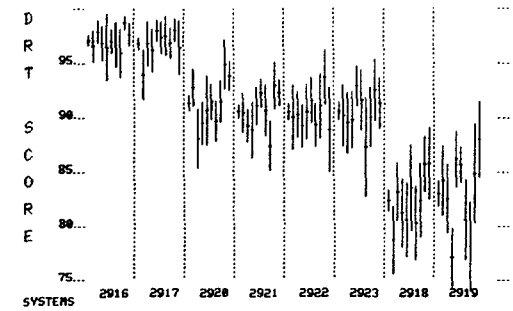

FIGURE 1B -FEMALES-
DRT RESULTS 02-JAN-85
TOTALS / ALL SPEAKERS
LISTENERS= 8
COMBINED FEATURE STATE

DATA ORDER	ITAPE	WLIST	DEPL
TOTAL SCORE			
SPEAKER VW	E514A	330B	+5.79
SPEAKER KS	E514A	334A	+3.22
SPEAKER MP	E514A	313A	+4.07
SPEAKER JS	E511B	317A	-0.28
SPEAKER LS	E511B	315B	+4.83
SPEAKER LV	E511B	325A	-0.23
SPEAKER AN	E512C	331B	+4.17
SPEAKER KG	E512C	335A	+5.31
SPEAKER LW	E512C	327A	-1.71

95% CI ON TOTALS FOR SE FROM L&S

MALES / EPL EQ & ORIGINAL / ALL PROCESSORS
SPEAKERS ARE LL BV PK JE CH RH AS DP IP

SYSTEM 2900	12-OCT-84 EPL EQ SETA 9M G/DYN
SYSTEM 2901	12-OCT-84 ORIGINAL SETA 9M G/DYN
SYSTEM 2908	24-OCT-84 EPL EQ SETA CVDS-16 9M G/DYN
SYSTEM 2909	24-OCT-84 ORIGINAL SETA CVSD-16 9M G/DYN
SYSTEM 2912	09-NOV-84 EPL EQ SETA APC 9M G/DYN
SYSTEM 2913	07-NOV-84 ORIGINAL SETA APC 9M G/DYN
SYSTEM 2902	12-DEC-84 EPL EQ SETA LPC-10 9M G/DYN
SYSTEM 2903	12-DEC-84 ORIGINAL SETA LPC 9M G/DYN

FEMALES / EPL EQ & ORIGINAL / ALL PROCESSORS
SPEAKERS ARE VW KS MP JS LS LV AN KG LW

SYSTEM 2916	09-NOV-84 EPL EQ SETA 9F G/DYN
SYSTEM 2917	16-NOV-84 ORIGINAL SETA 9F G/DYN
SYSTEM 2920	05-DEC-84 EPL EQ SETA CVSD-16 9F G/DYN
SYSTEM 2921	05-DEC-84 ORIGINAL SETA CVSD-16 9F G/DYN
SYSTEM 2922	07-DEC-84 EPL EQ SETA APC 9F G/DYN
SYSTEM 2923	07-DEC-84 ORIGINAL SETA APC 9F G/DYN
SYSTEM 2918	14-DEC-84 EPL EQ SETA LPC 9F G/DYN
SYSTEM 2919	14-DEC-84 ORIGINAL SETA LPC 9F G/DYN

TABLE 2
DRT SYSTEM KEY

SEX OF SPEAKER

	MALE	FEMALE
	EQUIVALENT PEAK LEVEL [dBm] Unequalized / Equalized Mean 7.81 / 8.75 S.D. 1.11 / 0.24 Range 6.5-9.6 / 8.4-9.0	EQUIVALENT PEAK LEVEL [dBm] Unequalized / Equalized Mean 5.66 / 8.45 S.D. 3.15 / 0.44 Range 2.4-10.9 / 8.0-9.2
UNPROCESSED	r = -0.03 DRT VARIANCE Unequalized / Equalized 1.06 / 0.26 F-ratio = 4.10*	r = 0.36 DRT VARIANCE Unequalized / Equalized 1.53 / 0.70 F-ratio = 2.18
CVSD	r = 0.03 DRT VARIANCE Unequalized / Equalized 2.14 / 3.61 F-ratio = 0.59	r = 0.34 DRT VARIANCE Unequalized / Equalized 3.23 / 4.72 F-ratio = 0.68
APC	r = -0.04 DRT VARIANCE Unequalized / Equalized 1.75 / 2.95 F-ratio = 0.59	r = 0.86 DRT VARIANCE Unequalized / Equalized 2.96 / 2.17 F-ratio = 1.36
LPC	r = -0.08 DRT VARIANCE Unequalized / Equalized 3.68 / 0.54 F-ratio = 6.81*	r = 0.13 DRT VARIANCE Unequalized / Equalized 14.05 / 5.90 F-ratio = 2.38

r = correlation coefficient for change in DRT vs. change in EPL, n=9
DRT VAR. = unbiased estimate of DRT variance across lists, n=9
F-ratio = test of null hypothesis that the variance across equalized lists equals variance across unequalized lists, df=(8,8)

TABLE 1
STATISTICS RELATING DRT SCORES AND EPL

MALES / EPL EQ & ORIGINAL / ALL PROCESSORS

SYSTEM	V	N	SU	BI	Q	C	TOTAL	
2900	97.2	99.4	99.0	99.3	94.9	99.7	98.2	MEAN
	0.58	0.22	0.27	0.23	0.60	0.17	0.14	S.E.
							0.6	DELTA
2901	96.9	98.7	98.5	98.6	94.1	99.3	97.7	MEAN
	0.65	0.35	0.40	0.33	0.65	0.23	0.20	S.E.
2908	96.2	98.4	95.5	91.2	83.0	97.1	93.6	MEAN
	0.74	0.46	0.69	1.12	1.15	0.54	0.33	S.E.
							1.5	DELTA
2909	94.5	97.4	93.1	90.1	81.1	96.2	92.1	MEAN
	0.86	0.74	0.79	1.20	1.16	0.71	0.38	S.E.
2912	95.1	97.9	92.2	95.1	83.6	95.9	93.3	MEAN
	0.73	0.43	0.89	0.68	1.14	0.77	0.34	S.E.
							0.7	DELTA
2913	93.5	98.1	90.3	95.3	82.6	95.9	92.6	MEAN
	0.87	0.46	0.96	0.67	1.20	0.72	0.33	S.E.
2902	92.4	93.8	86.6	91.3	77.1	89.7	88.5	MEAN
	0.99	0.87	1.19	0.85	1.23	0.79	0.40	S.E.
							0.4	DELTA
2903	92.2	92.0	87.3	92.0	76.3	88.5	88.1	MEAN
	0.89	1.15	0.97	0.93	1.15	1.04	0.47	S.E.

FEMALES / EPL EQ & ORIGINAL / ALL PROCESSORS

2916	95.4	98.8	97.8	99.4	93.4	97.8	97.1	MEAN
	0.67	0.36	0.43	0.22	0.75	0.57	0.25	S.E.
							0.3	DELTA
2917	95.2	99.1	96.5	99.0	93.6	97.5	96.8	MEAN
	0.73	0.26	0.66	0.32	0.75	0.53	0.28	S.E.
2920	94.3	99.3	93.5	89.1	76.5	96.0	91.4	MEAN
	0.71	0.32	0.73	1.21	1.29	0.69	0.37	S.E.
							0.8	DELTA
2921	93.3	99.0	92.4	86.4	76.8	95.8	90.6	MEAN
	0.87	0.29	0.96	1.15	1.04	0.63	0.33	S.E.
2922	92.4	98.1	90.1	87.0	79.8	96.2	90.6	MEAN
	0.82	0.42	0.92	1.16	1.06	0.71	0.39	S.E.
							0.0	DELTA
2923	91.7	98.2	88.4	86.9	82.8	95.9	90.6	MEAN
	0.95	0.42	1.00	1.17	1.11	0.70	0.43	S.E.
2918	89.5	89.7	76.1	79.8	74.6	84.9	82.4	MEAN
	1.13	1.29	1.53	1.31	1.50	1.08	0.51	S.E.
							-0.6	DELTA
2919	89.6	91.2	80.4	79.6	72.4	85.1	83.0	MEAN
	0.99	0.95	1.41	1.18	1.40	1.18	0.61	S.E.
SYSTEM	V	N	SU	BI	Q	C	TOTAL	

NOTE S.E. ACROSS LISTENERS AND SPEAKERS

TABLE 3
DRT FEATURE SCORES ACROSS SPEAKERS AND STATES

MALES / EPL EQ & ORIGINAL / ALL PROCESSORS

SYSTEM	LL	BV	PK	JE	CH	RH	AS	DP	IP	TOTAL	
	SPEAKERS										
2900	97.0	97.8	98.3	97.3	98.8	98.4	98.2	98.2	98.3	98.2	MEAN
	0.39	0.41	0.39	0.34	0.41	0.28	0.33	0.55	0.48	0.14	S.E.
2901	98.6	98.0	99.1	95.6	97.4	96.9	98.3	97.5	97.8	97.7	MEAN
	0.44	0.54	0.41	0.61	0.62	0.65	0.34	0.65	0.41	0.20	S.E.
2908	94.8	94.1	93.2	94.0	96.4	95.3	91.5	92.3	90.4	93.6	MEAN
	0.92	0.62	0.94	0.83	0.81	0.68	0.72	0.76	0.85	0.33	S.E.
2909	93.0	92.4	92.7	92.6	93.5	93.2	91.3	88.8	91.1	92.1	MEAN
	0.61	1.46	1.06	1.09	1.14	0.98	1.04	1.29	0.94	0.38	S.E.
2912	94.4	94.4	93.1	92.3	94.8	95.1	89.5	93.1	93.1	93.3	MEAN
	0.59	0.73	1.00	0.71	1.13	0.94	1.07	0.71	0.83	0.34	S.E.
2913	93.9	94.0	92.1	91.9	93.5	93.2	89.7	92.4	92.8	92.6	MEAN
	0.50	1.04	1.32	0.73	0.87	1.06	1.05	0.58	0.97	0.33	S.E.
2902	89.1	88.0	89.6	87.2	88.8	88.3	87.8	89.1	88.7	88.5	MEAN
	0.81	1.16	0.84	1.44	2.07	1.63	1.02	0.84	0.82	0.40	S.E.
2903	90.0	88.8	88.0	87.5	88.7	88.4	85.9	90.8	84.5	88.1	MEAN
	1.06	0.67	0.81	1.52	1.45	1.20	1.49	1.14	2.06	0.47	S.E.

FEMALES / EPL EQ & ORIGINAL / ALL PROCESSORS

SYSTEM	VW	KS	MP	JS	LS	LV	AN	KG	LW	TOTAL	
2916	96.6	97.9	96.9	96.5	97.0	96.7	96.0	98.7	97.7	97.1	MEAN
	0.61	0.44	0.65	1.29	0.46	0.87	0.93	0.26	0.43	0.25	S.E.
2917	94.0	96.9	96.2	98.0	97.4	97.5	96.9	98.0	96.5	96.8	MEAN
	0.98	0.86	0.83	0.41	0.62	0.76	0.59	0.46	1.04	0.28	S.E.
2920	92.8	88.2	89.6	90.8	91.5	89.7	91.5	94.9	93.9	91.4	MEAN
	0.72	1.15	0.84	1.36	0.69	0.80	0.82	0.95	0.57	0.37	S.E.
2921	90.5	89.3	88.9	91.1	92.3	90.8	87.5	93.0	92.3	90.6	MEAN
	0.77	0.65	1.10	0.74	0.55	0.99	0.96	0.92	0.52	0.33	S.E.
2922	90.1	90.4	89.3	90.6	91.7	89.5	91.1	93.8	88.9	90.6	MEAN
	1.26	0.90	0.83	1.08	0.90	0.85	1.26	1.06	1.65	0.39	S.E.
2923	90.2	89.6	89.8	93.0	91.7	87.4	90.1	92.6	91.4	90.6	MEAN
	1.20	1.18	1.09	0.78	1.18	1.94	1.16	1.20	1.00	0.43	S.E.
2918	78.8	83.2	81.2	80.6	83.6	80.3	82.4	85.8	85.8	82.4	MEAN
	1.35	1.14	1.35	1.47	1.69	1.48	1.45	1.08	1.41	0.51	S.E.
2919	84.2	82.6	77.1	86.2	85.7	80.6	78.1	84.9	88.0	83.0	MEAN
	1.37	1.36	1.16	1.09	0.70	1.59	1.76	1.94	1.47	0.61	S.E.

NOTE S.E. ON TOTAL SCORES ACROSS LISTENERS AND SPEAKERS

TABLE 4
DRT SPEAKER SCORES ACROSS FEATURES AND STATES